Data Architecture for Digital Twin of Commercial Greenhouse Production

Howard, Daniel Anthony; Ma, Zheng; Aaslyng, Jesper Peter Mazanti; Jørgensen, Bo Nørregaard

Published in:
2020 RIVF International Conference on Computing and Communication Technologies (RIVF)

DOI:
10.1109/RIVF48685.2020.9140726

Publication date:
2020

Citation for published version (APA):

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying this open access version.

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
Data Architecture for Digital Twin of Commercial Greenhouse Production

Daniel Anthony Howard  
Center for Energy Informatics  
The Maersk Mc-Kinney Moller Institute  
Odense, Denmark  
danho@mmmi.sdu.dk

Jesper Mazanti Aaslyng  
AgroTech  
Danish Technological Institute  
Taastrup, Denmark  
jeaa@teknologisk.dk

Zheng Ma  
Health Informatics and Technology  
The Maersk Mc-Kinney Moller Institute  
Odense, Denmark  
zma@mmmi.sdu.dk

Bo Nørregaard Jørgensen  
Center for Energy Informatics  
The Maersk Mc-Kinney Moller Institute  
Odense, Denmark  
bnj@mmmi.sdu.dk

Abstract—There is an increasing demand for industry specific solutions for optimizing production processes with the transitions towards Industry 4.0. The commercial greenhouse sector relies heavily on optimal use of energy with multiple new concepts introduced in recent years e.g. vertical farming and urban agriculture. Digital twins allow utilizing Internet of Things and big data to simulate the alternative operation strategies without compromising current operation. This paper aims to present the development of a digital twin of the commercial greenhouse production process as a part of the recently launched EUDP funded project Greenhouse Industry 4.0 in Denmark. This digital twin allows using big data and the Internet of Things to optimise the greenhouse production process and communicate with other digital twins representing essential areas in the greenhouse (climate and energy). This digital twin can estimate future states of the greenhouse by using past and real-time data inputs from databases, sensors, and spot markets. This paper also introduces a Smart Industry Architecture Model Framework for the discussion of the required data architecture of the digital twin for the greenhouse production flow which ensures a correct data architecture for the data exchange across all entities in the system.

Keywords—industry 4.0, commercial greenhouse, digital twin, data architecture, production process

I. INTRODUCTION

Currently, modern commercial greenhouses in Denmark and Western Europe are equipped with advanced facilities, e.g. movable screens for insulation during night and shading during day, supplemental lighting often fitted with energy-efficient light-emitting diodes (LED), screens for ventilation, pipe heating systems and CO2 supply connected to an advanced climate-control computer that enable the commercial greenhouse growers to optimize the climate continuously to reach predefined production goals (in time, quantity and quality) with the least possible energy expenses.

Therefore, during the last 20 years, several projects have focused on different aspects of energy optimization in the greenhouse production. The focuses have either been on implementing new technologies, such as infrastructures (glass and screen types), light-emitting diodes or other types of advanced equipment, or on developing various decision support systems and ICT systems for the improved climate control in order to balance the optimal plant photosynthesis and growth, or transpiration, with the use and cost of energy [1-5].

The innovative approach and functionality of the new project Greenhouse Industry 4.0 (GHI4.0) is a four-years project from September 2019 to October 2023 and is funded by EUDP (the national Energy Technology Development and Demonstration Program, Denmark), the project is a collaboration between the following parties: Danish Technological Institute, Center for Energy Informatics at University of Southern Denmark, Department of Food Science and AgroEcology at Aarhus University, Danish Horticulture, NB Data, Danish Cleantech Hub, Energy Denmark, the nurseries: Knud Jepsen, ByGrowers, and Hjortebjerg.

The project GHI4.0 enables the commercial greenhouse growers to become sustainable and flexible energy users (electricity, district heating, gas). Furthermore, the scope of the project is also to show both the possibilities and the challenges in including complex energy demanding and energy-producing industry with high flexibility to the smart grid. The case story will not only lift the Danish commercial greenhouse industry but also lead the way for other industries and businesses to take advantage of the knowledge and technology developed in the project.

There is an increasing penetration of intermittent renewable energy production in the Danish energy system [6]. To accompany the renewable transition, there has been an increasing focus on intelligent solutions that effectively exploit and adapt to the energy system. In Denmark, the greenhouse industry alone uses 4.2 PJ annually [7]. The greenhouse industry, as a labour and energy-intensive sector, to stay competitive against low-cost regions, need to apply a vertical integration and optimisation of the production processes with the Industry 4.0 Digital Twin (DT) concept. The project will enable the greenhouses to become increasingly energy-efficient while maintaining sustainable production. Furthermore, a DT can incorporate essential parameters to optimise the overall production plan of the individual greenhouse hence providing a complete approach for the commercial growers. The project thereby seeks to enable Danish greenhouse growers to transition from best-practice to early adopters of next-practice.
The Danish horticultural industry must continuously improve energy efficiency and production throughput and productivity, without compromising product quality or sustainability, to stay competitive in the global market. This is achieved by using a DT that combines modelling, artificial intelligence (AI) and Big Data analytics with Internet-of-Things (IoT) and traditional sensor data from the production and cloud-based enterprise data to predict how the physical twin will perform under different operation conditions. Deploying the DT concept support co-optimization of the production schedule, energy consumption, and labour cost, by considering influential factors including production deadlines, quality grading, heating, artificial lighting, energy prices (gas and electricity), and weather forecasts.

There are seven work packages in the project GHI4.0, and includes five research work packages:

- WP3: Digital Twin of greenhouse climate compartments.
- WP4: Digital Twin of greenhouse production flow.
- WP5: Digital Twin of greenhouse energy system.
- WP6: Greenhouse production Dashboard.

This paper presents the motivations and expected outcomes in regard to the development of a DT for the greenhouse production process (WP4).

The paper is organized as follows: firstly, the importance and potentials of developing DTs for greenhouse production is introduced with the literature review. Then, a in-depth description of DT development and communication is outlined. Thirdly, the methodology is described with the focus on the essential data and data architecture required for the DT operation followed by the discussion and conclusion.

II. IMPORTANCE OF DIGITAL TWIN FOR GREENHOUSE PRODUCTION

Currently, there is a desire to go towards a smarter energy system, allowing for better management of energy consumption. Research has been focused primarily on automation and management of buildings. However, in recent years there has been a growing interest in the development of DT models [8]. Gartner has appointed DTs as one of the top 10 strategic technology trends of 2017, 2018, and 2019 [9]. DTs have been proved usefully in optimising systems based on collected data and how a system responds to changes in setting. This can help stakeholders realise unexploited potentials and improve operational security [10].

Due to the complexity of production systems and the high energy demand of producing plants in the northern hemisphere, modern greenhouses are still energy-intensive. Therefore, the commercial greenhouse growers have to achieve high production levels with high monetary output to secure return on investment. This is, however, often a complex task. Decisions on finding the optimum of production and resource input for a perfect cultivation climate is a trade-off between many different needs and factors. The adjustment of one climate variable, e.g. temperature, affects most other climate variables, which further affects plant growth and development and energy costs.

With the use of DT simulation, Schultz et al. found a potential for a significant reduction of energy demand in manufacturing processes by 27 % [11]. Furthermore, with the increasing implementation of decentralised intermittent renewable energy sources increasing the flexibility on the supply and demand side becomes imperative for grid stability. It was shown by Beier et al. [12] that a manufacturing systems energy demand could be regulated based of the intermittent supply of renewable energy sources by implying a control strategy based on simulations performed in AnyLogic, and even showed improvements within the manufacturing process. Hence, there is a potential for DTs to help facilitate the transition towards a sustainable energy system.

The advantage and use of DTs were simply described by Bernard Marr in a Forbes article "pairing of the virtual and physical worlds allows analysis of data and monitoring of systems to head off problems before they even occur, prevent downtime, develop new opportunities and even plan
for the future by using simulations.” [13]. This builds on leveraging the rising amount of available sensory data for developing virtual models of physical systems. The development of a generic DT framework would enable factories to control production processes more accurately and respond to changes in market demand. Furthermore, it could help reduce the cost of inefficient production and improve the economic benefits and sustainable development capabilities of enterprises [9].

The GHI4.0 project focuses on smart production facilities (industry 4.0), energy systems and management, due to this consumer category having a large potential for flexibility [14]. The greenhouse system is shown in Fig. 1, the figure illustrates how a set of domain-specific DTs will be developed in the project. This paper mainly introduces the development of the DT for greenhouse production flow (WP4).

A. Importance of Digital Twin for Greenhouse Production Process (WP4)

The DT of the greenhouse production process presents a new approach to integrating greenhouse production flow with an energy management system. Greenhouse production has a high degree of flexibility in using different energy sources. The production (the plants) also tolerate a high degree of dynamics in the climate. Combining the flexibility and production tolerance with the right decisions, the commercial greenhouse growers have the possibility to be an active player in the future energy system.

This WP aims to create an artificial intelligence (AI) based simulation model of the greenhouse production flow for investigating the effects of co-optimizing production schedule, plant growth, energy consumption, and cost, by considering influential factors including production deadlines, quality assessment, (district) heating demand, gas and electricity prices, and weather forecasts. WP4 will incorporate all parts of the production process in the DT which will allow the commercial greenhouse growers to obtain an overview of the current operation of the production flow. The validation of the DT is established once the current production process is approximated at a sufficient level. Using the WP4 DT, optimisation possibilities within the production flow can be examined that will allow testing alternative strategies with minimal risk to the actual operation.

The importance of incorporating the production process DT in the modelling process is especially important in the aspect of production time. The DT of the production process will have inputs as the delivery of new seedlings which are set at specific times and outputs set as the customer delivery deadline, everything from the initial delivery to the deadline is controlled in the production process. The seedlings will enter an almost conveyor type production process until they are fully grown plants ready for delivery at a specific deadline.

During the growing stages, a set of parameters will heavily influence the overall production time and the final delivery time. This depends on the type of plant, the amount of sunlight and artificial lighting together with humidity, temperature etc. controlled by the climate system. However, as previously stated the plants are in a conveyor type process which means that accounting for the available space in each part of the system is important to ensure that no bottlenecks become present throughout the system and that the overall throughput is upheld.

Each batch of incoming plants will act as an individual agent which will have a set of prescribed parameters e.g. time for arrival, deadline, time to deadline, plant type, quality, time in the system etc. As the batches (agents) move through the system the parameters will be updated accordingly and can provide information to the DT for prioritisation of batches with e.g. upcoming deadlines. As the batches will consist of an array of different plant species each type will have varying susceptibility to flexibility potentials i.e. some plants have stricter requirements for lighting, humidity, and temperature. Hence all batches are not equal in terms of production process flexibility potential and the DT will enable pinpointing the optimisation potentials while considering all the previously discussed parameters so deadlines are upheld and no bottlenecks become present.

B. Importance of coupling with other Digital Twins (WP3 & WP5)

Together with the greenhouse production process DT, two other DTs will be developed representing the greenhouse climate system and the energy system in the greenhouse, shown in Fig. 1. It is imperative to couple the three DTs together to achieve insight into the consequences associated with a decision in one of the domains. Communication will be established between the energy system DT and the climate system DT and then from the climate system to the production process DT. The information exchange flow between the greenhouse DTs with the production process DT shown in Fig. 2.

![Diagram showing information exchange between greenhouse DTs](image)

As previously mentioned the production process will be modelled in batches prescribed by a set of parameters. To obtain the parameters for each batch, it is important for the production process to communicate with the climate system as this is the DT that will calculate and analyse the batches in regards to plant specifications. This will result in the overall system effectively considering prices on energy combined with product specific variables and production process restraints. E.g. choosing to increase the light intensity in some hours to speed up the production process may give a short-term monetary saving due to a low electricity price. However, it may subsequently create bottlenecks or the entire batches ready too early resulting in an overall monetary loss because the production schedule and deadline are not considered. Hence it is important to incorporate all the models into the Common Information Model (WP2) that will ensure that all parts of the greenhouse are considered for optimisation.

III. METHODOLOGY

The purpose of the DT development of a greenhouse production process is to document the effects of changing production deadlines, quality grading requirements, district heating demand, gas and electricity prices, and weather forecasts. Using a multi-agent simulation, the potentials of
energy efficiency and demand response participation can be investigated by simulating different scenarios of production flow with optimization.

The development of the DT will be done in collaboration with the Danish Technical Institute (DTI), NB Data, and three commercial greenhouse growers. The AI simulation platform AnyLogic will be used to create a generic multi-agent simulation model of the greenhouse production flow that includes all production steps from the delivery of initial seedlings until the shipment of the final plants. AnyLogic was chosen as the simulation software as it currently is the only simulation software supporting multimethod simulation modelling enabling the user to use discrete event, agent based, and system dynamics simultaneously [15].

To advance the DT parameters, A case study with interviews will be developed to gain insight into the behavioural patterns of the production line operators as well as the current operation strategies and practices. It will be combined with an array of parameters to accurately depict the production flow. Initially, the goal will be to establish a working DT of the existing system to ensure a correct working baseline, once the current system is modelled with sufficient accuracy, the DT can be used for setting up scenarios and testing alternative operating strategies. For instance, the greenhouse layout of a commercial greenhouse grower is shown in Fig. 3.

![Greenhouse layout](image)

Fig 3. Greenhouse layout [16]

To construct a DT for the greenhouse production process at a sufficient level, essential data about the entire greenhouse production is needed. The data includes the production schedules, number of cuttings arriving per batch, the arrival interval, and length of conveyor belts, deadlines, quality grading, weather forecasts, etc. Integration with the company ERP system could be a possibility for easily obtaining information about procurement, production, sales, and distribution. The DT relies heavily on available data and a continues flow of real-time in order to ensure continuous adaption and learning from the environment.

### A. Data and Data Architecture

A DT combines simulation, AI and big data analytics with IoT (Internet-of-Things) and traditional sensor data from the production and cloud-based enterprise data to predict how the physical twin will perform under different operation conditions. For instance, there are multiple types of data for the greenhouse production process (shown in TABLE I). Wang et al. [17] state that the shift towards Industry 4.0 entails a vast amount of sensory data produced by the sensors attached to the examined devices. In this new data-driven environment, the processing, storage, and utilization of the manufacturing data become increasingly important which makes the traditional one-way data architecture obsolete as it requires multi-scenario application.

<table>
<thead>
<tr>
<th>Day</th>
<th>Stage</th>
<th>Description</th>
<th>Lighting and temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cuttings</td>
<td>The cuttings are placed in pots manually on a conveyer belt</td>
<td>–</td>
</tr>
<tr>
<td>1-9</td>
<td>Rooting</td>
<td>The cuttings are sprayed with anti-fungi and grow small roots</td>
<td>22 – 24 °C 40 micromoles lighting: LED 17 hours light, 7 hours dark</td>
</tr>
<tr>
<td>10 – 28</td>
<td>Propagation</td>
<td>The cuttings with small roots are moved to a new area where they are grown from cuttings to small plants</td>
<td>19 °C (day) 21 °C (night) 100 + 25 micromoles: Combined lighting: HPS and LED 17 hours light, 7 hours dark</td>
</tr>
<tr>
<td>29 – 100</td>
<td>Short-day area</td>
<td>The plants are grown to the required height and size</td>
<td>19 °C (day) 21 °C (night) 100 micromoles: HPS 10 hours light, 14 hours dark</td>
</tr>
<tr>
<td>100</td>
<td>Sales</td>
<td>The plants are sold</td>
<td>–</td>
</tr>
</tbody>
</table>

Saqlain et al. [18] develop a Industrial Data Management System (IDMS) with five layers, and IDMS presents a system for real-time and scalable data collection, transmission, processing, and storage in industrial systems. A similar four-layered architecture for IoT (Internet-of-Things) devices was presented by Xu et al. [19] in a survey that systematically reviewed and summarized the current state-of-art in industry IoT. Combining the SGAM approach with the developed IoT models i.e. Industrial Data Management System (IDMS) and four-layered architecture for IoT provide a useful tool for obtaining data information by applying an IoT based framework with a domain-oriented framework to understand the required data flow required in different parts of the industry process.

To purposefully understand the required and exchanged data throughout the project the Smart Industry Architecture Model Framework (SIAM) is proposed (shown in Fig. 4). The SIAM framework incorporates the SGAM (Smart Grid Architecture Model) framework developed by the Smart Grid Coordination Group. SGAM framework includees
information flows across the primary domains in Smart Grids [20].

The SIAM framework consists of a five-layer digitalization application which are connected to the domains of the supply chain (including logistics, production and customer premises) and a set of different zones (including process, operations, enterprise, market). WP 4 - digital twin of greenhouse production process is relevant to the ‘production’ in the domain of supply chain. TABLE II shows the data and information exchange associated with the different zones and applications in regards to the production (Note: only relevant fields are described and fields with no immediate correlation are left blank).

IV. DISCUSSION

A generic DT of production processes requires the industry readiness for the implementation of smart solutions. The physical sensors and big data fulfill the requirement that can support the DT to perform optimisation of the production process.

A. Expected outcome and activities

The expected outcome of the WP4 will be a functional DT for the greenhouse production process that can be incorporated through a Common Information Model to deliver scenarios and consequences for the operation of the production process through co-optimisation with the other DTs. Activities will include a visit to the participating greenhouses in order to investigate the production process and the operational strategies. Once all information has been collected and analyzed based on the SIAM framework (shown in Fig. 4), the DT will be developed in AnyLogic. The DT of greenhouse production flow is expected to deliver the following:

- Analyses of production throughput speed.
- Calculation of expected production deadline and quality with co-optimization effects.
- Calculation of expected production deadline and quality with demand response participation.
- Analyses of energy efficiency and demand response potentials.
- Calculation of expected production deadlines and quality with demand response participation.

### TABLE II. SIAM FRAMEWORK OF GREENHOUSE PRODUCTION FLOW

<table>
<thead>
<tr>
<th>Application</th>
<th>Process</th>
<th>Operation</th>
<th>Enterprise</th>
<th>Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Layer</td>
<td></td>
<td>Possibility of partaking in the demand response program</td>
<td>Will contribute to the proper model restraints to prevent production bottlenecks</td>
<td>Will contribute to proper model restraints for the monetary value of workforce operation with respect to time</td>
</tr>
<tr>
<td>Function Layer</td>
<td>Will contribute to the proper model restraints to prevent production bottlenecks</td>
<td>Will contribute to the proper model restraints for quality of the batches</td>
<td>Will contribute to the proper model restraints for the monetary value of workforce operation with respect to time</td>
<td>Will contribute to proper model restraints to ensure timely delivery and throughput of the system</td>
</tr>
<tr>
<td>Information Layer</td>
<td>Space availability in the given areas.</td>
<td>Operating times and current states of batches. Input parameters from the climate system</td>
<td>Information on available workforce, wages and schedule.</td>
<td>Sales and shipment details for each batch.</td>
</tr>
<tr>
<td>Communication Layer</td>
<td>Sensor data transmitted through chosen communication technology. Supported by inputs from the climate system DT</td>
<td></td>
<td>Acquired through the company ERP system.</td>
<td>Acquired through the company ERP/CRM system.</td>
</tr>
<tr>
<td>Component Layer</td>
<td>Physical outline of the components in the production process. Placement and dimensions are essential parameters.</td>
<td>Placement of sensors in the system.</td>
<td></td>
<td>Workforce movement</td>
</tr>
</tbody>
</table>

It also applies to the readiness of commercial greenhouses that implement smart solutions in their production to ensure the transition from perceived best-practice to next-practice. In the famous words of Peter Drucker “What gets measured, gets managed” holds especially true for DT optimisation as it relies heavily on the available measurements throughout the system. The detail of DTs depends on the availability of the relevant data. Some data is not necessary to be included or into too detail, i.e. it is not necessary to know the electricity usage down to second if the electricity price is only available on an hourly basis.

Fig. 4. Smart Industry Architecture Model
The DT will use multi-agent simulation of the production flow to analyze the effect of changing production parameters for the DT of the greenhouse climate system. Furthermore, it will be used to simulate production and to optimize it with respect to both production, quality and use of energy.

B. Expected contribution

As the transformation to Industry 4.0 becomes greater, the need for industry-specific applications becomes increasingly important to fully utilise and benefit from the available data. The research conducted as a part of this project is expected to give a more holistic approach for optimisation of the greenhouse industry that will allow incorporating all parts of their value chain and a better decision-process. Furthermore, the research will focus on the development of a generic DT framework for the production processes implementation which will ease the development of future DT models for industry-specific purposes. At the core, this project represents an important addition to the already existing knowledge concerning DTs and Industry 4.0 as the project encompasses all parts of the system. Currently, the majority of research has focused heavily on the development of standalone DTs that allow optimisation in a given domain. However, as many of the different parameters in a industrial production are undoubtedly linked, it is imperative for industries to gain an overview of the mechanisms that link and influence the entirety of their production process. Hence this research will contribute to a more holistic understanding of DT coupling and how this can facilitate the shaping and improvement of the future industries.

V. CONCLUSION

The paper introduces a work package of the 4-years project- Greenhouse Industry 4.0 (from September 2019 to September 2023). This work package aims to develop the digital twin of greenhouse production flow. The expected result is hence a generic DT framework for modelling, controlling and optimizing cyber-physical systems of the production process. This will accumulate to a DT of the greenhouse industry 4.0 that will be installed and demonstrated in three commercial greenhouses in Denmark. Furthermore, the potentials for large-scale implementation will be examined.

Through the use of the SIAM (Smart Industry Architecture Model) framework and literature on data architecture and the Internet of Things, the required exchange of data and information in the different layers of the model could be examined as a foundation for the digital twin development of the greenhouse production process.

ACKNOWLEDGEMENT

This work is part of the project- Greenhouse Industry 4.0 funded by EUDP, Denmark.

REFERENCES


