Modeling a digital twin for the optimization of a self-supply energy system for residential use

Laura Rodríguez de Lope
Communications Engineering Dpt.
Universidad de Cantabria
Santander, Spain
laura.rdelope@unican.es

Víctor M. Maestre
Chemical & Biomolecular Engineering Dpt.
Universidad de Cantabria
Santander, Spain
victormanuel.maestre@unican.es

Luis Diez
Communications Engineering Dpt.
Universidad de Cantabria
Santander, Spain
luisfrancisco.diez@unican.es

Alfredo Ortiz
Chemical & Biomolecular Engineering Dpt.
Universidad de Cantabria
Santander, Spain
alfredo.ortizsainz@unican.es

Ramón Agüero
Communications Engineering Dpt.
Universidad de Cantabria
Santander, Spain
ramon@tlmat.unican.es

Inmaculada Ortiz
Chemical & Biomolecular Engineering Dpt.
Universidad de Cantabria
Santander, Spain
inmaculada.ortiz@unican.es

***Abstract*—The climate situation and the energy crisis have prompted a number of policies and strategies that foster the adoption of renewable energy sources. To tackle the intermittency and fluctuations associated with the operation of these sustainable energy sources, renewable hydrogen appears as an appealing solution to decarbonize different economic sectors. In this sense, the design and implementation of a hybrid renewable energy-hydrogen system has led to the first electrically self-sufficient social housing in Spain, located in the town of Novales (Cantabria). On the other hand, the digitization of this type of self-sufficient systems would allow automatic adaptation to changing situations, increasing energy efficiency. In this context, we introduce the design and initial implementation phases of a digital twin architecture that, using machine learning and artificial intelligence techniques, facilitates the optimization of the performance of the physical system by interacting with its control components. This involves the use of telemetry solutions that allow the capture and storage of data from the physical system itself, as well as from the environment, such as instance meteorological data. We also discuss some initial results of the digital twin, which features models of the electrical components of the physical system, based on both their logical behavior and machine learning techniques.**

***Index Terms*—digital twin, renewable energy, hydrogen.**

I. Introduction

The current juncture of climate and energy crisis has led to the implementation of policies aimed at promoting the use of renewable energy sources (RES) with the purpose of fostering energy independence through sustainable solutions.

Since the Conference of the Parties (COP) 21, held in Paris in 2015 [1], various roadmaps and strategies have been promoted to mitigate the harmful effects of climate change. In this sense, the current society's dependence on fossil fuels stands out as the predominant factor responsible for the global climate situation. Specifically, energy related activities contribute to $> 75\%$ of the overall emissions of equivalent carbon dioxide (CO2eq) [2], [3]. In this context, the European Union (EU) has approved the "Fit for 55" plan, which includes a 55% reduction in greenhouse gas emissions (GHG) by 2030. On the other hand, the current political instability, has brought a high inflation, and a shortage of fossil fuels, mainly natural gas and oil, imported from Russia [4].

In this scenario, the large-scale implementation of renewable energy sources becomes imperative to ensure a decarbonized energy system that simultaneously provides a degree of energy autonomy through efficient and sustainable solutions. However, it is essential to find effective technological solutions for energy storage that respond quickly, safely, and flexibly to the intermittent and fluctuating nature of RES. For this reason, the European Commission has approved the REPowerEU plan for the year 2022. According to the policies included therein, the EU aims to promote the energy independence of the continent by fostering the use of RES, increasing energy efficiency, and developing the hydrogen economy. In this context, the use of hydrogen as an energy carrier and raw material emerges as an efficient and sustainable solution for large-scale and seasonal energy storage. It presents itself as a suitable alternative to promote the presence of RES in the energy sector, and to facilitate the decarbonization of various sectors related to energy [5].

In particular, the residential sector becomes critical, due to its status as a massive energy consumer in the EU, contributing to 40% of the overall consumption. Additionally, it is currently a rather inefficient sector, largely due to its aging infrastructure, which negatively impacts its carbon footprint. On the other hand, unprecedented inflation in the economy has exacerbated the situation for the most vulnerable citizens, often inducing energy poverty. It thus becomes imperative enhancing the energy performance of the residential sector, to reduce its impact on climate change and to alleviate the rising costs of electricity bills, which affect the population's living standards [6].

We start from a hybrid renewable energy-hydrogen system in the town of Novales (Cantabria), which meant the first socially subsidized residence to achieve electrical self-sufficiency in Spain. Over such installation, we propose developing a digital twin (DT) of the pilot plant. Its main objective is to improve the performance of the renewable hybrid system (RHS) through specific algorithms developed and validated on the digital replica. This work describes the architecture of the DT, comprising the digital model of the pilot plant, the communication interface for data collection and signal transmission, and a module designed to facilitate data sharing with third parties.

In summary, the main contributions of this paper are:

- We introduce the design of a complete DT for a renewable energy-hydrogen system. To the authors' best knowledge, there do not exist previous works who have exploited the DT concept to model this type of systems.
- We use a real physical system, which is being used to tune the specification of the modules of the proposed DT. This is of utter relevance, since we exploit the real information about the behavior of such system, allowing us to use Artificial Intelligence and Machine Learning (AI/ML) to promote more accurate operation of the digital replica.
- We discuss initial results of our DT, which show the benefits of fostering AI/ML based techniques to mimic the behavior of the real counterparts, compared to model-based approaches.

The rest of this paper is structured as follows: Section II provides a description of the current state of the art on the use of digital twins in the energy sector. Section III details the architecture of both the physical system and the digital replica. Subsequently, Section IV focuses on the development of the models that constitute the DT, and Section V discusses the results obtained with the initial integrated implementations. Finally, Section VI concludes the paper, providing an outlook of our future work.

II. State of the Art

The concept of DT was originally coined at the turn of the century, in particular for industrial environments [7], although its use has recently spread to different sectors [8], [9], taking advantage of recent progress in digitization and the augmenting capabilities of communication and computing systems. As explained in [10], the increasing complexity in processes can only be replicated exploiting AI and ML techniques, whose relevance might be rather strong when applied in DTs.

In the case of the chemical sector, one scenario of particular interest is that of renewable energy sources. In this case, as mentioned by the authors in [11] and [9], there are very few works that have tried to apply the DT concept to this type of systems. Moreover, the authors of [12] conclude that there do not exist any in-depth studies on the use of DT in this sector.

One of the few papers applying the digital twin concept to the energy domain is that of Nguyen *et al.* [13]. The authors propose the use of a DT to improve the performance of power distribution systems, highlighting its ability to take optimal control decisions based on analyses that are carried out in real time. Similarly, Agostinelli *et al.* discuss in [14] the potential benefits of DT in the management of energy distribution and consumption in buildings, and they highlight the role that artificial intelligence techniques could play.

Focusing on the hydrogen production process, [15] proposes the use of a DT to address the uncertainties associated with the investment and operating costs of the system. However, they focus on how different factors might impact financial indicators, while we exploit the DT to improve the performance of the system.

As can be seen, despite their clear potential benefits, the use of digital twins in the energy realm in general, and in renewable hydrogen-based systems in particular, is still very limited. Hence, the approach that we introduce in this paper contributes to the integration of these two aspects, exploring the advantages that they might bring.

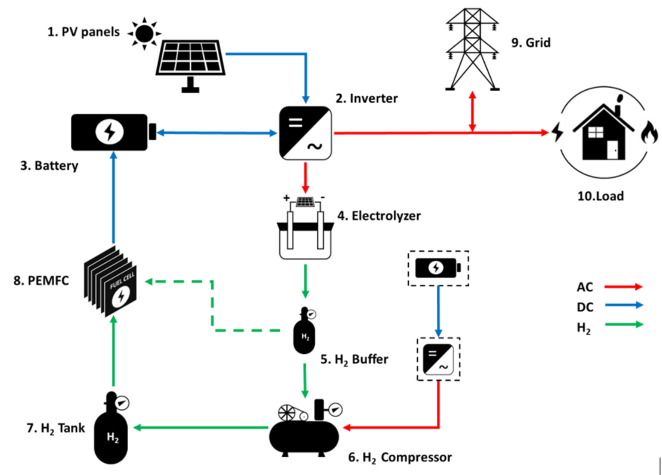


Fig. 1: Renewable hydrogen-based system architecture

III. Methodology

This section first describes the real renewable hydrogen system (physical system), identifying its elements and depicting its general operation. Then, the overall design of the digital twin is discussed, illustrating its behavior.

A. Pilot plant

The real system consists on a RHS, which was designed and deployed in the framework of the SUDOE ENERGY PUSH project[1]. It combines both renewable energies and novel hydrogen-based technologies to ensure complete electrical self-sufficiency of a social housing throughout the year. Fig. 1 illustrates the architecture of the pilot plant, where we also identify the electricity and hydrogen flows within the system. A through discussing of the pilot plant can be found in [16].

Photovoltaic panels (point 1 in Fig. 1) installed on the roof of the building collect solar energy to supply the house (point 10 in the figure) as a primary source. Whenever there is any excess energy after supplying the house, it is first stored in a set of lithium-ion batteries (point 3), which store energy

[1]https://www.sudoe-energypush.eu/

for short-term consumption. If the energy excess is high, it is used for the generation, compression, and storage of hydrogen for seasonal energy savings. The hydrogen is generated by an electrolyzer (point 4), which creates hydrogen with an electrolysis-based procedure, which is powered by electricity. This hydrogen is first stored in a buffer (point 5). When the buffer is complete, the hydrogen is compressed and stored in a high-pressure tank (points 6 and 7). In case that there is still excess electrical energy from the solar panels, after supplying these processes, it is fed into the grid (point 9). During periods of photovoltaic energy deficit, the batteries supply electricity to the house and, when they reach a certain discharge threshold, they are charged by a fuel cell (point 8), which finally covers the household's demand. This fuel cell generates electricity from the hydrogen stored both in the buffer or in the high-pressure tank.

The operation of the pilot plant has been fully automated and it is remotely controlled with the help of a programmable logic controller (PLC). In addition, the RHS operates with an energy management strategy based on the status of the stored autonomy, and is continuously monitored by means of a supervisory control and data acquisition (SCADA) system.

B. Digital twin architecture

Over the pilot plant described above, we introduce in this work the design and development of a DT where the plant components are characterized, and control solutions will be developed to address the automatic improvement of the real system parameters. For their modeling, machine learning (ML) and artificial intelligence (AI) techniques will be applied to the collected data when the behavior of the model-driven approach does not accurately mimic the real operation of the RHS.

To develop the DT, we propose an architecture with three main stages. The first one, focused on the physical-virtual interaction, is in charge of the collection of information from the real system, as well as the implementation of decision policies. Moreover, a module for the management and integration of data from external sources, such as weather forecasts or energy prices, is also envisaged. In a second stage, and taking advantage of the aforementioned modules, the DT model is implemented by a set of software libraries that replicate the behavior of the real system. In this regard, once the input/output and control variables of the main components of the RHS have been identified, we address its modeling. It will adopt a model-driven approach when the underlying behavior is well known, and data-driven AI/ML techniques otherwise. Finally, we plan to use the DT model to evaluate the performance of various control policies on the digital replica, including those based on weather forecasting.

Fig. 2 illustrates the complete DT high level architecture, including the underlying logic flow of its operation. As was mentioned above, the pilot plant uses a SCADA system for monitoring the performance of the real devices. Thus, the DT will interact with the SCADA system through the PLC to collect data and apply the appropriate control actions, as depicted in the points 1 and 4 of the figure.

The logic of the digital replica consists of a main component, the DT model, which captures the behavior and performance of the pilot, and the management and control system, in which control strategies are implemented. As can be seen in Fig. 2, we will follow a loop-based approach to ensure that the behavior of the real system is accurately captured (points 5 and 6 of Fig. 2): (i) analysis of the control strategies on the DT model to optimize the performance of the physical system; (ii) implementation of the strategy on the real pilot, by forwarding control commands that interact with the deployed SCADA system; (iii) the system will continue receiving feedback from the physical pilot (continuous monitoring) to further train the DT behavior for those cases where ML solutions are adopted.

Finally, although its design assumes the DT to run in a closed way, we will also exploit the data generated during its operation. In this sense, those pieces of information that are considered to be more relevant, will be made available to third parties and interested stakeholders, as shown in Fig. 2. Thus, the data generated by the DT will be made available at a data marketplace [17], such as that promoted by the FIWARE initiative, as well as at open access repositories such as Zenodo. In this sense, we need to adopt open data models, such as Smart Data Models (SDM), which facilitate the interoperability and reuse of information by third parties. In case there are no available models for the specific needs of the DT, new definitions will be proposed to extend the SDM repository, making the datasets that are created during the DT operation open and accessible.

Having outlined the general architecture of the system, the following sections delve into the DT logic modeling, and discuss the first results obtained.

IV. Modeling

The DT model is implemented as a set of independent and interconnected software modules, each one of them modeling one or more physical components of the physical RHS. This disaggregated solution facilitates the implementation and independent validation of each component. On the other hand, it also allows the replacement of the models applied to different specific components, without affecting the system as a whole.

We have identified the set of variables that influence the performance of each module, classifying them as control and input variables; we have also established the outputs generated by each of them. These input and output variables correspond to physical magnitudes of the system, while the control variables mimic decision signals, such as those generated by the PLC. Thus, when certain input and control variables are applied to a module, it generates the corresponding outputs, replicating the behavior of its physical counterpart. The different modules are connected so that the outputs of one module can act as input variables for other modules. In fact, some signals are fed back into the system, such as the State of Charge (SOC) of the battery and the pressures of the hydrogen storage block components, which are sent to the PLC for decision making.

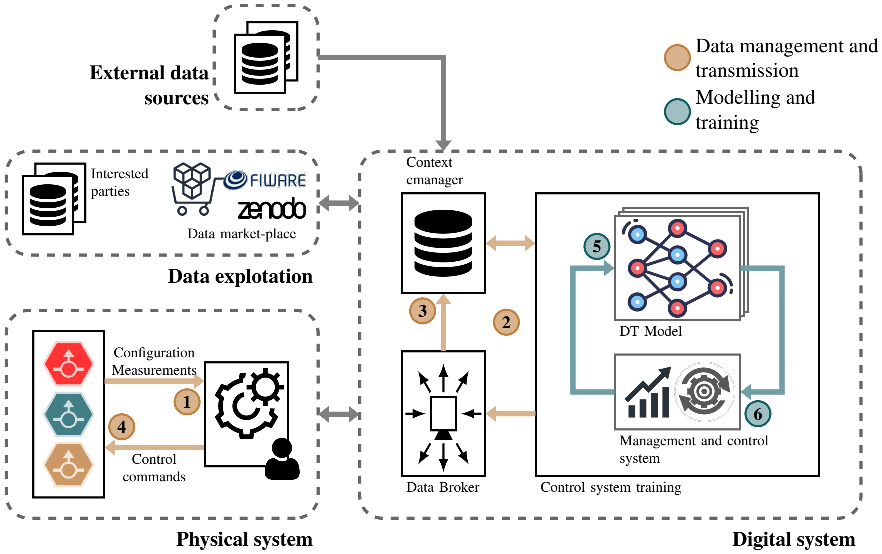


Fig. 2: Digital twin high level architecture

Fig. 3 shows the modules that we have identified, together with the corresponding variables, differentiating between external (grey), system (blue) and control (orange) variables. The former are independent of the system, while the latter are modeled by the DT, and their states depend on the external variables, the previous state of the system and/or the established decision policy. As can be seen in the figure, the DT only requires the power demanded by the dwelling and the photovoltaic (PV) power, which correspond to external variables, while the rest of variables are estimated by the DT itself. In the future evolution of the DT model other external data sources will be integrated, such as the weather forecast or electricity pricing, since they might have a direct impact over the overall system performance.

Although the first steps of this work have been focused on the implementation of the electrical components' modules, such as the PLC, the inverter and the battery, we describe below all the modules that constitute the proposed DT, while the input and output parameters for each of them are listed in Table I.

A. PLC

This is the main component of the system, as it contains the overall management logic of the system. This module takes as input variables the home power demand, and the generation of photovoltaic energy from the solar panels. These variables, together with other internal ones that establish the situation of the system, such as the SOC and the pressure of the hydrogen storage systems, are used to estimate the output variables, which in this case correspond to the control signals for switching on/off the rest of the modules. Initially, the logic implemented in the PLC follows the behavior that is currently configured in the real system, as described in Section III, to facilitate the validation of the other models. Subsequently, the implementation of different control algorithms will lead to the modification of this logic.

TABLE I: Input/Output variables of each component

Input variables	Output variables
PLC	
PV panels power	Electrolyzer on/off signal
Home demanded power	Compressor on/off signal
SOC	Fuel cell on/off signal
Buffer and tank pressures	
Inverter	
PV panels power	Charge/discharge battery power
Home demanded power	Imported/ exported grid power
Electrolyzer power	
Compressor power	
Battery	
Charge/discharge battery power	SOC
Fuel cell power	
Electrolyzer	
Electrolyzer on/off signal	Electrolyzer power
	Hydrogen generated
Hydrogen Storage	
Compressor on/off signal	Compressor power
Hydrogen generated	Buffer pressure
Hydrogen consumed	Tank pressure
Fuel cell	
Fuel cell on/off signal	Power generated
Hydrogen consumed	

B. Inverter

The inverter module uses as input variables the power generated by the PV panels, the power demanded by the house, as well as the power required by both the electrolyzer and the compressor. The proposed model distributes the energy excess/deficit, generated from the difference between the demand of the house and the production of the solar panels, to/from the different energy storage modules and, if necessary, to/from the grid.

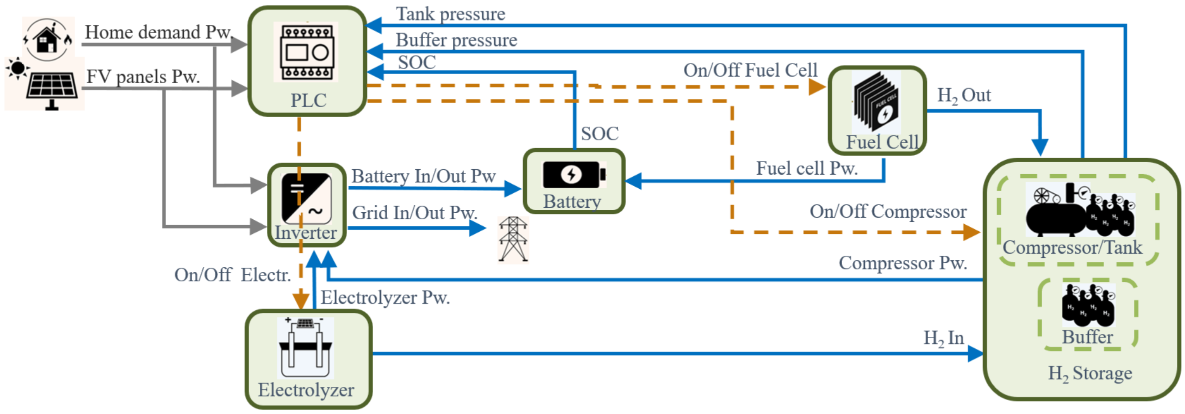


Fig. 3: Modules and variables of the DT

C. Battery

The battery module uses the charge/discharge power provided by the inverter module to update the SOC. In case the fuel cell is on, the generated power is also used as an input variable to establish the SOC. The output variable of this module, SOC, is fed back into the system, to be used by the PLC in its decision making processes.

D. Electrolyzer

The electrolyzer module models, from the on/off signal provided by the PLC, the flow of generated hydrogen and the power consumed by this module. These variables will also act as inputs for the hydrogen storage and inverter modules, respectively. Other variables, such as the ambient temperature, will also be considered in future phases, to analyze their potential impact over the behavior of this block.

E. Hydrogen storage blocks

This module consists of a first hydrogen storage block in a buffer and a second one, which includes a compressor and a storage tank. The on/off signal that goes from the PLC to the compressor establishes whether the hydrogen generated by the electrolyzer should be stored in the buffer, or in the compressed hydrogen tank. The hydrogen input/output flow determines the pressure on these blocks, which will be fed back to the PLC, as they are required for the control logic decision making.

F. Fuel cell

It receives the on/off signal from the PLC when the battery is below a certain threshold. It provides, as output variables, the consumed hydrogen and the generated power, which then update the status of the hydrogen storage and battery modules, respectively.

As an initial step, the aforementioned modules are implemented using a model-driven approach that replicates the logical behavior of the physical components according to their physical phenomena. In those cases where such algorithmic solutions are not as accurate as it would be desired, data-driven AI/ML models, trained with the collected data, are adopted. The initial analyses using AI/ML solutions have been performed based on three types of neuronal networks that allow both the modeling of complex functions and the prediction of time series:

- Dense Neuronal Network (DNN): Composed by a set of neurons, each of them applying a non-linear function to a combination of the module inputs. The non-linear function (activation) can be modified to identify those with a better behavior. At the moment, Rectified Linear Unit (ReLU) functions are used.
- Convolutional Neuronal Network (CNN): The operation within the neurons is a convolution of the corresponding inputs, and the used weights are being continuously updated. Each output is a linear combination of the current input and a number of previous inputs. In the training phase, the CNN learns the linear coefficients, i.e. the weight of prior values on the current output.
- Recurrent Neuronal Network (RNN): They are characterized by their memory.In particular, we use the well known Long Short-Term Memory (LSTM) neurons, which store inputs sequences. They obtain information from previous inputs to influence the current output.

V. Results

During the initial phase of the DT model development, the activity has focused on the implementation of the modules of the electrical components of the system: PLC, inverter and battery. The modeling has followed an algorithmic approach, using the logical operation of each of them, as well as the equations governing their behavior. Subsequently, other data-driven approaches are adopted, depending on the accuracy of the observed prediction. To validate that the proposed models provide an accurate behavior, capturing that of their physical counterpart, this section presents an analysis of the results of the aforementioned components, focusing on the interaction between the inverter and the battery.

The results shown below are based on the data obtained after sampling the physical system, measuring the state of each variable within a 7 day interval, with a sampling rate of 5 seconds, which allows sufficient granularity to capture abrupt changes in the state of the variables, such as peaks in the energy demand of the dwelling. For each sampling instant, the DT generates an estimation of the corresponding variables.

In the case of the inverter, the analysis compares the real and estimated values of power supplied/consumed to/from the battery, being the input signals the generated PV power, the power demand of the household and the estimated PLC signals. In the case of the battery, the SOC output variable is assessed by means of its charge/discharge power, which is used as input signal for the component.

A. Algorithmic model

This section presents the results obtained by the application of model-driven approaches to the DT modules.

Fig. 4 shows the instantaneous evolution of the battery discharge power during a sequence of $1e5$ samples. The shaded background represents the difference between the two input variables, the PV power and the household demand, whose scale is indicated on the right-hand axis of the figure (PV/home diff.). The red and green dots correspond to the real and estimated values of the battery discharge, respectively.

As can be seen in Fig. 4, the battery discharge occurs mainly during the night period, when there is an energy deficit (negative PV/home diff). During the day, the discharge power remains zero, with the exception of certain home demand peaks. It can be seen that the model behavior follows the same trend as that observed for the physical system, providing estimated values very close to the real ones in most cases. However, differences can be observed when sudden power discharges occur, corresponding to the peaks that can be observed in the figure. This is due to the fact that the grid is more reactive than the battery, and so reacts more quickly to sudden power demands. This phenomenon results in an import of energy from the grid, requiring less power to be discharged from the battery. This behavior of the real system is hard to predict with the model-driven approach, which always performs a maximum battery discharge before importing power from the grid. In addition, the particular values that would trigger the power consumption from the grid would also depend on the devices that are being used, which would make replicating its behavior harder.

Fig. 5 shows the relationship between the measured and estimated battery discharge power values during the sampling period. Each of the blue dots represents the discharge power estimated by the DT versus the one measured by the real system at a particular time instant. The ideal behavior is represented by a line (orange), where the estimated values and the real samples match. As can be observed, the model is able to accurately reflect the real behavior in most cases. However, for certain extreme values, either maximum or minimum, some mismatches might occur. On the left side of the figure it can be appreciated that the model provides values over the whole range, while the measurement system did not detect any discharges. This behavior is due to gaps in the monitoring process in the physical system, in particular by the PLC, during certain time intervals, where null measurements are provided. The aforementioned phenomenon, in which abrupt power demands occur, is also reflected in the upper right part of the figure, where it can be seen that the estimated discharge power reaches its maximum value, while the real one adopts intermediate values due to the faster reaction of the electrical grid to these unexpected power consumption demands.

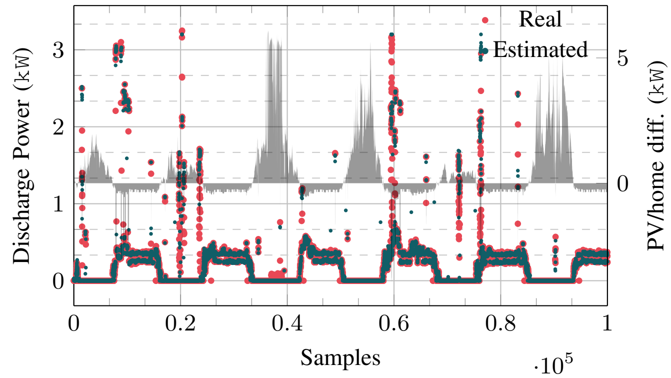
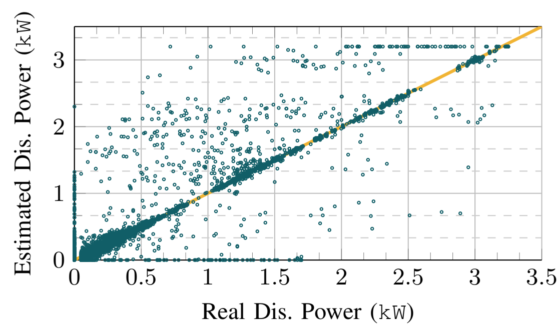


Fig. 4: Instantaneous battery discharge power. Real (red) and estimated (blue).

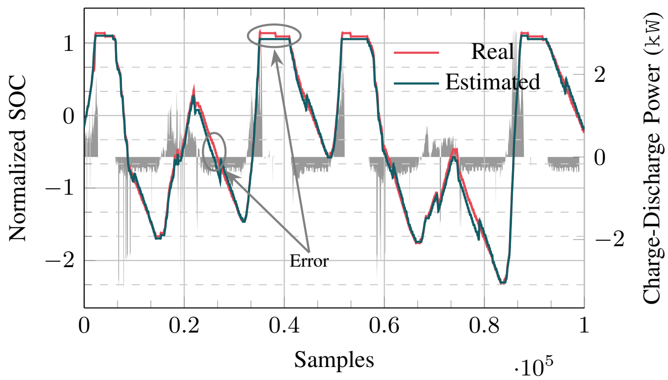


centering

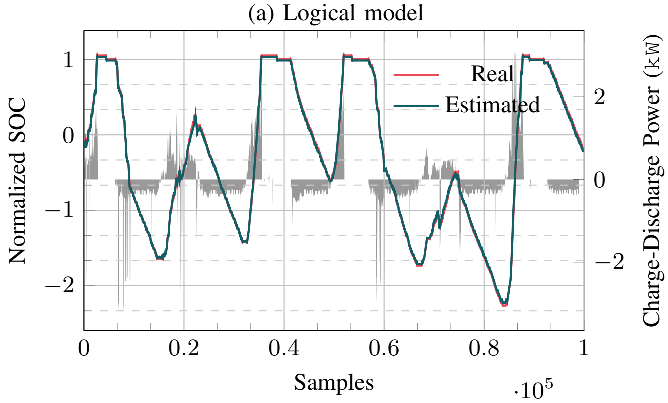
Fig. 5: Battery discharge power. Real vs. Estimated

To numerically characterize the deviation between the values provided by the DT and the real ones, the normalized root mean square error (NRMSE) has been calculated from all battery discharge power estimations, yielding a value of 0.0279, which evinces the high accuracy of the proposed model. Nonetheless, in the following subsection, we discuss the results obtained using machine learning techniques to replicate the system reaction to sudden demands, and we will see that they are able to reduce this deviation.

To analyze the behavior of the battery module, the SOC output signal has been compared with its real value, which is monitored in the physical system. Fig. 6a shows the time evolution of the real battery normalized SOC values, and the one estimated by the DT, with red and blue lines, respectively. For each sample s in a set of samples S, the scaled value $\hat{s}$ is obtained as $\hat{s} = (s - \mathbb{E}[S])/\sigma[S]$; where $\mathbb{E}[S]$ and $\sigma[S]$ hold for the average value and standard deviation of the set of samples S.



(a) Logical model



(b) DNN-based model

Fig. 6: Real vs. Estimated SOC for the Logical model (top) and the one based on DNN (bottom)

In order to better illustrate the observed behavior, the shaded gray background illustrates the difference of the input variables (charge and discharge power), whose scale is indicated in the the right-hand axis.

In this case, the NRMSE obtained for the entire measurement period is $2.42 \cdot 10^{-2}$. As can be noticed, despite the differences in high values of the SOC, the model proposed for the DT shows a rather appropriate behavior, capturing quite accurately the one of the corresponding component in the physical system. However, we can observe a recurrent mismatch between the real and predicted values when the SOC reaches its highest values (fully loaded battery), and during the battery discharge periods. These mismatches always happen in the same points of the SOC curve, and they could have a strong impact over the overall DT model.

B. Models based on neuronal networks

As discussed previously, model-driven approaches are not able to always fully capture the behavior of real systems, due to their non-predictable behavior. However, data-driven models can learn and reproduce these situations. Hence, for both modules, inverter and battery, three different types of neuronal networks (DNN, CNN and RNN) have been trained with the 50% of the real system measurements, 10% for evaluation during the training phase and 40% for testing the models.

TABLE II: NRMSE obtained for each model

Model	Discharge Power	SOC
Model-driven	2.79e-2	2.42e-2
Dense neuronal network	2.01e-2	1.50e-2
Convolutional neuronal network	1.42e-2	1e-2
Recurrent neuronal network	1.56e-2	6.3e-3

In order to illustrate the differences between the model-driven and data-driven approaches, Fig. 6b shows the predicted values obtained by the DNN. In particular, the DNN model takes as inputs the charge and discharge power values, as well as the previous SOC level. As can be observed, the results yielded by the DNN provide an almost perfect match with the real values. The data-driven approach does not only learns the normal and well known battery load and unload trends, but it also learns SOC behavior that does not correspond to a normal battery load calculation. First, opposed to the model-driven solution results, shown in Fig. 6a, the DNN avoids SOC saturation. Besides, it learns to follow the SOC unload process, even when it does not fully correspond to its expected behavior, predicted with the power charge/discharge. It is worth noting that the same behavior is also seen with the other neural networks used in this work.

In order to numerically measure the benefits of using data-driven solutions, Table II summarizes the NRMSE obtained with both the model-driven approach and the neural network models, to predict the battery discharge power and the SOC, to assess the performance of both the inverter and battery modules, respectively. From these values it can be concluded that the error decreases when applying neural networks on both modules. In the case of the battery discharge power, the convolutional neuronal network yields an error reduction close to a 50% compared to the model-based solution, slightly better than the other neural networks. However, in the case of the SOC, the recurrent neural network almost reduces the NRMSE an order of magnitude compared to the algorithmic-based solution. It is worth noting that, besides reducing the overall error, the data-driven models are able to avoid recurrent mismatches at particular points. For instance, the NRSME of the SOC prediction using the DNN is not as low as the one obtained with the RNN, but it was nevertheless able to correct the pathological behavior of the model-driven solution shown in Fig.6.

VI. Conclusions

This work presents the design and the first steps of the implementation of a DT for a hybrid renewable energy-hydrogen system, deployed in a house in the town of Novales (Cantabria, Spain).

The DT model features a number of functional modules that emulate the physical components of the real system. It uses, as external variables, the energy demand of the house, and the energy generated by the photovoltaic installation, predicting the value of the remaining variables.

We have discussed the initial implementation of the first modules that have been integrated into the DT, which represent the electrical components of the physical system. The obtained results indicate a good fit between the values measured in the real system and those obtained with the DT, with very low errors. On the other hand, it has been seen that the real components do not precisely behave as would be expected, but present some unexpected and unpredictable trends. To overcome this situations, data-driven solutions based on neural networks have been adopted. The results evince that the results yielded by the neural networks clearly outperform those obtained with the model-based option by reducing the overall error and avoiding pathological predictions.

The next steps we will tackle aim at extending the functionalities of the DT model, implementing the rest of the system components as independent software modules, which will be then connected to each other by means of the appropriate variables. Once the overall DT model has been validated, control policies will be proposed to optimize the system performance using both the system variables and external data sources (i.e. meteorological data). In addition, we will also foster an open access to relevant data, by developing the corresponding connections to various repositories and marketplaces. This would allow interested stakeholders to exploit the information generated by the proposed DT.

Acknowledgment

This work has been funded by the Spanish Government (Ministry of Science and Innovation) and the European Union (Next GenerationEU/RTRP) through the projects “Digital twin of a hybrid solar photovoltaic-hydro hybrid system for residential supply” (TED2021-129951B-C22) and ”Demonstration pilot of a solar-photovoltaic-hydrogen hybrid system for residential energy supply” (TED2021-129951B-C21), as well as by the Government of Cantabria through the project “Enabling Technologies for Digital Twins and their application in the chemical and communications sectors” (GDQuiC) of the program “Grants for research projects with high industrial potential of of technological agents of excellence for industrial competitiveness TCNIC”.

References

[1] “Paris agreement.” [Online]. Available: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en
[2] International Energy Agency, “CO_2 emissions from fuel combustion 2020 edition. Database documentation,” 2020. [Online]. Available: https://iea.blob.core.windows.net/assets/474cf91a-636b-4fde-b416-56064e0c7042/WorldCO2_Documentation.pdf
[3] H. Ritchie, M. Roser, and P. Rosado, “Co2 and greenhouse gas emissions,” *Our World in Data*, 2020, https://ourworldindata.org/co2-and-greenhouse-gas-emissions.
[4] European Council, “Fit for 55 - the eu’s plan for a green transition,” 2023. [Online]. Available: https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/
[5] ——, “Repowereu: Affordable, secure and sustainable energy for europe,” 2022. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/repowereu-affordable-secure-and-sustainable-energy-europe_en
[6] V. Maestre, A. Ortiz, and I. Ortiz, “Transition to a low-carbon building stock. techno-economic and spatial optimization of renewables-hydrogen strategies in spain,” *Journal of Energy Storage*, vol. 56, p. 105889, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352152X22018771
[7] M. Grieves, “Digital twin: manufacturing excellence through virtual factory replication,” *White paper*, vol. 1, no. 2014, pp. 1–7, 2014.
[8] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, “Digital twin in industry: State-of-the-art,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019.
[9] F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji, “Digital twin modeling,” *Journal of Manufacturing Systems*, vol. 64, pp. 372–389, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612522001108
[10] Q. Qi and F. Tao, “Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison,” *IEEE Access*, vol. 6, pp. 3585–3593, 2018.
[11] M. Liu, S. Fang, H. Dong, and C. Xu, “Review of digital twin about concepts, technologies, and industrial applications,” *Journal of Manufacturing Systems*, vol. 58, pp. 346–361, 2021, digital Twin towards Smart Manufacturing and Industry 4.0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612520301072
[12] A. Ebrahimi, “Challenges of developing a digital twin model of renewable energy generators,” in *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, 2019, pp. 1059–1066.
[13] S. Nguyen, M. Abdelhakim, and R. Kerestes, “Survey paper of digital twins and their integration into electric power systems,” in *2021 IEEE Power & Energy Society General Meeting (PESGM)*, 2021, pp. 01–05.
[14] S. Agostinelli, F. Cumo, G. Guidi, and C. Tomazzoli, “Cyber-physical systems improving building energy management: Digital twin and artificial intelligence,” *Energies*, vol. 14, no. 8, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/8/2338
[15] B. Gerard, E. Carrera, O. Bernard, and D. Lun1, “Smart design of green hydrogen facilities: A digital twin-driven approach,” *E3S Web of Conferences*, vol. 334, no. 2022, pp. 1–7, 2022.
[16] V. M. Maestre, A. Ortiz, and I. Ortiz, “Implementation and digitalization of a renewable hydrogen-based power system for social housing decarbonization,” *Chemical Engineering Transactions*, vol. 96, pp. 223–228, Nov. 2022. [Online]. Available: https://www.cetjournal.it/index.php/cet/article/view/CET2296038
[17] N. Hamed, A. Gaglione, A. Gluhak, O. Rana, and C. Perera, “Query interface for smart city internet of things data marketplaces: A case study,” *ACM Trans. Internet Things*, vol. 4, no. 3, sep 2023. [Online]. Available: https://doi.org/10.1145/3609336