35th European Modeling & Simulation Symposium
20th International Multidisciplinary Modeling & Simulation Multiconference

2724-0029 © 2023 The Authors.
doi: 10.46354/i3m.2023.emss.031

Snapshot Initialization: a New Method to Synchronize Dynamic Digital Twins Applied to Industrial Process

Louise da C. Ramos[1,2,*], Adrien Vassal[2], Gildas Besançon[1], Didier Paen[2], Jean-Christophe Blanchon[2], Stéphane Lagarde[3] and Thomas Piollet[3]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 11 Rue des Mathématiques, Saint-Martin-d'Hères, 38400, France
[2]CORYS, 44 Rue des Berges , Grenoble, 38000, France
[3]Framatome, 291 Route de l'Electrochimie, 38560 JARRIE, France

[*]Corresponding author. Email address: louise.da-costa-ramos@grenoble-inp.fr

Abstract

Digital twins are high-fidelity computational models vastly used to improve several industrial areas. In most applications, the digital twin requires synchronization to the real asset to properly represent the system. However, the expenses associated with the implementation of current synchronization methodologies for modeling vast industrial systems render such endeavors impracticable. In this work, a methodology to synchronize large-scale industrial digital twins through a snapshot file initialization approach is proposed. The method consists on creating an extensive data-base of snapshots files that represents the full parametric space of the system of interest. Each snapshot is linked to a period of measurements, which enables tracking the state of the system. A research approach is then applied to find the best snapshot file to initialize the model, synchronizing it to the real system based on measurements from the real asset. The methodology is demonstrated on an industrial process to separate zirconium and hafnium. The results show that, even if the real measurements are not precise and noisy, the approach can estimate a snapshot file to initialize the model with a error under 15 %.

Keywords: Digital Twin; Synchronization; Industrial Process

1. Introduction

Over the past decades, and still today, the industry is experiencing multiple technological breakthroughs such as Artificial Intelligence (AI), Internet Of Things (IoT), and Digital Twins (DT) (Brosinsky et al., 2019). The latter is the main focus of this article. Digital Twins are high-fidelity computational models that represent a real system in a complex or reduced form, being able to virtually copy or estimate the behavior of a real asset, enabling the evaluation, optimization, or prediction of a system with a reduced cost (Semeraro et al., 2021). Digital twins have been vastly used in industries, for instance, to design and test engines for airplanes (Popp and Schmidt, 2012; Enagi et al., 2017), to help the performance and automate industrial and manufacturing processes (Pérez Silva et al., 2020; Aguirre et al., 2020), to train industrial operators and reduce danger risk (Zoleykani et al., 2022), and to do optimal control and predict the behavior of different industrial processes (Vassal et al., 2022; Xia et al., 2021; Mounaam et al., 2020; Flood and Flood, 2022). Moreover, it has also been applied to support effective design of industrial production lines (Cimino et al., 2023), to improve human robot collaboration (Ramasubramanian et al., 2022) and to predict failure for certain types of tools (Luo et al., 2019).

The DT internal models can be based on physical equations (white box models), based on data (black box models), or hybrid (gray box models) (Brosinsky et al., 2019). In this article, a hybrid model is used. For most applications, the DT must correspond to the real asset, not only on the engineering and design properties but also on the state response of the system, requiring synchronization (Eckhart and Ekelhart, 2018). For instance, predictive control and optimization applications demand a DT synchronized to the real system, while operator training applications do not. Therefore, the research for methods to adjust models has increased over the past years, showing that most of the available synchronization techniques require primary knowledge from the system of interest, i.e. historical data from the system key variables (Boyd, 2001). For instance, parameter identification techniques have been vastly applied to calibrate DT for industrial fields, such as for power systems (Chen et al., 2022), or for industrial redundant manipulator robots (Urrea and Pascal, 2017). Furthermore, the synchronization of DT with industrial plants has been developed through techniques such as state observers design (Zipper and Diedrich, 2019), Artificial Intelligence, and Neural Networks (Akbarian et al., 2020; Brosinsky et al., 2019).

Remark 1: *It is worth stressing that in this work, synchronization refers to all methods that drive a DT to the same behavior as the one of the real asset, while calibration is an approach where parameters are automatically estimated to make the system behavior gets closer to the expected one.*

Nevertheless, to the best authors' knowledge, most of the available bibliography focuses on small-scale systems, processes, or machinery models (e.g. turbines, motors, boilers), which are simpler when compared to an entire plant, such as a refinery or a nuclear power plant. At a plant scale, the synchronization methodologies proposed for small systems might encounter several computational limitations, such as the computational cost and time to find a solution. One may also note that, for all of the available methodologies, the system of interest has a DT which equations are assumed to be known or available to be simulated coupled with the synchronization methods. However, for large-scaled nonlinear systems, which is the focus of this work, it is not feasible to presume that the modeling equations used on the DT are fully known, available, or even easy to simulate coupled with optimization algorithms due to their complexity.

Developing DT for large-scale industrial processes is a challenge itself, since such systems usually have several coupled components, low confidence and unknown parameters, are complex, highly non-linear, and have few and uncertain data available from the real asset (Schweiger et al., 2020). Nevertheless, the benefits that a DT brings to process engineering are such that the research on modeling solutions is an ongoing work. On the industrial process modeling, for instance, the physical-based model of a natural gas steam reforming plant (Salem et al., 2021), developed on ASPEN Hysys, has been used to generate process data to identify the optimal condition to maximize real production. (Salem et al., 2021) used a model at design condition, validated through steady state real data. In (Carrara et al., 2010) the steady state simulation of a steam reforming hydrogen production plant has been developed using ASPEN Plus commercial code (physical based model), coupled with field data from the plant, to theoretically study the energetic performance of the process. However, such works do not focus on dynamic simulations or how to validate the model behavior under real process charges.

Accordingly, for small scaled systems, it has been seen previously that parameter identification and synchronization techniques can be applied, nevertheless, for large-scaled DT such methods usually do not apply. Using the previously presented techniques, such as calibration or observers methods, demands coupling the DT equations with the observer, or extensively exploiting the parametric space for the unknown parameters, thus mapping the system behavior. However, most industrial high-fidelity DT are developed in software where equations are not available, and considering thousands of variables (physical and computational ones), such that these methods do not apply.

Inspired by such a problem, this work proposes an approach to synchronize large-scale nonlinear industrial DT through a *snapshot initialization*: the idea is to synchronize a DT by finding the best point to initialize it, such that the large-scale physics-based DT dynamics is similar to the real system. The approach is based on real measurements coming from the plant and already existing data from the DT. In the next sections of this paper, the proposed approach is detailed and tested for a real industrial plant. In section 2, the problem is stated, and section 3 presents the proposed methodology to address it. Section 4 then illustrates its application through an industrial example and results, while section 5 concludes the paper.

2. Problem Statement

The modeling of industrial processes through different physics-based simulation software (INDISS Plus, Modelica, Hysys, etc) is more and more applied in industrial fields (Schweiger et al., 2020). However, as previously discussed, there are several difficulties linked to the proper setup of a DT for a real industrial large scaled system. In this article, a DT for a zirconium purification process from Framatome (Barberis, 2016) is considered as an example. A model of the process developed with INDISS Plus software from CORYS is used. It is based on P&ID (Process & Instrumentation Diagram), control narrative, and equipment datasheest. The available model has been initially developed for operator training purposes, i.e., to teach operators how to handle the plant in the best possible way. Such a type of model presents hypothetical scenarios of the plant, and is not synchronized to the real behavior. For this work, the same model is used for engineering applications, such as state prediction, failure detection, etc. Consequently, the model must be synchronized to the plant through the plant's available information.

Figure 1 gives a scheme of the proposed synchronization solution, which is divided into three phases; (1) data extraction, (2) offline synchronization and (3) real and DT measurements comparison. In the first phase (1) information is extracted from the real plant, and a data treatment method is applied to make the real data homogeneous and clean. The second phase (2) is the synchronization of the system based on the available measurements. In this phase, an optimization methodology is applied to find the best initialization for the DT, such that the measurements from the plant and the model are as close as possible. The third phase (3), thus, is the validation of the DT based on the behavior of the measurement from the DT and the real asset. In this phase, the data from the real plant (blue line) and the DT (red dotted line) are compared, as exemplified in the plot.

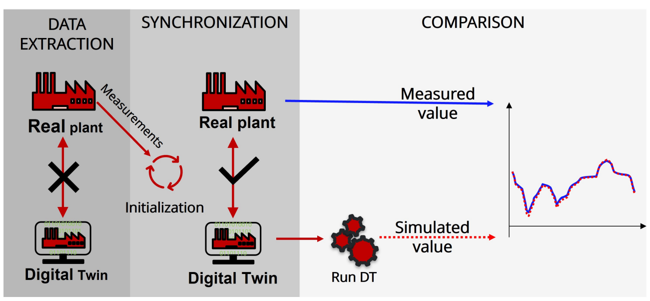

Figure 1. Problem scheme.

The main objective is to do an offline synchronization of the DT with the plant by finding the best state to initialize the DT, such that it leads to a behavior similar to the plant's behavior. On the modeling software, however, there are hundreds of internal variables, parameters, and states which are not fully known or accessible. Thus, a methodology based on snapshot files is proposed. Let us now take an overview of INDISS Plus and its main characteristics.

2.1. INDISS Plus

INDISS Plus is a high-fidelity process simulation platform that is based on fluid dynamics, rigorous thermodynamics and physical models (Thiabaud et al., 2011). The software model relies on first principles, such as the conservation of mass, momentum, and energy. INDISS Plus enables the modeling of any fluid industrial process from the main equipments to the safety system, covering the full project life cycle within the same platform. The process control system can be either modeled into INDISS Plus with dedicated libraries or imposed by a DCS (Distributed Control System) connected to it via OPC (Open Platform Communications). It includes processes, instruments, and safety systems, forming a virtual plant used for design, training, or other engineering applications (CORYS - INDISS Plus Team, 2022).

Even though INDISS Plus is a software for dynamic simulation and most of the physical parameters, variables, and boundary conditions are available, many internal states and parameters are hidden in the simulation, and can not be directly accessed. Such hidden values are, however, crucial for the simulation convergence. If a model from the software needs to be created but such values are not considered, the simulation will most likely diverge from the expected behavior. To overcome such limitations, the software gives the option to save information on a Snapshot File (SF). An SF is a large binary file that contains all information linked to the DT at the instant the snapshot is saved, i.e., the value of all variables, parameters, and internal states. Such a file can then be loaded into the software to restart the simulation at the same state as the one saved into the SF (CORYS - INDISS Plus Team, 2022).

Moreover, INDISS Plus enables the user to extract the value of model measurements over time. Also, the software outputs the name of the SF linked to each instant of simulation, such that, each time a new SF is saved during the simulation, the user can track when the snapshot was saved and the measurement values linked to each SF. It is worth stressing that the model output is what enables tracking the snapshot files during the simulation, without the need of opening the file and interpreting it. The interpretation of a snapshot file would demand a high cost concerning industrial large-scale processes, since hundreds of variables and parameters, and hundreds of hidden states would have to be decoded. Dynamically speaking, the cost of decoding SF for each saved instant is unfeasible.

This work proposes a methodology based on the snapshot files and the extracted output from INDISS Plus. In the following section, the industrial test case and the DT are presented.

2.2. Industrial Problem

The test case here is the industrial Framatome jarrie plant, which operates at the front end of the global process of converting zirconium (Zr) ore into nuclear and other industrial products (https://www.framatome.com/en/implantations/jarrie). The process consists in separating Zr and hafnium (Hf), which is a challenge since the two elements have similar chemical properties. To obtain pure Zr, a distillation process is applied, being composed of multiple steps which are presented in Figure 2 (Barberis, 2016).

The process is composed of four main parts: (1) the feed preparation (2) the distillation column (3) the liquid bottom collector drum (4) the reboiler. The process is a distillation that separates zirconium chloride and hafnium chloride which are initially mixed together. The distillation must use a solvent to dissolve the Zr and Hf chloride.

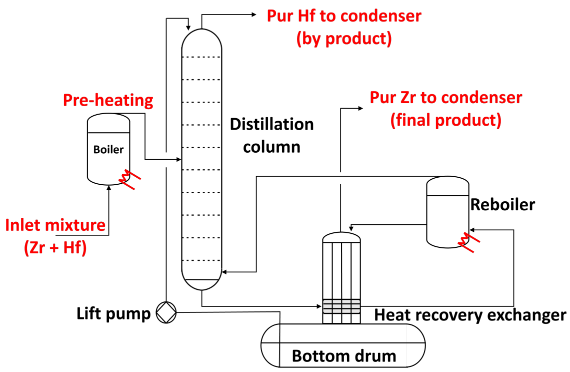

Figure 2. Framatome zirconium process scheme.

This solvent is a mixture of molten salts. The more volatile hafnium chloride is collected as a vapor phase at the top of the distillation column and routed to a condenser. The less volatile component, the zirconium chloride is extracted from the liquid coming from the reboiler and routed to a condenser. The liquid is pumped from the bottom drum to the top of the column. It is important to note that the schematic diagram presented herein provides only a broad overview of the system. The actual industrial plant is composed of numerous additional equipment, which have not been depicted in this diagram for intellectual property reasons.

This work uses a subsystem from the jarrie plant as an industrial example. It is composed of the third and fourth parts of the plant; the bottom of the distillation column, the bottom drum, the heat exchanger, and the reboiler. A model for a such subsystem is already available on INDISS Plus. For the test, a set of 25 measurements from the real system is given for a period of one week, where the process behavior varies. It is worth stressing that such a process has a slow dynamical response, and thus the measurements are given for a week.

3. Proposed Methodology

Our methodology consists in three main steps, as shown in Figure 3; (1) extract and pre-process data, (2) search measurements in the snapshot data library, and (3) returns the 'best' snapshot to synchronize the DT. In the following sections, each step is explained in more detail.

3.1. Plant Data Pre-Processing

The first step to enable the synchronization of the DT is to obtain a set of data from the plant. In this case, 25 measurements are available, but each one has a range of variation, a different sampling time step, noise and missing data. Then, it is necessary to clean and homogenize all the available measurements, i.e., complete the missing data points, set the same sampling time for all measurements, and clean the measurements outliers and noise.

Once the data is homogeneous, a Principal Component Analysis (PCA) is applied to understand the relation between the different variables, enabling the reduction of the model complexity by keeping only the variables that have different behavior. For instance, sometimes similar measures are taken close to each other, most likely repeating the same information. The use of PCA enables the understanding of the data and existing relations, and also gives the possibility to take out redundant information from the data set and to detect errors on the measurements. Such an approach must be applied when the system of interest is extensively big, and the number of available measurements is high. For the sake of brevity, the data processing is not further discussed in this paper.

3.2. Initialization Snapshot Selection

The goal here is, given a set of snapshots generated with INDISS Plus and a period of measurements from the real plant, to select the snapshot that better corresponds to the behavior of the measurement. There are two important steps here; (1) how to create the snapshot library and (2) how to find the best snapshot.

Remark 2: *In this methodology, the Snapshots are considered as two-component objects. Each snapshot is composed of an INDISS Plus snapshot file reference and a set of temporal measurements, as represented in Figure 3 at step 2. Such an approach enables to link each snapshot to a state of the system, characterized by its measurements. Again, the use of this particular snapshot definition enables the application of identification methodologies without knowing what is exactly inside a snapshot file, thus facilitating the research and decreasing the cost compared to other estimation methods.*

It is worth mentioning that the model used for the synchronization is assumed to have a behavior close to the steady state behavior of the plant. When the DT is not representative of the steady state behavior of the system, it is necessary to do a calibration on the model, but this is a research subject itself, and not the focus of this work.

3.2.1. *Creation of Snapshot Library*

First, it is necessary to build a database of snapshots. Two methodologies are proposed; one based on plant commands and the other one on sampling approach.

The plant command approach consists of using infor-

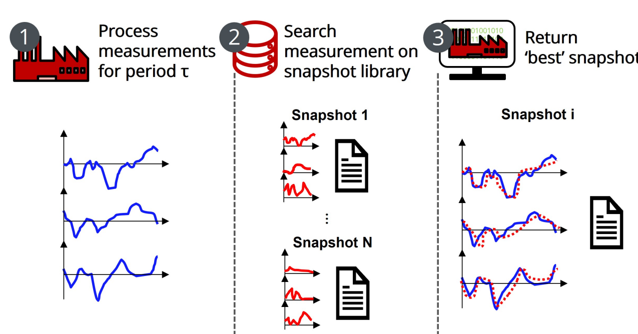

Figure 3. Method scheme.

mation given by the plant measurements to explore the regions of variation of the inputs, i.e. apply the same boundary conditions and commands given by the plant at the DT. For that, it is necessary to have extensive historical data of the plant, with variations, to impose the same values on the DT. For instance, if two command set-points are available in between the available measurements, those values can be used as an imposed input on the DT, as it is going to be demonstrated in Section 4.

The sampling-based approach is similar to that of the well know Design of Experiment (DoE), where the variables that impact the system output are known, and those are varied in a combined form, assuring that the parametric space is extensively explored. To apply such a method, it is necessary to know the limit variation range for each of the inputs, i.e., if there are n inputs $(x_0, x_1, \cdots, x_{n-1})$, it is necessary to know the range for each of the inputs; i.e. $((x_0^{min}, x_0^{max}), (x_1^{min}, x_1^{max}), \cdots, (x_{n-1}^{min}, x_{n-1}^{max}))$. For industrial systems, the knowledge of such values must come from specialists in the process of interest, such that it remains a physical and representative value.

Once the ranges are set, a statistical method, such as regularly spaced or Latin Hypercube Sampling (LHS) (Stein, 1987), is used to create a list of values for each variable. The number of created points depends on the sensitivity to the process of interest, which can be highly sensitive, thus needing a lot of points, or slightly sensitive, needing fewer points to map the behavior of the system. Once each input variable has a set of values, the DT is run according to those values , varying one input value at a time. This approach guarantees that all of the possible behaviors for the system of interest are explored and can be retrieved in the snapshot search. A further study of the impact of the SF distribution is a future work of interest.

3.2.2. *Initialization Snapshot Research Method*

Now that the snapshot database is created, the last step is to find the best snapshot to initialize the DT based on real measurements taken from the plant. To estimate the '*best*' SF from the database, a Root Mean Square (RMS) method is used, such that the smaller distance between the real and the DT temporal measurements is returned as the snapshot to initialize the model. Precisely, suppose that the n available plant measurements are $\Upsilon = \left[Y^0, Y^1, \cdots, Y^{n-1}\right]$, where each measurement Y^i, $i \in [0, n-1]$, is a vector with the temporal variation of the measurement, i.e., $Y^i = \left[y_0^i, y_1^i, \cdots, y_\tau^i\right]^T$, for y_t^i the measurement i for instants $t \in [0, \cdots, t_\tau]$. t is the indexation for each measurement instant, and τ is the number of samples.

The DT measurements have the same structure, being; $\hat{\Upsilon} = \left[\hat{Y}^0, \hat{Y}^1, \cdots, \hat{Y}^{n-1}\right]$, where each $\hat{Y}^i$ corresponds to a vector of measurements $\hat{Y}^i = \left[\hat{y}_0^i, \hat{y}_1^i, \cdots, \hat{Y}_\Gamma^i\right]^T$, with Γ the total number of snapshot samples. To find the best snapshot to initialize the model, an RMS between the real data and the snapshot set of measurements is done for all of the snapshots, and the minimization of the mean RMS value for all measures gives the best snapshot. Once the distance ($dist(Y^i, \hat{Y}^i)$) is computed (Eq. 1), the minimum value ($dist_{min}$) is identified.

$$\overline{dist}(Y^i, \hat{Y}^i) = \frac{1}{n} \sum_{i=0}^{n} RMS(Y^i, \hat{Y}^i) \tag{1}$$

Once the minimum value is obtained, the associated SF can be automatically selected from the database, and given to the DT to initialize the model.

Such a method is heavily based on the measurements that are available from the process of interest. To properly define and set the measurements from the plant, it is important to have special technical knowledge of the process. It is worth stressing that the choice of the period of sampling measurements (τ) has an impact on the results, and needs to be chosen based on the system's dynamic response time. In this work, a trial and error approach with industrial knowledge is applied to find the optimal sampling period, but further analysis of the impact of such a variable is to be developed.

4. Results

Let us now see the proposed methodology applied to the industrial plant subsystem, presented in Section 2.2. The first step is to clean and homogenize the data. Since the historical data is extensive, the commands and boundary conditions for the system can be identified, such that the same conditions are imposed on the DT. Once the commands and boundary conditions are defined, the DT is run under such inputs, to generate a set of snapshot files and data measurements, as explained in Section 3.2.1, at the plant commands approach. In this work, one SF was saved for each hour of simulation, leading to a total of 150 SF.

For the subsystem of interest, four variables are used as inputs: the temperature setpoint of the boiler, the boundary conditions of pressure and temperature at the distillation column bottom, and the pressure at the heat exchanger output. Then, INDISS plus is simulated with the imposed inputs and, for the sake of brevity, a subset of ten measurements is used for comparison and discussion here.

Figure 4 gives the comparison between the real measurements (blue curve) and the simulated ones (red curve) for the same imposed inputs. All the measurements are normalized with respect to the real measurement values, and the time step on the horizontal axis is given in seconds [s]. The comparison between the curves helps validate the model and create the snapshot library. Overall, the behavior of the DT is similar to the real expected one, and few discrepancies are seen. The difference between data might have two causes: first, the real measurements come from transmitters that could have some internal error, leading

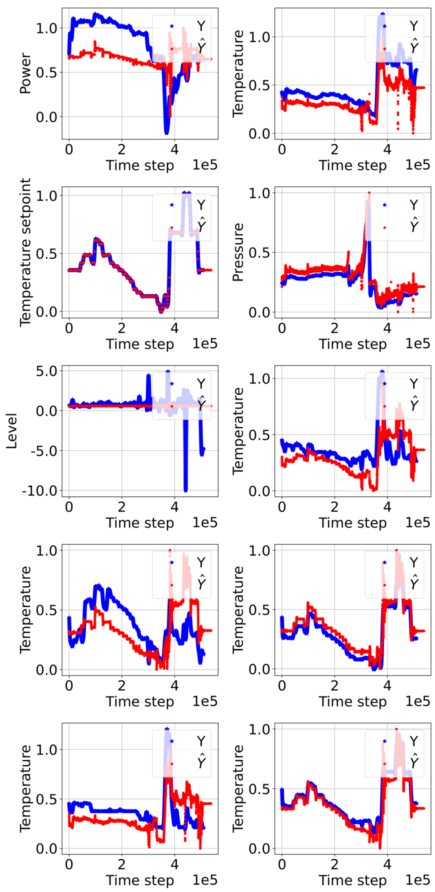

Figure 4. Comparison between the real plant (blue line) and for the DT (red line) measurements for the same imposed input command values.

to a false measurement value, linked to an offset on the mean measurement value. However, the dynamic behavior of each measurement can be trusted. Second, the DT model, in this case, represents a section of the whole system, that should interact with the full system, and the modeled subsystem only represent partially the physical impact coming from the external sections. Also, the model presents the ideal physical behavior, and it might underestimate some thermal and power losses, which cannot be controlled in reality. However, it is considered that the simulated data are close enough to the plant state to proceed with the snapshot research method described in 3.2.2.

In Figure 4, two of the four imposed inputs are depicted, representing a specific subset of the measured variables; the temperature setpoint of the boiler and the pressure of the bottom of the column, both given at the second row from the figure top. The corresponding real and DT curves for the two inputs are superposed, meaning that those are identical all along the simulation. The Power measurement (top row on the left) and the outlet temperature of the heat exchanger (fourth row from the top) are the measurements with remarkable differences in comparison to the real curve. The power given to heat the fluid on the boiler is higher on the real measurement, and this is due to the losses that the boiler has, which is ideal in the simulation, while it is impacted by many factors in reality. That being said, the mean behavior of the power measurement is similar. Similar remarks are made about the temperature, since the heat exchanger modeled presents ideal behavior, while the real one has other external factors that are not modeled. Nevertheless, the mean behavior for the temperature is similar. All the other measurements present low error in comparison to the real data, giving a mean and dynamic behavior close to reality.

Once the model and snapshot library are ready, a set of measurements from the real data is randomly chosen to test the methodology presented in Section 3.2.2. Here, the period for taking the plant measurements is of 45 minutes, meaning that, for a sampling time step of 60 seconds, 45 instants are taken for each measurement. The choice of such 45 minutes window of measurement is done by trial and error. Here, three validation cases are discussed, for different sets of measurements.

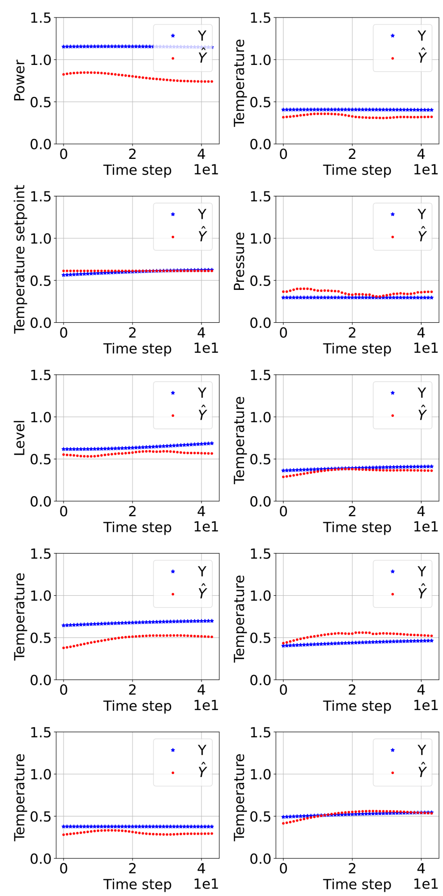

Figure 5. Comparison between plant (blue line) and the DT measurements found through the RMS minimization (red line) for the first validation case.

Generally, all the variables have a good match for the plant and the model measurement, less than 5 % error.

However, there is a remarkable difference between the plant and DT values for the power and the heat exchanger temperature. Such an error is linked to the transmitter measurement precision and to the ideal hypothesis that the model assumes, as previously discussed. Even though there are measurements that are noisy or not precise, the method is still able to find the best match of DT measurements, since it is computed over a period of measurements, thus accounting for the dynamic behavior of the measure, and because it averages the error between each of the measurements. The average error, for this first validation case, is 1 %. The summarized result of the method for the three validation cases is given in Table 1.

Table 1. Validation result summary

Case	Expected SF	Found SF	Minimum RMS	Expected vs Found error
1	01d03h00m	01d03h00m	1 %	0 %
2	02d22h00m	02d18h00m	7 %	0.26 %
3	05d10h00m	04d22h00m	13 %	0.32 %

Table 1 presents the summary of the results for three validation cases. It gives the validation case in the first column, the reference to the expected snapshot in the second column, the predicted snapshot reference in the third column, the minimum error found on the RMS (Eq. 1) on the fourth column, and the error between the expected and the found snapshot behaviors on the last column. The expected and found snapshot references correspond to the instant where the snapshot is saved during the simulation, in days (d), hours (h), and minutes (m), such that the reference looks like: ##d##h##m. For instance, the first validation case expected snapshot corresponds to 01d03h00m of simulation, i.e. taken after $0.97e5$ seconds of simulation. The expected SF was taken based on the superposition between the plant and the DT curves; since the same inputs (commands and boundary conditions) are imposed on both systems, each instant of simulation is assumed to correspond to the same instant of the plant data, thus being linked to a saved SF from the simulation. The use of such snapshot reference enables to superficially validate if the predicted SF from the initialization snapshot research method is close to the expected one. To quantify the error between the expected and predicted SF, both snapshots were loaded and run on INDISS Plus, and their corresponding measurements were compared by computing the RMS relative error, shown in the fifth column of table 1.

The predicted snapshots for the two first validation cases are remarkably close to the expected one, and the minimum RMS error is lower than 10 %. In figure 5, one may see the comparison between the measurements from the plant (Y) and from the DT ($\hat{Y}$) of the best match for the first validation case. The difference between most of the measurements is very little, presenting a good fit with a relative error of 1 %. Nonetheless, for the Power and the Temperature of the heat exchanger (fourth row from the top), a gap is seen between the two curves, which is due to the impact of external factors that are not taken into account on the ideal model used on the DT. Similar behavior was seen for the second validation case. The third validation case, however, presented a slightly higher error of 13 % for the best distance match. Such an error is due to the plant measurements which have a non-physical behavior, presenting a value that does not represent the real dynamic of the system, but it is not shown here for the sake of brevity. Such a problem, might be linked to the lack of precision of a measurement transmitter, and even if the best match error is higher, the comparison between the found and expected SF presents an error of 0.32 %.

5. Conclusion

A methodology to synchronize large scale industrial DT through a snapshot file approch has been proposed. The method consists in using an existing DT to generate a database of snapshot files that are linked to a set of measurements. The approach assumes that the DT software enables the recording of the model variables and parameters through snapshot files, which can be loaded into the software to impose a certain state over the DT. A RMS is used to estimate the best snapshot file to initialize the DT such that it is synchronized to the real asset behavior.

When applied to real data from the industrial plant subsystem from Framatome, the approach presents remarkable results, even if real data are not precise and contains noise. Here, the plant command approach is used to generate the snapshot database, since the amount of data from the plant is extensive and variable. However, when the history of the plant behavior is small or lacks variability, the sampling-based approach can be used, and the study of such a method is a perspective of this work. Furthermore, it has been seen that the period of measurement coming from the plant impacts the result, and thus more attention should be given to the choice of the measurement time window. Here, a period of 45 minutes of sampling seems optimal, and such a relatively long time is linked to the slow dynamics of the studied system.

The proposed approach has several advantages in comparison to existing ones. For instance, the method is not linked to the DT model or equations, since it is based on SF, and the better distributed and populated the snapshot database is, the better the precision of the results. Also, no training is necessary, different from machine learning training algorithms, since the approach is based only on the minimization of the distance between the plant and the DT measures. It has also been seen that the more the SF on the database, the more the research method takes time to compute. Thus, the study of other approaches for highly populated snapshot database is a perspective of this work. Moreover, the presented method is now being tested on the full jarrie plant, giving a highly complex system and DT example.

References

Aguirre, A., Huerga, D., and Albajez, I. (2020). Simulation analysis of a new semi-continuous chemical facility with automated production lines: an industrial case study. In *Proceedings of the 32rd European Modeling and Simulation Symposium (EMSS 2021)*, pages 186–192.

Akbarian, F., Fitzgerald, E., and Kihl, M. (2020). Synchronization in digital twins for industrial control systems. *ArXiv*, abs/2006.03447.

Barberis, P. (2016). Métallurgie du zirconium et du hafnium. *Techniques de l'ingénieur Élaboration et recyclage des métaux*, base documentaire : TIP554WEB.(ref. article : m2360).

Boyd, D. W. (2001). *CHAPTER 2 - Systems Modeling Principles*. Academic Press.

Brosinsky, C., Song, X., and Westermann, D. (2019). Digital twin - concept of a continuously adaptive power system mirror. In *International ETG-Congress 2019; ETG Symposium*, pages 1–6.

Carrara, A., Perdichizzi, A., and Barigozzi, G. (2010). Simulation of an hydrogen production steam reforming industrial plant for energetic performance prediction. *International Journal of Hydrogen Energy*, 35:3499–3508.

Chen, X., Zhao, T., and Chen, Q. (2022). An online parameter identification and real-time optimization platform for thermal systems and its application. *Applied Energy*, 307:118199.

Cimino, A., Gnoni, M. G., Longo, F., and La Rosa, A. (2023). Digital twin (dt) based methodology to support effective design of industrial production lines. *Procedia Computer Science*, 217:1896–1907. 4th International Conference on Industry 4.0 and Smart Manufacturing.

CORYS - INDISS Plus Team (2022). *INDISS Plus Manual*. CORYS.

Eckhart, M. and Ekelhart, A. (2018). A specification-based state replication approach for digital twins. In *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy*, page 36–47.

Enagi, I., Al-attab, K., and Zainal, Z. (2017). Combustion chamber design and performance for micro gas turbine application. *Fuel Processing Technology*, 166:258–268.

Flood, I. and Flood, P. D. L. (2022). Intelligent control of construction manufacturing processes using deep reinforcement learning. In *Proceedings of the 12th International Conference on Simulation and Modeling Methodologies, Technologies and Applications - SIMULTECH,*, pages 112–122.

Luo, W., Hu, T., Zhang, C., and Wei, Y. (2019). Digital twin for cnc machine tool: modeling and using strategy. *Journal of Ambient Intelligence and Humanized Computing*, 10:1129–1140.

Mounaam, A., Harmen, Y., Chhiti, Y., Souissi, A., Salouhi, M., and el Khouakhi, M. (2020). Unisim-design simulation and analysis of a sulphuric acid manufacturing plant with double absorption process. In *Proceedings of 8th International Conference on Simulation and Modeling Methodologies, Technologies and Applications - SIMULTECH*, pages 91–100.

Popp, M. and Schmidt, G. (2012). Rocket engine combustion chamber design concepts for enhanced life. In *32nd Joint Propulsion Conference and Exhibit*.

Pérez Silva, A., Rodríguez, V., Flores, I., and Soler, F. (2020). Simulation of elements of industry 4.0 applied to the production processes of a company dedicated to the manufacture of plastic products. In *Proceedings of the 32rd European Modeling and Simulation Symposium (EMSS 2021)*, pages 84–92.

Ramasubramanian, A. K., Mathew, R., Kelly, M., Hargaden, V., and Papakostas, N. (2022). Digital twin for human robot collaboration in manufacturing: Review and outlook. *Applied Sciences*, 12(10).

Salem, M., Shoaib, A., and Ibrahim, A. (2021). Simulation of a natural gas steam reforming plant for hydrogen production optimization. *Chemical Engineering and Technology*.

Schweiger, G., Nilsson, H., Schoeggl, J., Birk, W., and Posch, A. (2020). Modeling and simulation of large-scale systems: A systematic comparison of modeling paradigms. *Applied Mathematics and Computation*, 365:124713.

Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. *Computers in Industry*, 130:103469.

Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151.

Thiabaud, P., Gros, F., Tourdjman, J., and Guaraldo, M. (2011). Benefits of Life Cycle Dynamic Simulation for FPSO Projects and Offshore Developments. volume All Days of *SPE Brasil Offshore*. SPE-143532-MS.

Urrea, C. and Pascal, J. (2017). Parameter identification methods for real redundant manipulators. *Journal of Applied Research and Technology*, 15:320–331.

Vassal, A., Ghribi, A., Millet, F., Bonne, F., Bonnay, P., and Bernaudin, P.-E. (2022). Spiral2 cryomodule models: A gateway to process control and machine learning. *Frontiers in Physics*, 10.

Xia, K., Sacco, C., Kirkpatrick, M., Saidy, C., Nguyen, L., Kircaliali, A., and Harik, R. (2021). A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. *Journal of Manufacturing Systems*, 58:210–230.

Zipper, H. and Diedrich, C. (2019). Synchronization of industrial plant and digital twin. In *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1678–1681.

Zoleykani, M., Abbasianjahromi, H., Banihashemi, S., Tabadkani, A., and Haji Rasouli, A. (2022). Extended reality (xr) technologies in the construction safety: systematic review and analysis. *Construction Innovation*, pages 413–417.