

# Preliminary Phases of Implementing a Digital Twin Solution in the Food Industry: a Case Study

Marco Menegon, Alberto Di Loreto, Laura Piazza\*

Università degli Studi di Milano, Department of Environmental Science and Policy, Via Mangiagalli, 25, 20133, Milano, Italy.  
[laura.piazza@unimi.it](mailto:laura.piazza@unimi.it)

In the current era of digital transformation, food companies are increasingly tasked with developing systems that enhance sustainability and operational efficiency. Small-to medium-sized food plants, in particular, often rely heavily on experience-based procedures, which can be financially suboptimal and lack the flexibility needed to adapt to changing demands. The digitalization and automation of production processes pose a critical challenge, offering significant opportunities to improve quality, competitiveness, and sustainability across the sector. Among the emerging technologies, the digital twin stands out as one of the most promising solutions. This paper presents results from a broader project aimed at enhancing productivity and efficiency while aligning with quality and sustainability objectives through the implementation of digital twins. Focusing on the industrial production process of vegetable broth, this study examines the initial phases of implementing a digital twin. These phases involve adopting a preliminary protocol that enables the reconstruction of material and energy flows within the process, the development of a Process Flow Diagram (PFD), the collection and analysis of process variables, and the resolution of a system of energy and material balance equations. This methodology addressed information gaps caused by limited sensor availability, facilitating the creation of a comprehensive dataset essential for digital twin implementation. Building on this dataset, the subsequent phase applies statistical methods, such as data reconciliation, to minimize errors and further enhance data accuracy. The refined dataset is then integrated into specialized simulation software, enabling the implementation of the digital model and the identification of optimization solutions. Additionally, the study highlights that integrating advanced sensor systems directly within the plant yields higher-quality data compared to traditional technical plant data or measurements obtained from offline sensors. This underscores the importance of investing in modern sensor technology to support the successful adoption of digital twin solutions in the food industry.

## 1. Introduction

Beyond assuring safety and quality, improving production efficiency, optimizing resource utilization (both material and energy-related), and enhancing logistics are among the top priorities of the food industry. Agriculture is a major user of water, primarily for irrigation in order to enhance the yield and quality of crops. It is therefore an essential driving force in the management of water use.

The agri-food sector is highly resource-intensive, characterized by substantial consumption of raw materials, water, and energy. Notably, it accounts for 32% of total gross water abstraction in the European Union for agricultural activities (EuroStat, 2025) and contributes approximately 29% of global CO<sub>2</sub>-equivalent greenhouse gas emissions (FAOstat, 2025). In response to these environmental and efficiency challenges, a recent Italian survey involving 365 food companies revealed that 82% have already initiated innovation processes by adopting Industry 4.0 technologies. These digital solutions enhance production efficiency, traceability, and logistics, enabling companies to develop new products and optimize existing ones. Moreover, they support reductions in time, energy, and resource consumption, while reinforcing control over critical steps of the production process, ultimately contributing to greater sustainability (Osservatorio Digital Innovation PoliMi, 2023).

An emerging tool that is increasingly adopted in Food Industry 4.0 is the “digital twin”. A digital twin is a virtual replica of a real-world process, continuously updated through sensor data and advanced big data analytics (Verboven et al., 2020).

The integration of digital sensors, connected via IoT systems, allows companies to monitor and control key process variables and KPIs in real time. Additionally, these tools support strategic decision-making aimed at improving efficiency and sustainability. As the deployment of sensors increases across different stages of production, the volume of collected data grows significantly. This requires food industry professionals to develop advanced analytical skills to effectively filter, manage, and utilize high-quality information for reliable digital twin modelling (Akyazi et al., 2020).

Digital twin modelling demands a multidisciplinary approach. In addition to data analytics, technical expertise is essential for defining the physical facts of processes. Data become truly valuable when they can explain and predict the overall system behaviour. In this regard, mathematical modelling plays a crucial role, particularly when implemented in computational environments. (Verboven et al., 2020).

“Mass and energy balances” help bridge information gaps caused by limited sensor availability—a common issue in small and medium-sized food plants. By integrating raw plant data with results of mass and energy balances, it be can possible to create comprehensive datasets that are fundamental for digital twin implementation.

The initial phases of digital twin development begin with defining a process flow diagram that outlines all unit operations. This is followed by data collection and the integration of raw data with statistical techniques, such as “data reconciliation”, to minimize errors and enhance data quality and reliability. Once a robust dataset is established, the next step is to integrate this high-quality data into specialized technical software for digital twin creation.

This case study examines the industrial production process of vegetable broth, aiming to develop a structured protocol for implementing digital twins in small-to-medium-sized food plants. The focus is on the preliminary phase of the digitalization process, highlighting the technical skills required in the food industry to support process engineers and data analysts in the successful deployment of digital twin technology.

## 2. Materials and methods

An industrial ready-to-eat food production company was chosen as a case study among the Italian industrial food industry. In particular, the focus was on the brick-packaging vegetable broth production line, where the initial stages of a protocol were completed to create a digital twin with an energy-optimization focus.

Here, we present the results outlining the preliminary phases of a protocol for developing a digital twin aimed at energy optimization. These phases include:

- Mapping the process and drafting a flow-sheet process.
- Analyzing the production process by collecting relevant information. Process variables, such as temperature and flow rate, were directly measured via on-line sensors. On the other hand, model parameters like specific heat and density were calculated from the average composition of the ingredients declared on the label.
- Processing data using a steady-state mathematical models to estimate missing values through mass and energy balances.
- Organizing the gathered information into a structured dataset.

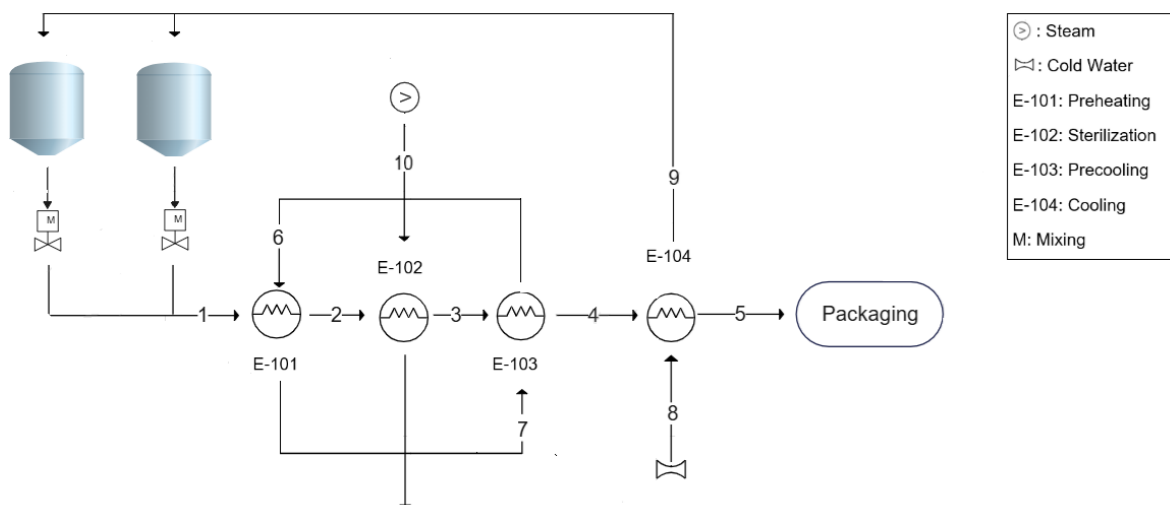


Figure 1: Schematic liquid broth production process. Number from 1 to 9 represent mass flow.

The company's vegetable broth production process consists of a few key steps and is validated by a temperature check at the end of the sterilization holding phase to ensure product safety and quality. Figure 1 provides a schematic representation of the entire process, illustrating the main operational phases. The mass flow was identified by numbers, from 1 to 9, while the energy action was indicated with the letter "E". Within the complete process, the sterilization phase has been identified as critical in terms of energy and resource consumption.

The raw materials are manually dosed using a calibrated scale, while the water used in the process, recovered from the heat exchanger, is reused at 40°C and measured by a float sensor in the mixing tank, which stops filling upon reaching 2800 L. The ingredients are added to a "mixing vat," where water from the cooling of sterilized broth or from the network circulates, facilitating the dissolution and homogenization of the ingredients. Mixing takes place in two insulated 3000 L tanks, which operate alternately to ensure a continuous flow toward packaging. Before proceeding to the next phases, the broth is filtered through a 1.5 mm stainless steel sieve, removing any suspended solids and preventing physical contamination. Subsequently, a volumetric pump conveys it to the heat exchanger.

UHT sterilization, E-102 (Table 2), aimed at eliminating *Clostridium botulinum* type A, represents a critical control point (CCP) in the process, as defined by the principles of the HACCP system established under the European Union Hygiene Package, specifically Regulation (EC) 852/2004, which mandates the identification and control of CCPs to ensure food safety throughout the production chain. The treatment occurs in a tubular heat exchanger, maintaining the broth at 134±3°C for 36 seconds. The exchanger is divided into six sections, including two six-meter sections for preheating E-101 (Table 2), a heat exchange section with steam at 158°C (6 bar), a 22-meter holding tube ensuring the required treatment time (approximately 36 s at 4200 L/h), two six-meter sections for pre-cooling E-103 (Table 2), and a final six-meter section dedicated to cooling E-104 (Table 2).

Preheating and pre-cooling occur in a closed circuit with service water (inlet at 92°C, outlet at 50°C), while final cooling uses counterflow network water at 18°C (16 m<sup>3</sup>/h), bringing the broth temperature to 35°C before packaging.

The raw data (Table 1) were collected using various types of instrumentation and sensors, ensuring a reliable dataset for process analysis and optimization. Where on-line data were not provided by in-line thermocouple, a digital thermometer offline, Dual-Laser Infrared Thermometer, Fisherbrand™ model 90012-60, resolution ± 0,1°, calibrated in December 2023 was used. In line data quality, as defined by the ISO 8000 standard, is a critical factor in the development of a digital twin. In particular, the use of multiple measurement methodologies can introduce biases due to variations in data quality associated with different measurement instruments, potentially affecting the consistency and reliability of the dataset (Unsworth et al. 2011).

*Table 1: Results of operative conditions monitoring (temperature T, flow rate V)*

	Flow	T (°K)	V (L/h)	Sensors
1	Broth input	313	4235 ± 39	Digital thermometer offline
2	Preheated broth	-	4235 ± 39	-
3	Sterilized broth	405,5 ± 0,31	4235 ± 39	Flowmeter and thermocouple
4	Precooled broth	-	4235 ± 39	-
5	Cooled broth	308 ± 0,9	4235 ± 39	Thermocouple
6	Preheating water	365	14000	Digital thermometer offline
7	Precooling water	323	14000	Digital thermometer offline
8	Cooling water in	291	16000	Digital thermometer offline
9	Cooling water out	313	16000	Digital thermometer offline
10	Steam	431	-	Expansion valve

In the case study analyzed in this work, the initial dataset was constructed using three distinct data acquisition methods, classified according to a hierarchy of reliability. The raw dataset was obtained by measuring temperature and volumetric flow rates every 20 seconds over a 20-minute production run, according to the company's internal GMP protocol. Online sensors, integrated directly into the production line, provided higher-

precision data compared to offline instruments and data internally by the company derived on their historical dataset due to greater measurement sensitivity, higher sampling frequency, and automatic data collection. Specifically, the thermocouple system has a resolution of  $\pm 0.01$  °C, whereas the offline thermometer's resolution is  $\pm 0.1$  °C. Temperature and flow rate measurements during the sterilization process were acquired through in-line operational software. Conversely, temperature data for service fluids and the broth exiting the mixing tanks were collected using offline sensors. Finally, service fluid volumes were estimated from raw data internally processed by the company based on their historical dataset to complete the dataset, although these estimates have lower data quality and reduced accuracy.

## 2.1 Mass and energy balances

From schematic liquid broth production process (Figure 1), it is possible to derive an equation's system of mass and energy balance (Table 2), a fundamental in process engineering. Mass balances are conducted for each individual operation and based on the principle of conservation of mass while energy balances are based on the principle of conservation of energy.

The energy balances within the fluids are measured with the equation:

$$Q = \dot{m} \cdot c_p \cdot \Delta T \quad (1)$$

where  $\dot{m}$  represent the mass flow rate of the fluid,  $c_p$  is the specific heat capacity of the fluid and  $\Delta T$  is the temperature variation within the single operation. The specific heat capacity ( $c_p$ ) and density ( $\rho$ ) are calculated considering the broth composition.

The energy supplied by steam during the sterilization phase is calculated by considering the enthalpy ( $H_v$ ) of saturated steam at 6 bar and 158°C. This value is obtained from the Mollier diagram, and the energy is determined using the following equation:

$$Q = \dot{m}_v \cdot H_v \quad (2)$$

where  $\dot{m}_v$  represent the steam flow rate and  $H_v$  is the enthalpy of saturated steam (David, 2008).

## 3. Results and Discussion

By solving the balance equations, gaps in the dataset caused by the partial deployment of sensor technology could be filled, thus improving the completeness and reliability of the initial raw dataset (Table 1). Unmeasured process variables were estimated through a balancing approach, where known values were incorporated into the equations, and the unknown variables were isolated and solved accordingly.

Table 2: system of energy and mass balance equations

Mass Flow	Energy Flow
$m_1 = m_2 = m_3 = m_4 = m_5$	$E_{101} = m_w \cdot c_{pw} \cdot (T_6 - T_7) = m_b \cdot c_{pb} \cdot (T_2 - T_1)$
$m_6 = m_7$	$E_{102} = m_v \cdot H_v = m_b \cdot c_{pb} \cdot (T_3 - T_2)$
$m_8 = m_9$	$E_{103} = m_w \cdot c_{pw} \cdot (T_6 - T_7) = m_b \cdot c_{pb} \cdot (T_3 - T_4)$
$m_v = \frac{E_{102}}{H_v}$	$E_{104} = m_w \cdot c_{pw} \cdot (T_9 - T_8) = m_b \cdot c_{pb} \cdot (T_4 - T_5)$
-	$E_{101} = E_{103}$

The broth specificity heat capacity ( $c_p$ ) and density ( $\rho$ ) were experimentally determined from the average composition of the broth declared on the label through the equation 3 e 4:

$$c_p = c_{pw}m_w + c_{pp}m_p + c_{pc}m_{pc} + c_{pf}m_f + c_{ps}m_s \quad (3)$$

where ( $c_p$ ) is the specific heat capacity of the overall product, has been calculated as the sum of the contributions from its individual components, specifically,  $c_{pw}m_w$  represents the contribution of water, the specific heat capacity and its mass;  $c_{pp}m_p$  corresponds to the contribution of proteins, their specific heat

capacity and their mass;  $c_{pc}m_{pc}$  accounts for the carbohydrates, their specific heat capacity and their mass;  $c_{pf}m_f$  represents the contribution of fats and finally,  $c_{ps}m_s$  refers to the contribution of salts.

$$\frac{1}{\rho_{tot}} = m_w \frac{1}{\rho_w} + m_p \frac{1}{\rho_p} + m_c \frac{1}{\rho_c} + m_f \frac{1}{\rho_f} + m_s \frac{1}{\rho_s} \quad (4)$$

where  $m_w \frac{1}{\rho_w}$  represents the contribution of water,  $m_p \frac{1}{\rho_p}$  the contribution of proteins,  $m_c \frac{1}{\rho_c}$  that of carbohydrates,  $m_f \frac{1}{\rho_f}$  the contribution of fats and  $m_s \frac{1}{\rho_s}$  the contribution of salts.

The broth specificity heat capacity resulted equal to 4.1 kJ/kg·K, while density is equal to 1.01 kg/L. The mass of the broth remains constant throughout the different stages of the investigated phases of the process, as it is not subjected to unit operations that could alter its flow rate, as resulting from the historical process data recorded (company communication)

The service water used for heat recovery circulates within a closed-loop system, maintaining a constant flow rate. In contrast, the cooling water is used to supply the upstream system in the mixing tanks; therefore, its flow is considered steady both at the inlet and outlet of the heat exchanger, within the process phase under investigation

The thermophysical properties of service water were assumed as a specific heat capacity of 4.186 kJ/kg·K and a density of 1.0 kg/L.

Regarding the sterilization unit operation, which employs saturated steam at 6 bar and 158°C, the corresponding enthalpy value of  $2.1 \times 10^3$  kJ/kg was obtained from standard thermodynamic reference tables based on the Mollier diagram, which describes the properties of saturated steam (David, 2008). Given the thermal energy required for broth treatment, E-102, (Table 2), the mass of steam necessary for the sterilization phase was determined as an average value over the analyzed time interval.

By solving the system of equations, the complete dataset was obtained (Table 3). The temperatures of the preheated and pre-cooled broth were determined using equations E<sub>102</sub> and E<sub>103</sub> (Table 2), resulting values of 326 K and 393 K, respectively.

During the resolution of the equation system, a discrepancy was identified between the calculated and theoretical values for the water flow rate in the heat recovery circuit. This inconsistency is likely due to differences in data quality and reliability, requiring recalculation to obtain a more accurate and coherent representation of the process. To address this, the flow rates of the service fluids (heat recovery water and cooling water) were excluded from the initial computation and subsequently treated as unknown variables. The recalculated values enabled the completion of the initial dataset, resulting in a cooling water flow rate of 14000 L/h and a heat recovery water flow rate of 1248 L/h (Table 3). The results will be re-evaluated in a later phase of the project using appropriate data reconciliation protocols.

*Table 3: Complete dataset to be used for the digital twin creation phase*

	Flow	T (°K)	V (L/h)
1	Broth input	313	4235
2	Preheated broth	326	4235
3	Sterilized broth	405,5 ± 0,31	4235 ± 39
4	Precooled broth	393	4235
5	Cooled broth	308 ± 0,9	4235
6	Preheating water	365	1248
7	Precooling water	323	1248
8	Cooling water in	291	16000
9	Cooling water out	313	16000
10	Steam	431	2 x 10 <sup>5</sup>

The resulting dataset was then utilized in the subsequent stages of digital twin development, which included statistical and analytical data analysis. These analyses further supported the recalculation of the heat recovery water flow rate, challenging the accuracy of its theoretical value.

The construction of a complete and reliable dataset for digital twin development required an in-depth analysis and comprehensive understanding of the production process. This necessity arose because raw data were often characterized by poor quality and limited reliability, primarily due to inconsistencies in data collection methodologies. Addressing these challenges was crucial to ensuring accurate process representation and effective model development. The findings highlight the crucial role of advanced sensor technologies and specialized software application in ensuring high-quality data acquisition. This is particularly evident in the measurements of broth flow rate and temperature at the end of the sterilization holding phase, which were found to be precise and reliable. Conversely, the theoretical data related to the flow rates of preheating and precooling fluids were ultimately found to be inaccurate, failing to provide a realistic representation of the process.

#### 4. Conclusion

This research focused on the initial phases of a protocol for the development of a digital twin (DT) aimed at optimizing the energy efficiency of a food industry production process in the light of process sustainability. The study involved reconstructing the material and energy flows within the liquid vegetable broth production process, developing a process flow diagram, collecting and analyzing process parameters recorded by the company, and solving a system of equations for mass and energy balances to describe the individual operations in detail. This approach helped compensate for information gaps due to the limited availability of sensors, enabling the creation of a comprehensive and detailed dataset for the construction of the digital twin.

The complete dataset used for the DT construction was derived from an initial raw dataset, which was built by extracting values that describe the process. The initial data were obtained using different data collection methodologies, which presented varying levels of precision and reliability. The study of the process, along with the resolution of a system of balance equations, allowed for the completion of missing information and the correction of erroneous data that did not accurately represent the process.

For the development of a digital twin in the food industry, this preliminary approach, combined with food technology expertise, is essential for accurately understanding and describing the processes. The study focused on a relatively simple industrial-scale process to enhance comprehension and demonstrate the feasibility of applying enabling technologies within the food sector, which remains underdeveloped in this area.

This research aims to promote the adoption of digital twin technology in the industry, driving economic benefits, sustainability, and, in the future, potential improvements in product quality. By incorporating quality factors into mathematical models, the study paves the way for holistic process optimization, balancing efficiency, sustainability, and product performance.

#### Acknowledgments

Project funded under the Italian Ministry of University and Research – PRIN 2022 – project “Digital food processing and optimization for sustainability (DIGITALFOOD)” (Grant number 2022XJ4Z42).

#### References

- Akyazi T., Goti A., Oyarbide A., Alberdi E., Bayon F., 2020. A Guide for the Food Industry to Meet the Future Skills Requirements Emerging with Industry 4.0. *Foods*, 9, 492
- David R. Lide, 2008. *CRC Handbook of Chemistry and Physics*. CRC Press, Taylor and Francis Group. Boca Raton (FL)
- EuroStat, 2025. C39 Water abstraction in agriculture /WEI+;
- FAOstat, 2025. Greenhouse gas emissions from agrifood systems,
- ISO-8000:2022, Data Quality- Part1: Overview.
- Osservatorio Digital Innovation PoliMi, 2023. Report: Adozione delle tecnologie digitali nel settore Agroalimentare italiano.
- Unsworth K., Adriasola E., Johnston-Billings A., Dmitrieva A., Hodkiewicz M., 2011. Goal hierarchy: Improving asset data quality by improving motivation. *Reliability Engineering & System Safety*, 96(11), 1474-1481
- Verboven Pieter, Thijs Defraeye, Ashim K. Datta, Bart Nicolai, 2020. Digital twins of food process operations: the next step for food process models. *Current Opinion in Food Science*, Volume 35, Pages 79-87., Pages 79-87.