

Digital twins in the green life sciences

Willem Jan Knibbe, Lydia Afman, Sjoerd Boersma, Marc-Jeroen Bogaardt, Jochem Evers, Frits van Evert, Jene van der Heide, Idse Hoving, Simon van Mourik, Dick de Ridder & Allard de Wit

To cite this article: Willem Jan Knibbe, Lydia Afman, Sjoerd Boersma, Marc-Jeroen Bogaardt, Jochem Evers, Frits van Evert, Jene van der Heide, Idse Hoving, Simon van Mourik, Dick de Ridder & Allard de Wit (2022) Digital twins in the green life sciences, NJAS: Impact in Agricultural and Life Sciences, 94:1, 249-279, DOI: [10.1080/27685241.2022.2150571](https://doi.org/10.1080/27685241.2022.2150571)

To link to this article: <https://doi.org/10.1080/27685241.2022.2150571>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 27 Nov 2022.



Submit your article to this journal [↗](#)



Article views: 398




View related articles [↗](#)



View Crossmark data [↗](#)



Digital twins in the green life sciences

Willem Jan Knibbe , Lydia Afman, Sjoerd Boersma, Marc-Jeroen Bogaardt, Jochem Evers, Frits van Evert, Jene van der Heide, Idse Hoving, Simon van Mourik, Dick de Ridder and Allard de Wit

Wageningen University & Research, Wageningen, The Netherlands

ABSTRACT


Digital twins provide a new paradigm for the integrated use of sensor data, process-based and data-driven modelling, and user interaction, to explore the behaviour of individual objects and processes. Digital twins originate from an engineering context and were developed for machines and mainly physical and chemical processes. In this paper, we further develop an understanding of digital twins for the green life sciences, which also include biological and social processes. We report on three use cases, in precision farming, greenhouse control and personalized dietary advice, focusing on practical benefits and challenges of digital twins compared with other research methods. This research extends earlier more conceptual research on digital twins in this domain. We find benefits in increased accuracy and impact because of the real-time data connection of digital twins to their real-life counterparts. Specification, availability and accuracy of relevant data sources are still major challenges. Specifically, when using digital twins for personalized advice, further research is needed on nontechnical aspects so that users will comply with the advice from the digital twins. We have outlined four directions of future research and expect that further research will include data-driven modelling to simulate the complex character of living objects and processes and at the same time develop approaches to limit the amount of required input data.

ARTICLE HISTORY Received 24 April 2022; Accepted 17 November 2022

KEYWORD Digital twins; precision farming; greenhouse control; personalized dietary advice

1. Introduction

Digital twins are digital representations of systems or processes with three specific characteristics distinguishing them from other representations such as data sets or simulation models. First, they represent one specific instance of a system or process, which is duplicated by the twin. Second, this instance is regularly monitored, and the resulting data are used to update the twin so that instance and twin are kept synchronized. Third, a twin allows user interaction to investigate its hypothetical time evolution under different

CONTACT Willem Jan Knibbe  willemjan.knibbe@wur.nl

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

scenarios. Digital twins employ in this manner advanced scientific knowledge for interactive and quantitative comparison of different scenarios for feedback, control or intervention. In particular, this can be done in real time, and for unique real-world instances. This dynamic and responsive character of digital twins distinguishes them from, e.g. digital shadows (Bergs et al., 2021; Kritzinger et al., 2018). Figure 1 illustrates this new paradigm.

This figure also shows several optional properties of a twin. For instance, advanced visualization is often used with digital twins, it is efficient and robust to build the twin as a modular system, it may provide information on multiple scales of the actual object or process, and it is interesting and challenging to consider the twin to interact with this object or process. However interesting, we do not consider such properties as essential to the definition of a digital twin.

Essentially, digital twins allow for simulation. Increasingly, data-driven methods are used for simulation as an alternative to models based on explicit process understanding (“process based”). Evolving through regression analysis, advanced statistical methods and various types of machine-learning algorithms, an enormous body of research with application to widely varying domains clearly shows the additional benefits of data-driven methods (Raghu

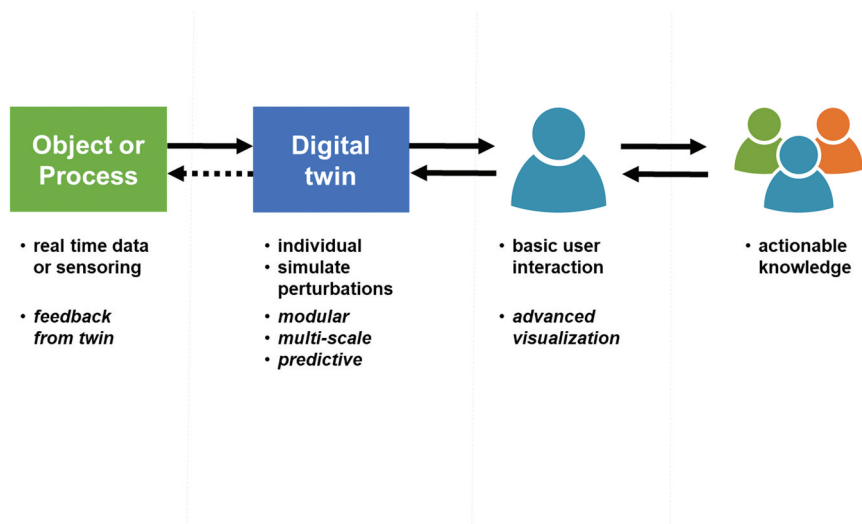


Figure 1. The digital twin is a representation of an actual object or process, monitored in real time, with the flow of information or data from the actual object or process to its twin. The twin itself represents a single instance of this object or process and can simulate its time evolution under normal or perturbed conditions. It offers facilities for basic user interaction and in this manner provides actionable knowledge, the added value of the twin. We consider these properties as essential to distinguish digital twins from other digital representations. The dotted arrow illustrates an optional feedback loop.

& Schmidt, 2020). Often, such methods adopt a two-phase approach. The first phase employs large amounts of (historical) data to calibrate, or “train”, the algorithms. In the second (application) phase, these are applied to new data for answering relevant questions.

The simulation part of digital twins can be based on data-driven as well as process-based methods. In both cases, digital twins can take advantage of their real-time data connection, where real time refers to time scales relevant for the object or process under investigation. Real-time data help to verify the quality of the simulations, to – regularly and automatically – adjust and/or recalibrate the model, or (for process-based methods) support regular parameter tuning.

The characteristics of digital twins, combined with the possibility to include different kinds of simulation models, offer many opportunities to quantitatively investigate complex systems and processes. Digital twins were conceived for product lifecycle management (Grieves, 2016). In NASA Technology Roadmaps, their definition explicitly links them to physical systems (Rosen et al., 2015). They have also found extensive application in industry, such as in manufacturing (Kritzinger et al., 2018). Generally, digital twins appear to be developed mostly in physical, well-controlled, technological domains.

Wageningen University & Research hosts research and education in a broad field that we will indicate here with the term “green life sciences”. This includes the living environment, with subjects such as the terrestrial environment, climate, human and social processes, the entire system of agriculture, food, nutrition, and health, as well as plant and animal sciences. Research focuses on systems and processes crucial for human society which require, due to their complexity, advanced analytical and data-driven solutions for quantitative investigations (Athanasiadis et al., 2020).

Given their benefits in modelling physical/technical entities, we are interested in the benefits digital twins can provide in the area of complex living systems and processes. In this work, we investigate the potential of digital twins in the green life sciences. In particular, we wish to know if the digital twin characteristics, applicable to non-living systems, can as well be applied in our domain. In addition, we intend to establish if these digital twin characteristics provide benefits compared to other methods; for example, a better understanding and so a better ability for deciding on how to obtain desired outcomes for the systems and processes investigated.

Here, we specifically investigate the practical challenges and benefits of digital twins in the green life sciences over alternative approaches. We have chosen three use cases, namely, digital twins of farm operations, greenhouse operations and human digestion. In the following sections, we first provide a brief overview of the domain of green life sciences. Then, we discuss the use of digital twins in this domain. We continue with a description of our research questions for the three use cases, followed by the analysis and results pertaining to these questions and end with conclusions.

2. Green life sciences

The mission of Wageningen University & Research (WUR) is “To explore the potential of nature to improve the quality of life”. Its three strategic domain areas are Society and well-being; Food, feed and biobased production; and Natural resources and living (“Finding Answers Together Strategic Plan” 2019–2022”, 2019). These areas overlap in scientific content but have a distinct focus. WUR combines university research and education with public- and private-funded research institutes in a single organization. Research covers a broad spectrum of subjects coherently focused on the complex living systems and processes of the human environment. The spectrum includes social sciences (e.g. economics, consumer behaviour, philosophy), exact sciences (e.g. physics and chemistry of food production, climatology, hydrology), biology (plant and animal sciences, human nutrition and health) and technology (bioprocess engineering, environmental technology, food safety). Fundamental and technology-oriented research at WUR is always aimed at applications in the three strategic domains. We refer to these as the “green life sciences”, distinct from the medically oriented “red life sciences” that tend to have a stronger focus on the internals of living beings.

3. Digital twins in the green life sciences

WUR embraces the opportunities and challenges of data science and artificial intelligence for its mission. As one of its initiatives, WUR has defined a strategic research programme “Digital Twins”. Earlier research has already explored the application of digital twins in the green life sciences. Their potential and challenges have been investigated in a.o. use cases in agriculture (Pylianidis et al., 2021), (Verdouw & Kruize, 2017), in smart farming (Verdouw et al., 2021), a review in horticulture (Ariesen-Verschuur, Verdouw and Tekinerdogan, 2022), design patterns including case studies (Tekinerdogan & Verdouw, 2020) and an exploratory investigation of animal farming (Neethirajan & Kemp, 2021). Ethical aspects related to digital twins have also been researched (Korenhof et al., 2021; Van der Burg et al., 2021).

Application-oriented research underlines the potential of Digital Twins. For instance, (Alves et al., 2019) leverage wireless environmental sensing technology (air, soil), with facilities for data analysis and visualization, as a first step in realizing a digital twin for farm resource management. Further possibilities are investigated by (Elijah et al., 2021), such as the incorporation in farming in combination with robotics, crop growth models and food preservation technology. (Nasirahmadi & Hensel, 2022) describe the state of the art and give a framework for digital twins in agriculture, to support farmers. In a somewhat broader context, (Kamilaris et al., 2021) provide several examples, such as a system setup for a beekeeping digital twin in an urban environment, and

digital twin technology combined with digital elevation maps and augmented reality as preparatory information for hiking tourism.

The examples mentioned above mostly focus on perspectives, requirements and implications, or provide prototype applications. The seven digital twin maturity categories presented by Uhlenkamp et al. (2022; context, data, computing capabilities, model, integration, control, human-machine interaction) are at most partially applicable, indicating that these examples are generally in early maturity stages. A different framework for distinguishing maturity levels (Metcalf et al., 2022) paints the same picture. Note, however, that in (Uhlenkamp et al., 2022) less than 2% of reviewed applications (2 out of 131) is within agriculture, and less than 4% (5 out of 131) within logistics (which can be considered part of the food chain). The number of applications in the green life sciences appears small compared to other domains, which motivated us to research practical challenges and benefits.

In the WUR Digital Twin research programme, three flagship projects investigate practical challenges and benefits of digital twins of (1) the nitrogen cycle at a farm, (2) operations of a tomato greenhouse and (3) the post-meal human fat response in the blood circulation. These flagship projects are our use cases for investigating digital twins in the green life sciences.

Below, we first provide some background on one of our flagship projects, the nitrogen cycle at a farm, to illustrate how the digital twin concept fits in the continuous development of our research.

3.1. Digital twins as a next step in research

As an example, Tipstar is a dynamic model of the growth and development of potatoes developed at WUR (Jansen, 2008). Recommendations for irrigation and nitrogen fertilization based on real-time simulation run with Tipstar resulted in agronomic and economic performance at least as good as that obtained by successful farmers (Jansen et al., 2003). Application of Tipstar on commercial scale was, at the time, precluded by the cost and labour associated with collecting the input data needed to run the model. The widespread availability and affordable cost now of new sensors (e.g. satellites and drones) and new IT infrastructure (e.g. software for registration of farm management data, access to weather data) inspired the flagship project for a digital twin at the farm, which includes Tipstar as a simulation model implemented in the open-source real-time data service platform Farmmaps (<https://farmmaps.eu>).

The development of Tipstar is exemplary for those in the domains of the other flagship projects, on greenhouses and human food responses, which are discussed below. Other examples include decision-support in cotton (McKinion et al., 1989) and maize (Sela et al., 2018) farming.

On grass-based dairy farms, proper prediction of biomass and its nutritional value is considered essential for modern day grassland management. To provide for this, a web application for farmers use (GrassSignal) has been developed and is being tested in practice. The approach here is to combine model prediction of biomass with sensor data (Hoving et al., 2019). The use of a growth model is seen as the basis for predicting biomass and nutritional value. Biomass prediction is being adjusted with spectral measurements or grass height. Concerning spectral measurements vegetation index WDMI red was the best predictor of yield.

3.2. Benefits and challenges for digital twins

While technological developments now allow the development of digital twins, the question remains what exactly their advantages and disadvantages are compared to other representations widely used in our research domains: simulation models, control systems and scientific advice. Simulation models formalize scientific knowledge for computational purposes. Decision support systems incorporate knowledge in decision rules to provide actionable knowledge supporting human control of systems or processes, potentially reacting to changing conditions. Scientific advice details recommended actions for specific situations, often in oral or written communication. Compared to each of these representations, digital twins offer specific possibilities:

- Digital twins differ from simulation models in their real-time data connection with a particular instance of a system or process. This can provide more accurate simulations than generic models of systems or processes lacking real-time data. It does, however, require a sufficiently accurate data connection with a particular instance, and a method to feed these data consistently as input information for the simulations.
- Digital twins differ from decision support systems in their ability to represent over time the system or process to be controlled. Including a digital twin in a control system makes it possible to investigate in more detail the options of human intervention to optimize system/process operation. It does, however, require more detailed and relevant information on the particular controlled system or process, in real time.
- Digital twins differ from scientific advice as they can efficiently provide interactive advice, for different scenarios based on actual data and specific for an individual object or process. This advice can in principle be automated and easily accessed to provide actionable knowledge, for instance, through mobile equipment. Collecting user feedback can reduce the risks of incorrect advice by adapting to individual situations. This does require efficient and effective user interaction and compliance.

In this paper, we investigate the benefits of digital twins compared to simulation models, as part of decision support systems, and as a tool for providing scientific advice. Evidently, digital twins may be used in many additional ways not considered here, for instance, for communication and training (interactively visualizing potential scenarios, “flight simulator”) or for product development (virtually testing various product or production alternatives, “digital laboratory”).

4. Research questions

In this section, we introduce the flagship projects and the related research questions.

The “Digital Future Farm” (DFF) project is based on an existing data platform and simulation environment for precision agriculture. We focused on the challenges of the real-time data connection to investigate the applicability of digital twins, and we analysed simulation results to study their benefits. An important measure of benefit is whether farmers can make better tactical and strategic decisions with output of the digital twin, compared to making decisions with output of non-digital twin simulation models. In this case, we investigated how the concept of a digital twin enables the next step in precision agriculture.

The “Virtual Tomato Crop” (VTC) project builds on existing work in climate control for greenhouses. This is an active area of research, with many challenges in combining data and model simulations to improve decision-making. We used this case to investigate benefits of including digital twins as part of automatic control systems. Specifically, we researched the challenges involved in integrating digital twins in more complex systems.

The “Me, My Diet and I” (MMD) project aims to increase the health benefits of dietary advice by adopting digital twins to personalize recommendations. We investigated whether algorithm-based dietary advice can support advice by health professionals. This research extends current knowledge on the influence of personal characteristics on human response to food intake, as well as on the receptiveness to dietary advice offered either directly by professionals or given by a digital twin as intermediary.

5. Analysis and results

We adopted different approaches for the different use cases researched in this paper. This section provides information on background, methods, and results for each.

5.1. Case digital future farm

5.1.1. Background

Agriculture has potentially large negative impacts on the environment. Society places regulatory pressure on farmers to limit the use of input to reduce these impacts. To optimize the use of inputs and reduce losses, farmers need precise, real-time information about the status of crops, soils and livestock, as well as information about the likely outcome of management decisions. The potential of dynamic crop growth models fed with real-time data to support farmers' decisions has long been recognized. Several monitoring and/or decision support systems are available, of which FarmMaps and Beregeningssignaal (<https://www.zlto.nl/beregeningssignaal>) are examples. However, decision support including real-time data connections based on dynamic models is currently not widely used.

5.1.2. Methods

We adopted a model-oriented approach for this case study. In particular, we focused on incorporating currently available simulation models in the digital twin by taking in real-time available sensor information and connecting to novel forms of visualization, as a basis for decision support. This case study focuses on our approach to data assimilation.

5.1.3. Challenges

We find four main reasons currently limiting the widespread adoption of decision support systems:

- A fragmented modelling landscape: different models need to be combined to support complex decisions by farmers, e.g. to include the effects of crops, soils, and livestock. However, such different models are not usually readily interoperable. For instance, it requires major efforts to harmonize definitions and units before starting the integration.
- Calibration of models for specific commercial farms: advice is only useful if it is sufficiently specific to add to the expertise and insights of farmers. This requires accurate calibration of models, a major effort given the complexity and diversity of the processes involved.
- Models never capture all relevant processes: they are necessarily an abstraction and simplification from reality and their results need to be interpreted with care before they can be used as a solid basis for decision making. It requires in-depth investigations to determine the validity of model results as a basis for decision making.
- IT infrastructure: with the growth of technological possibilities, the complexity of design and development increases exponentially due to

the many available options for design principles, technical standards, and methods of implementation.

In the Digital Future Farm (DFF) project we develop a digital twin representing a particular arable or dairy farm, with a focus on the nitrogen cycle (Van Evert et al., 2021). This digital twin comprises the components necessary to simulate interventions and to generate actionable knowledge to (1) reduce nitrogen use of farms while maintaining or increasing crop yields and (2) minimize nitrogen losses on farm. The DFF includes various models currently in use at Wageningen University & Research for the different processes on a farm. In addition to process-based models, we develop data-driven methods for selected sub-systems (not further discussed here).

The data needed by the DFF digital twin are varied and many. For crop farming, soil models are needed that require information about the soil. For The Netherlands, this is available online from BOFEK (De Marke et al., 2021). In the US, this information can be obtained from SSURGO (<https://websoilsurvey.nrcs.usda.gov>); SoilGrids (<https://soilgrids.org>) aims to provide this information for the entire world. Both soil and crop models need information about daily weather, especially temperature, irradiation, precipitation, humidity and wind-speed. This information can typically be obtained online and in real-time from local weather stations, national weather services, or global providers such as Weather Corp. Crop and soil models also need information about farmer management activities, such as tillage, planting, fertilizing and irrigation. This information is typically stored in a so-called Farm Management Information System (FMIS), of which many commercial versions are available. Farm operations are often entered manually into an FMIS and therefore it is a challenge to have timely access to accurate data. Further, FMIS from different vendors are often not interoperable. Standards such as Agro-XML (www.agroxml.de), EDI-Teelt++ (www.agroconnect.nl), and others are being developed but none of them seems already to have gained widespread attention. In The Netherlands, the FMIS of Dacom (<https://www.dacom.nl/nl>) is widely used and can be used in our DFF to provide input to crop and soil models.

Many sensors and sensing systems are available to make real-time observations on crops and soils (Van Evert, 2019) for updating the digital twin. Reflectance measurements are of prime importance for crops, as they provide information about aboveground biomass and nitrogen content in crops. Reflectance can be measured with sensors borne on satellites, UAVs or drones, or tractors. Moisture measurements are similarly important for soils, and in particular automated soil moisture sensors provide useful information. Accurately measured soil nitrogen content is equally important given our focus on the nitrogen cycle; thus, far this requires destructive analysis but automated detection via Vis-NIR spectroscopy is under active investigation (Gebbers, 2018).

Exact determination of crop yields from yearly recurring crop harvesting can provide valuable information for recalibration of crop and soil models, and so provides opportunities for investigation potential sensing deviations. Especially interesting are yield maps, which provide yield as a function of the position in the field. Yield maps can be generated by combine-harvesters for crops such as wheat and maize. The quality of yield maps for potatoes depends on soil and harvesting conditions. In the future, we expect possibilities for more accurate and extensive yield determination through analysis of visual information from cameras.

In dairy farming, grassland use or consumption is an important parameter in the nitrogen cycle. This is, however, particularly difficult to determine due to irregular animal behaviour and may require, e.g. Global Navigation Satellite System receivers on the animals. Reflectance measurements may be an alternative, similarly to the use in crops such as wheat and potato. The sensitivity seems, however, to be somewhat less, perhaps because grass has a permanent stubble layer which influences the measurement; another explanation may be that the low amount of above-ground biomass, due to frequent harvesting, tends to provide a signal which is a mix of the grass and the surface layer below. This reduces the sensitivity to grass characteristics. Like crop farming, dairy farming provides opportunities for model recalibration, e.g. using cow sensors to predict fresh grass uptake (Schils et al., 2019) or measured quantities and quality of produced milk.

5.1.4. Nitrogen fertilization in potatoes

The baseline for nitrogen fertilization in potatoes in The Netherlands is to apply all nitrogen just before or just after planting. This frequently results in either an over-supply or an under-supply of N, because the actual demand is influenced by spatial variability in the field and by the growing conditions during the season. Wageningen UR has developed a recommendation system where two-thirds of the recommended amount of N is applied at planting; around July 1st the amount of N taken up by the crop is then measured via canopy reflectance and compared with the amount of N that would have been taken up by a crop growing without N limitation (determined with model simulations) and the difference is then applied as side dress N. This system maintains yield and results on average in a 15% reduction in N use (Kempenaar et al., 2017; Van Evert et al., 2012).

An important limitation is that it does not consider nitrate leaching, mineralization of organic matter, and other processes that influence the availability of N in the soil. For example, after several weeks without rain the system will recommend a high side dress N rate even though crop growth is limited by lack of water (and not by lack of nitrogen).

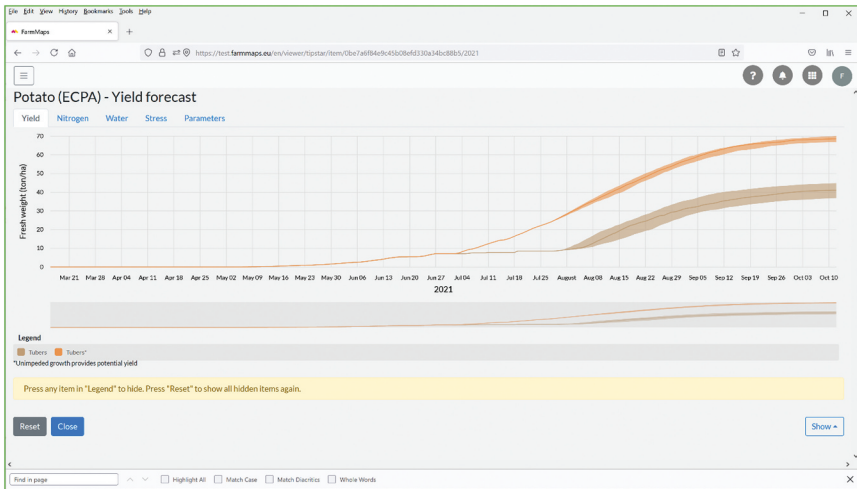


Figure 2. Screenshots from <https://farmmaps.eu> showing a simulation and forecast with the Tipstar potato model for a field near Wageningen, The Netherlands, in 2021. Simulated fresh tuber yield. The crop was planted on 15 April 2021 and the forecast was made on 14 July. From planting to 14 July, observed weather was used for the simulation. For the two weeks following 14 July, forecast weather was used. From 1 August, a stochastic simulation using 30 years of historic weather was performed, thus from 1 August the line becomes a plume.

This is where using a dynamic simulation model can be expected to lead to more accurate recommendations. For this purpose, we have made it possible to run Tipstar in real-time for potatoes in The Netherlands, with soil hydrologic data derived from the Dutch national soils database (Heinen et al., 2021), daily weather (and two-week forecast) from The Weather Company (<https://www.ibm.com/weather>), and manual crop registration. This approach is implemented on the FarmMaps platform (Figures 2 and 3) and can be (but has not yet been) applied to other crop or dairy-related models (e.g. WOFOST, WatGro, Beregeningssignaal, Grassignaal) as well.

As part of our approach, an Ensemble Kalman Filter (EnKF) has been implemented in the DFF, based on earlier work (Evensen, 2003; De Wit et al., 2012; De Wit & van Diepen, 2007). The EnKF provides a solution for estimating the state of the system by combining sensor information and model predictions considering the uncertainty in both. The goal is to provide a consistent time evolution of the simulated system which is adjusted by observations when they come available. This approach including an EnKF is not restricted to Tipstar and can be used to assimilate real-time observations into any of the simulation models used in the DFF.

The use of Tipstar and EnKF is demonstrated in Figure 4. Data assimilation improved the tracking of aboveground growth (as indicated by Leaf Area

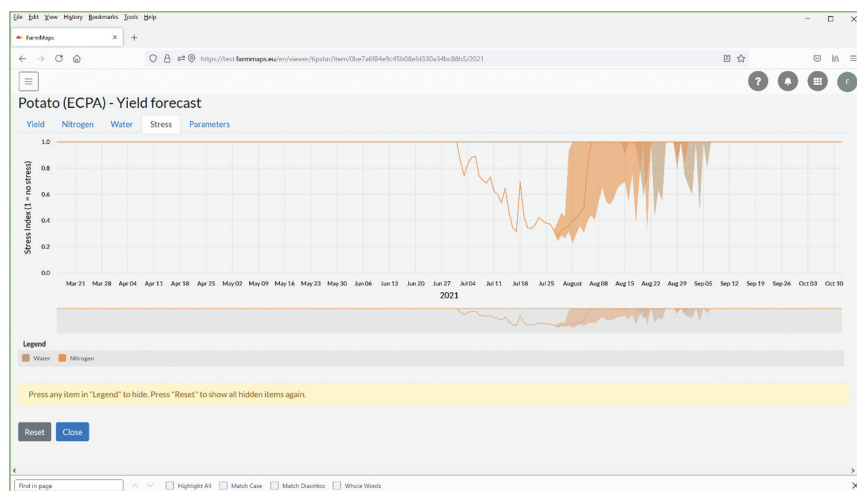


Figure 3. Same as Figure 2 but for simulated reduction of crop growth due to water-stress and N-stress. It shows that the crop is running out of N around the time of the forecast. At this point, the crop still looks healthy, and while the farmer cannot observe problems the model predicts these will occur soon.

Index, or LAI) but had negligible effect on the tracking of tuber growth. Further work is needed to determine how effective data assimilation is here in practice. A possible improvement could come from adding tuber weight as a state variable in the analysis and let the EnKF adjust tuber weight as well. We encounter several further research questions related to the required sensor information. The choice of measurement quantities, frequencies, and accuracies needs to be further investigated. We also note that in these investigations different prediction horizons need to be considered. Effects on prediction of state variables a few weeks in the future (supporting tactical decisions) may well differ from effects on end-of-season predictions (supporting strategic decisions).

As an illustration of the fact that benefits of a real-time data connection are strongly dependent on the context, we note that, e.g. crop growth models have many parameters, some of which influence the model's behaviour at any point in time, but there are also parameters that have an influence only during specific periods of the growing season. Figure 5 shows the effect of two parameters: one parameter determines LAI at emergence and therefore influences the simulation in the beginning of the growing season; another parameter determines the life span of individual leaves and therefore starts to influence the simulation only when the first (oldest) leaves reach their maximum age and start dying.

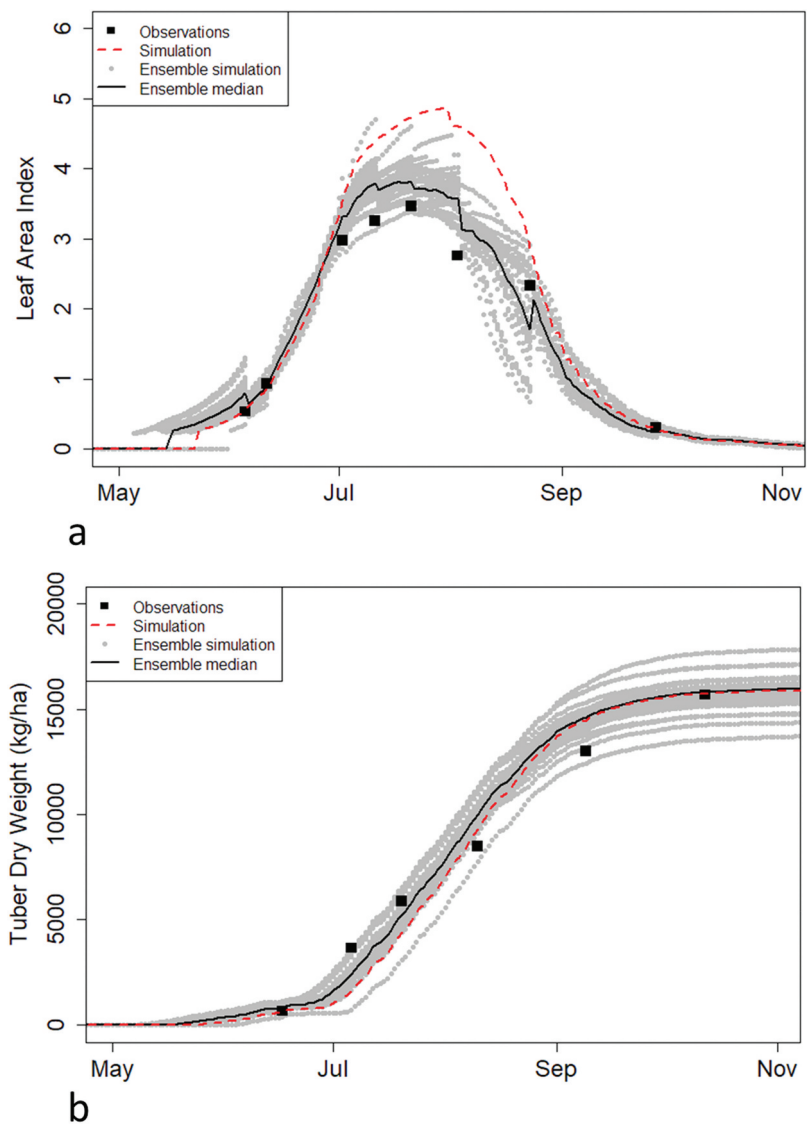


Figure 4. Simulation of potato growth with EnKF and Tipstar for Leaf Area Index (a) and for Tuber dry matter yield (b). The dashed red line shows the default simulation for this field. An ensemble of simulations ($n = 50$) is represented by grey-dotted lines; the ensemble median is shown with a solid black line. The filled square symbols show the LAI observations based on Sentinel-II imagery which were used to update the ensemble, and destructively measured tuber weights, for (a) and (b), respectively.

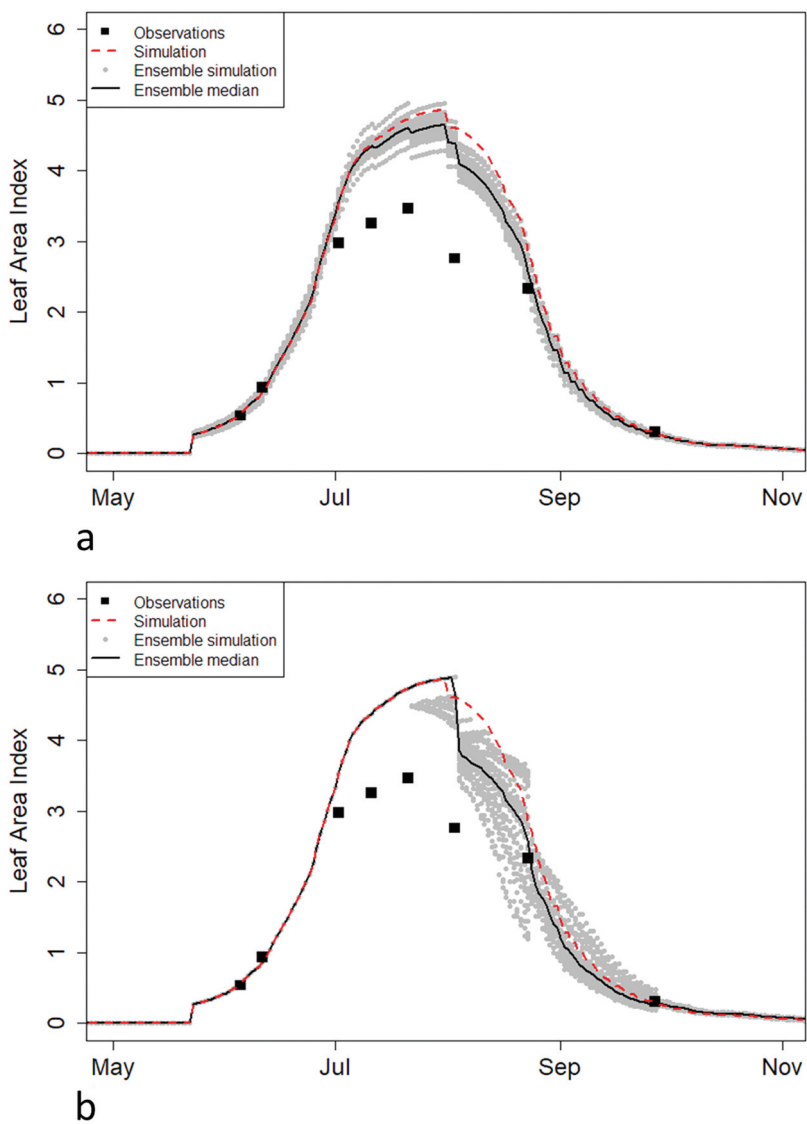


Figure 5. Variation in the EnKF simulation created with a parameter that influences LAI at emergence (a) and with a parameter that affects the lifespan of leaves (b).

5.1.5. Summary

The case of Tipstar on FarmMaps and the data assimilation examples demonstrate the technological readiness of the digital twin approach for arable farming. The connection to real-time data has been demonstrated for a limited application, but for most arable farmers this remains a major obstacle. We take the approach of demonstrating the usefulness of the arable farming digital

twin, in the expectation that this will provide a stimulus for growers and technology providers to solve the operational data connection challenges.

5.2. Case virtual tomato crop

5.2.1. Background

The VTC digital twin is based on a detailed simulation model of the dynamics of climate and individual tomato plants in a greenhouse. Tomato is not only a valuable export crop in the Netherlands but also widely studied as it represents many challenges also found in other greenhouse crops. The model simulates relevant crop processes (light capture, photosynthesis, assimilate allocation, growth of leaf area and stem extension, yield), greenhouse climate dynamics (temperature, humidity, CO₂), climate control measures (opening of windows, heating, CO₂ injection, lighting, and screening), and crop handling (harvesting, leaf pruning). The digital twin represents the individual tomato plants in the greenhouse. The model configuration (parameters settings) associated with a specific cultivar and greenhouse configuration are tuned based on real-time sensor measurements taken from the real tomato crop.

Potential key advantages of using a digital twin follow from exploiting the detailed information available from real-time data streams. Merging actuator and sensor data with process-based greenhouse crop modelling enable high precision monitoring of system states by application of filtering algorithms (Hameed, 2010; Van Mourik et al., 2019), analogously to the approach followed for the DFF. In this manner, model parameters can be calibrated to specific system properties for an individual greenhouse or crop cultivar (Speetjens et al., 2009), or even to individual plants. Then, with detailed monitoring, accurate prediction, and improved control performance, external factors such as yield rate and energy efficiency can be optimized.

5.2.2. Methods

In this case study, we adopted a systems-oriented approach, investigating the challenges resulting from including a digital twin into greenhouse control systems. We focused on required information and accuracy to outline the further development of a digital twin as part of different types of control systems.

5.2.3. Improving decision support with a digital twin

In academia and in the greenhouse support industry, there is great interest in model-based automated decision support, which can be provided in three types (Van Mourik et al., 2021): 1) monitoring the current state of crop and climate variables by updating model states and parameters with sensing information, 2) predictions about future development of crop and climate under various scenarios of crop handling and climate control, and 3) high-level control advice about optimal setpoints, such as for indoor temperature

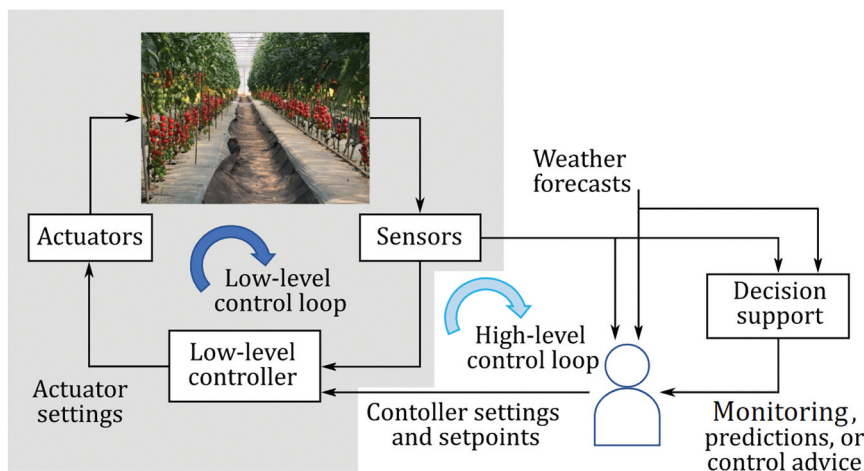


Figure 6. The role of decision support in greenhouse operational management. Two control loops are shown: a low-level management loop (with a low-level controller) with input consisting of settings and setpoints provided by the grower, and a high-level loop, in which a grower determines those settings and setpoints based on sensor information, forecasts and automated decision support (observation, prediction or control advice). Source photo: <https://european-seed.com>.

and pruning rate, and additional settings or constraints, such as maximum indoor humidity or minimum leaf area. Type 3 is currently mainly provided by consultancy, but this can in principle be automated, and even developed into a fully automated controller. For predictions (type 2) but also for control advice (type 3), forecasts about crop development, weather, and market prices need to be combined for optimizing, e.g. yield rate, crop quality and rewards, and energy use. The role of decision support in greenhouse management is visualized in Figure 6.

Decision support by monitoring (type 1) with a digital twin requires a method to correctly estimate the current state of the system from the sensor information. Decision support by prediction (type 2) with a digital twin additionally requires consistent state updating as input for model predictions. Decision support by control (type 3) further requires a control algorithm to automatically compute optimal management of input and actions, to provide advice, or even fully automatic control. Figure 7 briefly illustrates the three different decision support types including the integration of a digital twin.

During implementation, we encountered several research questions, which we think are of immediate interest to academics and industry involved in developing intelligent sensing and actuation technology including digital twins.

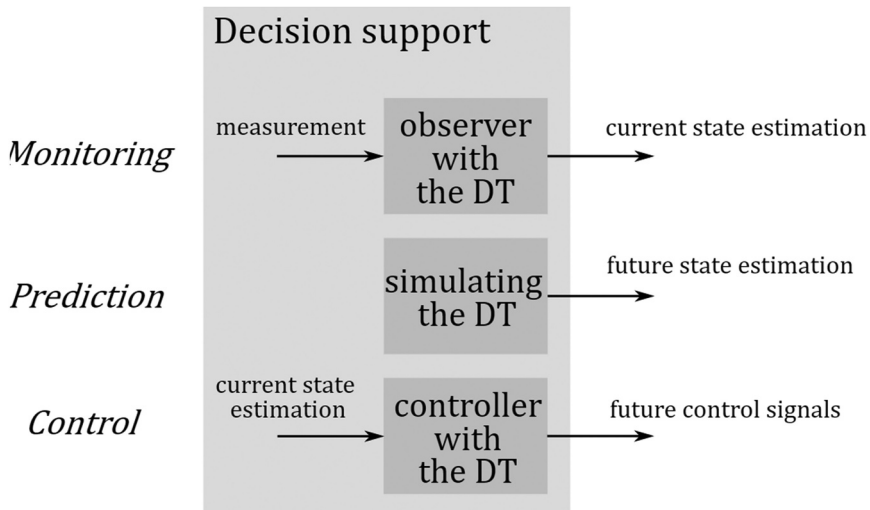


Figure 7. Role of the digital twin in decision support. Monitoring: the current state is estimated with an observer algorithm that employs the digital twin, together with sensor measurements. Prediction: the digital twin predicts the evolution of the state over a given time horizon. Control: using the current state estimation, together with the predictions of the digital twin, a control algorithm calculates the future control signals that optimize the operational management loop.

First, it needs to be determined exactly what system information needs to be monitored. Some system parameters cannot yet practically be measured in real-time (e.g. plant sugar levels) or require expensive sensing equipment. It is of interest to know what information needs to be monitored to improve the performance of decision support, and which information can be discarded, before e.g. investing in sensing equipment.

Then, the influence on system performance needs to be determined of measurement and actuation errors, their propagation through the controlled system, their interaction, and their influence on the accuracy of state and parameter estimations, and model predictions. This not only relates to the desired accuracy of sensing equipment (as mentioned above) but also to the level of detail needed in model calculations.

The convergence of estimated parameter values to the true values, its dependence on sensor choice, and measurement accuracy needs to be determined. In other words, how can we evaluate the correctness of a state estimation and how efficient is such a procedure? Related but different is the question what the benefits are of precise actuation (e.g. designing a lighting and pruning control strategy per individual plant) compared to conventional control (each plant receives the same amount of artificial light, and all plants have the same leaf area which is kept constant over time).

Continuing this line of thought, it needs to be clarified what the relation is between model spatial resolution, performance, and accuracy of predictions. In other words, how much would a high-resolution model such as a 3D crop model that predicts individual plant development (Vos et al., 2010), or a spatially distributed climate model (Reichrath & Davies, 2002; Saberian & Sajadiye, 2019), improve performance over a basic 1D model (such as a canopy crop model, or a spatially homogeneous climate model)?

Lastly, we note that it needs to be determined how the different types of high-level control (Figure 7) will perform. Although previous studies on model predictive control have indicated huge potential in performance increase (Henten, 1994; Kuijpers et al., 2021; Van Beveren et al., 2015; Van Ooteghem, 2010), most studies are not comparative, and involve predictive controllers directly controlling the actuators. This does not reflect the control loop encountered in practice and carries the risk that mispredictions could easily result in violations of bounds and jeopardize crop quality (e.g. incorrect weather predictions leading to too high or low indoor temperature). Much is still unknown about the performance of model predictive control in practice and about the tradeoffs between different control strategies (optimal, robust or risk sensitive).

Such questions can be answered through extensive simulation studies on a system that includes a controller based on a digital twin, algorithms for estimating states and parameter values, and a control algorithm. The outcomes can subsequently be validated with an experimental setup in a real greenhouse.

5.2.4. Challenges in controlled systems design

Through hands-on simulations and literature study, we have explored the challenges of designing a digital twin-based control system, consisting of a high-resolution functional-structural plant (FSP) model (Vos et al., 2010) combined with a greenhouse climate model (Vanthoor et al., 2011), contemporary algorithms for monitoring (particle filter), parameter estimation and model predictive control for the high-level control loop (see, Figure 6). The following challenges were identified, per type of decision support:

Monitoring: whereas a detailed model has the potential for accurate predictions, too much complexity may present a bottleneck as parameters may become interdependent and/or very sensitive to noise, leading to highly uncertain predictions. A similar challenge exists for state estimation. The issue of determining optimal model complexity is known as the bias-variance tradeoff (Hastie et al., 2009).

Prediction: computational demand may pose a burden here. Uncertainty analysis related to the research question on error propagation will easily require thousands of runs to simulate nonlinear parameter uncertainty propagation through sampling (Van Mourik et al., 2014). Inference of parameters may require hundreds to thousands of model integrations per update. For research practice, dozens to hundreds of trial runs for method development

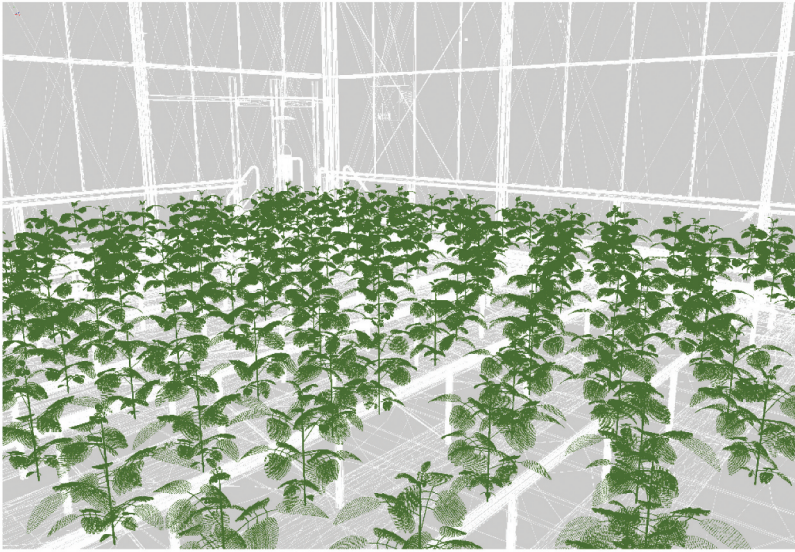


Figure 8. Simulated tomato plant growth in a virtual greenhouse environment. Visualisation created by Katarina Streit.

and debugging are required. Further, prior information about credible parameter ranges is indispensable. Otherwise, unrealistic values might be obtained that could jeopardize prediction reliability. Even after extensive literature search, or with tailored experiments, it is very likely that not all ranges can be determined.

Control: in the plant model, the number of states changes over time due to development of new stems, leaves and fruits (see, [Figure 8](#)). Conventional control algorithms are designed for systems with fixed numbers of states. Possible solution strategies exist, in the form of hybrid systems control, however issues about stability and optimization remain (Camacho et al., 2009). Another possible solution is a controller design based on a fixed number of summarized states, e.g. leaf area index and weight of organs and fruits (Vanthoor et al., 2011), which is made up of multiple other states. As an additional challenge, the solution space of the control problem grows exponentially with the number of states (Van Mourik et al., 2021). This makes conventional optimization such as dynamic programming infeasible for models with many states.

5.2.4.1. Recommendations for obtaining a digital twin with suitable model complexity. To find solutions for the challenges mentioned above, we recommend an emulation approach, with a controlled system containing a simplified model as a starting point, and only a low-level controller ([Figure 6](#)). The dynamics and performance of this basic system can be compared with that of a high-resolution system that serves as a proxy for

the real system. An iterative procedure, in which the outcomes of the simplified and high-resolution system are compared, can be used to determine which processes are essential and thus should be added to the simplified model to achieve the same system behaviour within an acceptable margin. After a model of suitable complexity has been selected or constructed, it should be assessed whether this model is suitable for monitoring, parameter estimation, and control design. Should the model then still be too complex, some techniques might offer a solution.

First, alternative AI optimization methods, such as reinforcement learning (Hemming et al., 2019; Zhang et al., 2021) to solve an MPC control problem within reasonable time. And second, model approximation techniques to reduce model complexity while preserving the essential behaviour. Examples are (piecewise) linearization, timescale decomposition (Van Straten et al., 2010), transfer function approximation (Van Mourik, 2008), regularization methods (Hastie et al., 2009), and likelihood approximation (Van Mourik et al., 2014).

5.2.5. Summary

Digital twins provide a promising extension to decision support for operational management of greenhouses, by combining high-resolution models with real-time data streams and algorithms for estimation and control. We argue that a high-resolution control system *in silico* offers the opportunity to provide answers to fundamental research questions that are of immediate interest to the greenhouse industry. However, the complexity of a high-resolution model introduces several serious challenges, that prohibit straightforward application of estimation and control algorithms required for automated decision support. To balance model complexity with respect to feasibility and resolution, we recommend to further explore emulation, AI-based optimization, and model approximation.

5.3. Case Me, My diet and I

5.3.1. Background

The world-wide increase in development of chronic diseases and the parallel rise in healthcare cost urgently calls for development of personalized smart healthcare tools to improve personal health and to reduce the pressure on the healthcare system. Many cases of chronic disease, such as type 2 Diabetes Mellitus and cardiovascular disease, are related to lifestyle. An effective way to prevent and improve these diseases is via dietary interventions. Dietary interventions and responses, however, are not equally effective for everyone (Berry et al., 2020; De Caterina et al., 2020; O'Connor et al., 2014). Although nutritionists have long been aware that what works for one person may not work for another, nutritional

advice is still given at a population-level via general nutritional guidelines reliant on the group mean (one-size-fits-all). To improve the effectiveness of dietary intervention there is an urgent need for more tailored precision-based personalized dietary advice. However, knowledge as well as technology on tailored precision-based dietary interventions is lacking.

Recent studies using continuous glucose monitoring have demonstrated significant individual differences in short-term post-meal responses. This provides a better understanding of individual differences in long-term dietary intervention responses. High variability and sometimes even opposite glucose (sugar) responses between subjects to the same foods were found (Zeevi et al., 2015). In that study, based on continuous glucose monitoring of 800 persons, an algorithm for predicting the personal post-meal glucose response to foods was developed, based on individual characteristics including the microbiome.

Post-meal lipid responses play a similarly important role in the development of metabolic diseases as post-meal glucose responses. A high postprandial lipid response is associated with an increased risk for development of cardiovascular diseases. From several studies performed so far, we know that people respond with a different blood triglyceride (fat) response to the same meal. The PREDICT study (Mazidi et al., 2021) has shown variations in postprandial glucose and triglyceride responses in 1000 subjects, including (real) twins, and revealed main factors associated with the variation in these responses. Effects of food intake have extensively been investigated (Bonham et al., 2019; Lee et al., 2020).

The flagship project “Me, My Diet and I” develops a digital twin for personalized dietary advice. This advice aims to reduce long-term personal post-meal triglyceride and glucose responses in a healthy overweight middle-age population. In particular, the aim is to reduce the amplitudes of the fluctuations to remain within healthy bounds, by enabling users to adapt their pattern, composition, and quantities of food intake for this purpose. For such a twin to be effective, the digital twin should be able to a) predict the personal postprandial triglyceride response and b) predict the best diet to reduce this triglyceride response on the long run. Both predictions will preferably be based on easily measured personal characteristics such as BMI, age, fat distribution, cardiometabolic blood markers, physical activity, and dietary intake.

5.3.2. Methods

In this case study, we adopted a development approach, addressing the various components required to provide personalized dietary advice. We included first results of a compliance study given the particular challenges related to personal health information.

5.3.3. Development approach

This digital twin is a representation of an individual with relevant personal traits to predict the postprandial triglyceride response, and to provide personalized dietary advice to reduce this postprandial response in the long run. Compliance to the advice is essential for the effectiveness of the twin, and this has shown to be dependent on personal characteristics (Dijksterhuis et al., 2021). We include therefore personal information such as behaviour, preferences, and values in the twin to be able to adapt the advice to individual preferences. The effectiveness of personalized advice requires sufficient information. It has been shown (Celis-Morales et al., 2015) that using a limited set of genotypes and blood biomarkers does not result in more compliance to dietary advice. We expect benefits from our approach based on a digital twin because of the integration of all relevant data into a digital twin, so that effective personalized advice can be given.

The simulation capabilities of the twin will be used to predict the effect of changes in diet on the long run. This can only be done with existing or new data of long-term dietary intervention studies that include triglyceride and glucose responses upon a postprandial test before and after the intervention. In principle, this can be accomplished with computational modelling, based on existing knowledge on triglyceride and glucose uptake and clearance. Examples of this are computational models such as physiology-based mathematical models (PBMMs) that capture mathematical representations of key metabolic processes (Erdős et al., 2021). Additionally, data driven approaches can be used: several methods (e.g. random forest or gradient boosting) could be used to build prediction models based on available input data. Predictions from the mechanistic model could be used as part of input to an incremental learning algorithm that adapts the output to the specific person. The complexity of the learning algorithm will be limited by the amount of data gathered for that person. We envision a phased development to investigate a combination of both approaches.

The building of a rudimentary version of the twin is the first phase in our approach. Once this twin is in place, it can be applied, validated, and sequentially optimized, which constitutes the second phase in our approach. For this phase, regular real-time data collection is crucial and although continuous triglyceride monitoring sensors are not yet available, other measures such as continuous glucose monitoring, physical activity and regular dietary intake using a dietary tracking app will be part of the twin. To monitor long-term effectiveness of the personalized dietary advice on postprandial triglycerides responses finger prick and measurements of triglycerides in dry blood spots may be used. We envision self-learning methods with this information to optimize the personalized dietary advice. The actionable knowledge of our twin is the personalized dietary advice based on phenotype of a person, which cannot be obtained without the twin.

The advantage of a digital twin instead is that the advice is generated and given in an automatic way without interference of a health professional such as a dietician. This eliminates barriers people may experience before contacting health professionals. This does not imply that it cannot be used by health professionals; it gives healthcare providers a unique look into their clients' personal metabolic situation which helps to further personalize their therapy. This also illustrates the need for and importance of knowledge input and expertise of experts such as dieticians and behaviour and communication professionals in the digital twin to prevent undesired, not compliant, and unhealthy personalized dietary advice.

5.3.4. Compliance survey

It is of fundamental importance for our approach that users accept and adopt personalized advice. To identify users' views on benefits and risks of following advice from a digital twin, we performed small-scale explorative interviews. Here, we share the preliminary qualitative results of this brief study for illustrative purposes and to indicate directions for further research. This will be necessary to establish more quantitative results and so further guide the development of the digital twin.

The interviews indicated that physical and mental barriers may play an important role in a user's decision on following the advice. Some people experience measuring and collecting personal biological data as a physical burden (e.g. wearing a continuous glucose metre). In addition, there is fear of finding unintentional health outcomes due to the required measurements (medical by-catches). For some people, the decision to follow this advice depends on the type of organization managing the advice: a private (commercially motivated) or public (societally motivated) actor.

In general, we found that people stress the need to have control on stakeholders – GP, dietician, health insurer, scientific researcher, employer, supermarket, etc. – that may have access to the personal information used for the advice. We found a dividing line for access of the information between on the one hand healthcare professionals (dieticians, doctors) and scientists (publicly funded researchers), and on the other hand private parties such as employers, insurers, supermarkets, food manufacturers, where the first group is more easily granted access than the second. In our interpretation, the first group apparently more easily inspires trust than the second group. Legal requirements and limitations, such as formulated in the EU's General Data Protection Regulation (GDPR), already provide guarantees that prevent misuse of personal information. Our findings at least support the public need for such guarantees in developing personalized dietary advice. At the same time, such regulations pose strict requirements on the development of such advice.

Furthermore, we found that most respondents insist that this advice should not be obligatory by the government or any (insurance) company.

A strong advice from a GP would on the other hand positively influence several respondents to follow the advice. Finally, and rather evidently, we found that the provided personalized dietary advice must be correct, reliable, and effective to be followed over longer periods of time.

5.3.5. Summary

We have described the rationale behind personalized dietary advice and the advantages expected from adopting digital twins for composing such advice. We have outlined several steps to realize this type of digital twin. Based on a small explorative study, we have indicated several nontechnical aspects that require attention for a successful launch of personal dietary advice based on digital twins. Our research continues to further develop this promising concept.

6. Conclusion

We investigated the potential benefits and challenges of digital twins for quantitative research on complex living systems in the green life sciences. Specifically, we compared the use of digital twins as alternative representations to simulation models, automatic control systems and forms of scientific advice:

The Digital Future Farm use case illustrates the potential for improved accuracy in predicting yield and growth of potatoes compared to traditional individual simulation models. The choice of inputs, acquisition frequencies and measurement accuracies as well as additional data sources influence this improvement in accuracy and requires further research. Filtering algorithms to obtain consistent temporal input given the occurrence of measurement inaccuracies are essential. The technological readiness for this use case is demonstrated by connecting to real-time hydrological, meteorological, and crop information data sources. It is precisely this connection to real-time data of digital twins that improves the utility of actionable knowledge over simulation models.

The Virtual Tomato Crop use case shows how digital twins can extend the capabilities of decision support for operational greenhouse management by considering individual plant characteristics and behaviour to provide detailed actionable knowledge. Integrating digital twins in greenhouse control provides additional benefits by providing more extensive information on temporal behaviour of greenhouse crops. We distinguish three types of decision support, with different requirements: state estimation for monitoring control, state updating for predictive control, and automatic calculation of control scenarios for decision support. Remaining research questions are how to determine and select data sources for information on plants and greenhouse climate, and what level of detail should be chosen for simulations and user interaction. We propose an emulation approach with a simplified model in to answer these questions.

Table 1. Summarized findings of the three case studies.

Case	Main benefits	Practical challenges	Further research
Digital Future Farm	Improved prediction accuracy by in real-time observed state updates	Real-time data connection, assimilation of observed and simulated data	Measurement quantities, frequencies, and accuracies, prediction horizons
Virtual Tomato Crop	More detailed actionable knowledge from more complete observational data	Input data selection, assimilation of observed data with control state information	Iterative emulation approach with additional AI-based optimizations
Me, My Diet & I	Increased impact of dietary advice for improved individual health	Availability of real-time personal data on health and behaviour	Receptiveness and compliance to personalized advice

Finally, the use case Me, My Diet and I illustrate how digital twins can increase the impact of professional advice, specifically personalized advice to keep postprandial lipid responses within healthy bounds. The use of digital twins only improves impact if the advice is sufficiently reliable and accurate for individuals, which requires extensive personal information. Obtaining this is challenging, for technical reasons as well as legal and social limitations. Compliance to the advice is an equally important factor determining impact. With an exploratory survey we identify, factors for compliance relevant for further research, such as personal convictions of users, stakeholders having access to personal information, and context of the advice.

The combined and summarized findings of the three case studies are shown in Table 1. Further research topics broadly fall into four directions, three of which are indicated in the Table

The first considers availability, accuracy, and frequency of sensor data on the input side of digital twins. This was most challenging for the Digital Future Farm, due to its particular approach (providing real-time sensor data to simulation models), but is similarly applicable to other cases. This direction of research is of fundamental importance for the development of digital twins, as the quality of their produced actionable knowledge is directly dependendt on fulfilling input requirements.

The second considers (temporary or permanent) emulations for the simulation part of digital twins. Given the integration of the digital twin in control systems, this was met most strongly in the case of the Virtual Tomato Crop, but is also relevant in other cases. This direction of research provides interesting opportunities for combining data-driven methods with process-based modelling in the search of computational efficiency, which may be required to obtain digital twins responses within practical time frames.

The third considers the reception of and compliance with the advice as provided by digital twins. Evidently, this followed from the investigations in

case of Me, My Diet, but applies to other cases as well. Although each particular digital twin may require a specific approach for investigating reception and compliance, ultimately, its societal benefit will largely depend on it. Therefore, we see this as equally fundamental as the other two directions of future research.

It has recently been noted in an engineering context that an overabundance of data may well lead to efficiency bottlenecks (Savage, 2022). The author further remarks that data do not come for free, implying that even if there are no limits of principle to increase the number of sensors providing more detailed information to digital twins, there are practical limits to this number, such as budget limitations. In the green life sciences, we also recognize the fundamental importance of accuracy and availability of data for digital twins but expect a lack of data rather than overabundance. For instance, manual registrations (of operations on farms and greenhouses) and privacy considerations (for personal medical details) may even require alternative approaches to compensate for non-available information. This can also be seen reflected in Table 1, where availability of data is a recurring practical challenge. At the same time, due to the complex behaviour of living objects and processes, we expect that data requirements in this domain will exceed those in engineering domains, for example, as discussed for the use case Me, My Diet and I which requires extensive data driven modelling to obtain sufficiently accurate simulations. We consider finding solutions for the relative scarcity of data in the green life sciences as the fourth direction of future research.

Given the main benefits, practical challenges, and further research listed in Table 1, it seems too early to draw conclusions regarding the societal relevance of digital twins. We expect digital twins to become an increasingly pervasive tool in the green life sciences. In each of the cases, we investigated, a fully operational digital twin will have consequences – on farming, greenhouse management, and the provision of dietary advice. It is important to have an agenda for responsible research and innovation regarding these consequences (Van der Burg et al., 2021).

In conclusion, we find promising benefits of digital twins in the domain of green life sciences. We illustrate this in three use cases, but obviously many similar use cases can be developed, in animal, environmental, food, plant, and social sciences. We expect that further research will include data driven modelling to simulate the complex character of living objects and processes and to develop approaches to limit the amount of required input data.

Acknowledgements

We gratefully acknowledge financial support provided by the Dutch Ministry of Agriculture, Nature, and Food Quality, and the Executive Board of Wageningen University & Research. We express our gratitude to the anonymous reviewers for their valuable comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Willem Jan Knibbe  <http://orcid.org/0000-0002-7177-9352>

References

- Alves, R. G., Souza, G., Maia, R. F., Tran, A. L. H., Kamienski, C., Soininen, J.-P., Aquino, P. T., & Lima, F. (2019). 'A digital twin for smart farming', in *2019 IEEE Global Humanitarian Technology Conference (GHTC). 2019 IEEE Global Humanitarian Technology Conference (GHTC)*, Seattle, WA, USA: IEEE, pp. 1–4. <https://doi.org/10.1109/GHTC46095.2019.9033075>
- Ariesen-Verschuur, N., Verdouw, C., & Tekinerdogan, B. (2022). Digital Twins in greenhouse horticulture: A review. *Computers and Electronics in Agriculture*, 199, 107183. <https://doi.org/10.1016/j.compag.2022.107183>
- Athanasiadis, I. N., Frysinger, S. P., Schimak, G., & Knibbe, W. J. (Red.). (2020). *Environmental Software Systems. Data Science in Action: 13th IFIP WG 5.11 International Symposium, ISESS 2020, Wageningen, The Netherlands*, February 5–7, 2020, Proceedings. Cham: Springer International Publishing (IFIP Advances in Information and Communication Technology). <https://doi.org/10.1007/978-3-030-39815-6>
- Bergs, T., Gierlings, S., Auerbach, T., Klink, A., Schraknepper, D., & Augspurger, T. (2021). The Concept of Digital Twin and Digital Shadow in Manufacturing. *Procedia CIRP*, 101, 81–84. <https://doi.org/10.1016/j.procir.2021.02.010>
- Berry, S. E., Valdes, A. M., Drew, D. A., Asnicar, F., Mazidi, M., Wolf, J., Capdevila, J., Hadjigeorgiou, G., Davies, R., Al Khatib, H., Bonnett, C., Ganesh, S., Bakker, E., Hart, D., Mangino, M., Merino, J., Linenberg, I., Wyatt, P., Ordovas, J. M., ... Spector, T. D. (2020). Human postprandial responses to food and potential for precision nutrition. *Nature Medicine*, 26(6), 964–973. <https://doi.org/10.1038/s41591-020-0934-0>
- Bonham, M. P., Kaias, E., Zimberg, I., Leung, G. K. W., Davis, R., Sletten, T. L., Windsor-Aubrey, H., & Huggins, C. E. (2019). Effect of Night Time Eating on Postprandial Triglyceride Metabolism in Healthy Adults: A Systematic Literature Review. *Journal of Biological Rhythms*, 34(2), 119–130. <https://doi.org/10.1177/0748730418824214>
- Camacho, E. F., Ramírez, D. R., Limón, D., de la Peña, D. M., & Álamo, T. (2009). 'Model Predictive Control techniques for Hybrid Systems', *IFAC Proceedings Volumes*, 42 (17), pp. 1–13. <https://doi.org/10.3182/20090916-3-ES-3003.00003>
- Celis-Morales, C., Livingstone, K. M., Marsaux, C. F. M., Forster, H., O'Donovan, C. B., Woolhead, C., Macready, A. L., Fallaize, R., Navas-Carretero, S., San-Cristobal, R., Kolossa, S., Hartwig, K., Tsirigoti, L., Lambrinou, C. P., Moschonis, G., Godlewska, M., Surwiłło, A., Grimaldi, K., Bouwman, J., ... Mathers, J. C. (2015). Design and baseline characteristics of the Food4Me study: A web-based randomised controlled trial of personalised nutrition in seven European countries. *Genes & Nutrition*, 10(1), 450. <https://doi.org/10.1007/s12263-014-0450-2>
- De Caterina, R., Martínez Hernández, J. A., & Kohlmeier, M. (2020). *Principles of nutrigenetics and nutrigenomics: Fundamentals of individualized nutrition*. Academic Press, eds.

- De Marke, PR - PRC, PPO/PRI AGRO Field Technology Innovations, Team Internationale Productie & Gewasinnovatie, Schröder, J. J., Hilhorst, G. J., Oenema, J., Verloop, J., & van den Berg, W. (2021). *BOFEK2020 - Bodemfysische schematisatie van Nederland: Update bodemfysische eenhedenkaart*. Wageningen Environmental Research. <https://doi.org/10.18174/541544>
- de Wit, A., Duveiller, G., & Defourny, P. (2012). Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agricultural and Forest Meteorology*, 164, 39–52. <https://doi.org/10.1016/j.agrformet.2012.04.011>
- de Wit, A. J. W., & van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1–2), 38–56. <https://doi.org/10.1016/j.agrformet.2007.05.004>
- Dijksterhuis, G. B., Bouwman, E. P., & Taufik, D. (2021). Personalized Nutrition Advice: Preferred Ways of Receiving Information Related to Psychological Characteristics. *Frontiers in Psychology*, 12, 575465. <https://doi.org/10.3389/fpsyg.2021.575465>
- Elijah, O., Rahim, S. K. A., Emmanuel, A. A., Salihu, Y. O., Usman, Z. G., & Jimoh, A. M. (2021). 'Enabling Smart Agriculture in Nigeria: Application of Digital-Twin Technology', in *2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)*. 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), Abuja, Nigeria: IEEE, pp. 1–6. <https://doi.org/10.1109/ICMEAS52683.2021.9692351>
- Erdős, B., van Sloun, B., Adriaens, M. E., O'Donovan, S. D., Langin, D., Astrup, A., Blaak, E. E., Arts, I. C. W., & van Riel, N. A. W. (2021). Personalized computational model quantifies heterogeneity in postprandial responses to oral glucose challenge. *PLOS Computational Biology* Edited by, J. Gallo. 17(3): e1008852. <https://doi.org/10.1371/journal.pcbi.1008852>
- Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343–367. <https://doi.org/10.1007/s10236-003-0036-9>
- 'Finding Answers Together Strategic Plan 2019-2022'. (2019). Wageningen University. (Accessed: 18 March 2022) <https://www.wur.nl/en/about-wur/strategic-plan.htm>
- Gebbers, R. (2018). Proximal soil surveying and monitoring techniques. In U. K. Silsoe Solutions & J. Stafford (Eds.), *Burleigh Dodds Series in Agricultural Science* (pp. 29–78). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2017.0032.01>
- Grieves, M. (2016). Origins of the Digital Twin Concept. *Unpublished*. <https://doi.org/10.13140/rg.2.2.26367.61609>
- Hameed, I. A. (2010). USING THE EXTENDED KALMAN FILTER TO IMPROVE THE EFFICIENCY OF GREENHOUSE CLIMATE CONTROL. *International Journal of Innovative Computing, Information and Control*, 6, 2671–2680.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*; Springer New York (Springer Series in Statistics). <https://doi.org/10.1007/978-0-387-84858-7>
- Heinen, M., Brouwer, F., Teuling, C., & Walvoort, D. J. J. (2021). *BOFEK2020 - Bodemfysische schematisatie van Nederland: Update bodemfysische eenhedenkaart*. Wageningen Environmental Research.
- Hemming, S., de Zwart, F., Elings, A., Righini, I., & Petropoulou, A. (2019). Remote Control of Greenhouse Vegetable Production with Artificial Intelligence—Greenhouse Climate, Irrigation, and Crop Production. *Sensors*, 19(8), 1807. <https://doi.org/10.3390/s19081807>
- Henten, E. J. V. (1994) *Greenhouse climate management: An optimal control approach*.

- Hoving, I. E., van Riel, J., Holshof, G., Plomp, M., Agricola, S., van Boheemen, K., & Roerink, G. (2019). *Schatten van grasopbrengst op basis van spectrale reflectie, grashoogte en modellering: Onderzoeksresultaten van een maaiproef op zand- klei en veengrond 2016-2017*. Wageningen Livestock Research. <https://doi.org/10.18174/508117>
- Jansen, D. M. (2008). 'Beschrijving van TIPSTAR: Hét simulatiemodel voor groei en productie van zetmeelaardappelen'. Plant Research International. <http://edepot.wur.nl/27135>
- Jansen, D. M., Davies, J. A., & Steenhuizen, J. W. (2003). 'Testen van Tipstar in de praktijk'. Plant Research International. <https://edepot.wur.nl/27135>
- Kamilaris, A., Wohlgemuth, V., Karatzas, K., & Athanasiadis, I. N. (Red.). (2021). *Advances and New Trends in Environmental Informatics: Digital Twins for Sustainability*. Springer International Publishing (Progress in IS). <https://doi.org/10.1007/978-3-030-61969-5>
- Kempenaar, C., Been, T., Booij, J., van Evert, F., Michielsen, J.-M., & Kocks, C. (2017). Advances in Variable Rate Technology Application in Potato in The Netherlands. *Potato Research*, 60(3–4), 295–305. <https://doi.org/10.1007/978-3-030-61969-5>
- Korenhof, P., Blok, V., & Kloppenburg, S. (2021). Steering Representations—Towards a Critical Understanding of Digital Twins. *Philosophy & Technology*[Preprint]. 1751–1773. <https://doi.org/10.1007/s13347-021-00484-1>
- Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11), 1016–1022. <https://doi.org/10.1016/j.ifacol.2018.08.474>
- Kuijpers, W. J. P., Katzin, D., van Mourik, S., Antunes, D. J., Hemming, S., & van de Molengraft, M. J. G. (2021). Lighting systems and strategies compared in an optimally controlled greenhouse. *Biosystems Engineering*, 202, 195–216. <https://doi.org/10.1016/j.biosystemseng.2020.12.006>
- Lee, D. P. S., Low, J. H. M., Chen, J. R., Zimmermann, D., Actis-Goretti, L., & Kim, J. E. (2020). The Influence of Different Foods and Food Ingredients on Acute Postprandial Triglyceride Response: A Systematic Literature Review and Meta-Analysis of Randomized Controlled Trials. *Advances in Nutrition*, 11(6), 1529–1543. <https://doi.org/10.1093/advances/nmaa074>
- Mazidi, M., Valdes, A. M., Ordovas, J. M., Hall, W. L., Pujol, J. C., Wolf, J., Hadjigeorgiou, G., Segata, N., Sattar, N., Koivula, R., Spector, T. D., Franks, P. W., & Berry, S. E. (2021). Meal-induced inflammation: Postprandial insights from the Personalised REsponses to Dietary Composition Trial (PREDICT) study in 1000 participants. *The American Journal of Clinical Nutrition*, 114(3), 1028–1038. <https://doi.org/10.1093/ajcn/nqab132>
- McKinion, J. M., Baker, D. N., Whisler, F. D., & Lambert, J. R. (1989). Application of the GOSSYM/COMAX system to cotton crop management. *Agricultural Systems*, 31(1), 55–65. [https://doi.org/10.1016/0308-521X\(89\)90012-7](https://doi.org/10.1016/0308-521X(89)90012-7)
- Metcalfe, B., Boshuizen, H., Bulens, J., & Koehorst, J. (2022). 'A hierarchy of Digital Twin maturity for the life sciences', In preparation [Preprint].
- Nasirahmadi, A., & Hensel, O. (2022). Toward the Next Generation of Digitalization in Agriculture Based on Digital Twin Paradigm. *Sensors*, 22(2), 498. <https://doi.org/10.3390/s22020498>
- Neethirajan, S., & Kemp, B. (2021). Digital Twins in Livestock Farming. *Animals*, 11(4), 1008. <https://doi.org/10.3390/ani11041008>
- O'Connor, E. M., O'Herlihy, E. A., & O'Toole, P. W. (2014). 'Gut microbiota in older subjects: Variation, health consequences and dietary intervention prospects',

- Proceedings of the Nutrition Society*, 73(4), pp. 441–451. <https://doi.org/10.1017/S0029665114000597>
- Pylianidis, C., Osinga, S., & Athanasiadis, I. N. (2021). Introducing digital twins to agriculture. *Computers and Electronics in Agriculture*, 184, 105942. <https://doi.org/10.1016/j.compag.2020.105942>
- Raghu, M., & Schmidt, E. (2020). 'A Survey of Deep Learning for Scientific Discovery', *arXiv:2003.11755 [cs, stat]* [Preprint]. (Accessed: 5 August 2021) <http://arxiv.org/abs/2003.11755>
- Reichrath, S., & Davies, T. W. (2002). Using CFD to model the internal climate of greenhouses: Past, present and future. *Agronomie*, 22(1), 3–19. <https://doi.org/10.1051/agro:2001006>
- Rosen, R., von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About The Importance of Autonomy and Digital Twins for the Future of Manufacturing. *IFAC-PapersOnLine*, 48(3), 567–572. <https://doi.org/10.1016/j.ifacol.2015.06.141>
- Saberian, A., & Sajadiye, S. M. (2019). The effect of dynamic solar heat load on the greenhouse microclimate using CFD simulation. *Renewable Energy*, 138, 722–737. <https://doi.org/10.1016/j.renene.2019.01.108>
- Savage, N. (2022). Virtual duplicates. *Communications of the ACM*, 65(2), 14–16. <https://doi.org/10.1145/3503798>
- Schils, S., der Werf, V., Eekeren, V., Dixhoorn, V., & den Pol- Dasselaar, V. (2019). Amazing grazing: A public and private partnership to stimulate grazing practices in intensive dairy systems. *Sustainability*, 11(20), 5868. <https://doi.org/10.3390/su11205868>
- Sela, S., Woodbury, P. B., & van Es, H. M. (2018). Dynamic model-based N management reduces surplus nitrogen and improves the environmental performance of corn production. *Environmental Research Letters*, 13(5), 054010. <https://doi.org/10.1088/1748-9326/aab908>
- Speetjens, S. L., Stigter, J. D., & van Straten, G. (2009). Towards an adaptive model for greenhouse control. *Computers and Electronics in Agriculture*, 67(1–2), 1–8. <https://doi.org/10.1016/j.compag.2009.01.012>
- Tekinerdogan, B., & Verdouw, C. (2020). Systems architecture design pattern catalog for developing Digital Twins. *Sensors*, 20(18), 5103. <https://doi.org/10.3390/s20185103>
- Uhlenkamp, J.-F., Hauge, J. B., Broda, E., Lutjen, M., Freitag, M., & Thoben, K.-D. (2022). Digital twins: A maturity model for their classification and evaluation. *IEEE Access*, 10, 69605–69635. <https://doi.org/10.1109/ACCESS.2022.3186353>
- Van Beveren, P. J. M., Bontsema, J., Van Straten, G., & Van Henten, E. J. (2015). Minimal heating and cooling in a modern rose greenhouse. *Applied Energy*, 137, 97–109. <https://doi.org/10.1016/j.apenergy.2014.09.083>
- Van der Burg, S., Kloppenburg, S., Kok, E. J., & Van der Voort, M. (2021). Digital twins in agri-food: Societal and ethical themes and questions for further research. *NJAS: Impact in Agricultural and Life Sciences*, 93, 98–125. <https://doi.org/10.1080/27685241.2021.1989269>
- van Evert, F. K., Booij, R., Jukema, J. N., ten Berge, H. F. M., Uenk, D., Meurs, E. J. J. (., van Geel, W. C. A., Wijnholds, K. H., & Slabbekoorn, J. J. (2012). Using crop reflectance to determine sidedress N rate in potato saves N and maintains yield. *European Journal of Agronomy*, 43, 58–67. <https://doi.org/10.1016/j.eja.2012.05.005>
- van Evert, F. K., Berghuijs, H. N. C., Hoving, I. E., de Wit, A. J. W., & Been, T. H. (2021). '110. A digital twin for arable and dairy farming', in *Precision agriculture '21. 13th European Conference on Precision Agriculture*, Budapest, Hungary: Wageningen Academic Publishers, pp. 919–925. https://doi.org/10.3920/978-90-8686-916-9_110

- van Evert, F. K. (2019). Data for developing, testing, and applying crop and farm models. In *Burleigh Dodds Series in Agricultural Science*. Burleigh Dodds Science Publishing, 385–418. <https://doi.org/10.19103/AS.2019.0061.19>
- van Mourik, S. (2008). *Modelling and control of systems with flow*. PhD. University of Twente. <https://doi.org/10.3990/1.9789036526173>.
- van Mourik, S., ter Braak, C., Stigter, H., & Molenaar, J. (2014). Prediction uncertainty assessment of a systems biology model requires a sample of the full probability distribution of its parameters. *PeerJ*, 2, e433. <https://doi.org/10.7717/peerj.433>
- van Mourik, S., van Beveren, P. J. M., López-Cruz, I. L., & van Henten, E. J. (2019). Improving climate monitoring in greenhouse cultivation via model based filtering. *Biosystems Engineering*, 181, 40–51. <https://doi.org/10.1016/j.biosystemseng.2019.03.001>
- van Mourik, S., van der Tol, R., Linker, R., Reyes-Lastiri, D., Kootstra, G., Koerkamp, P. G., & van Henten, E. J. (2021). Introductory overview: Systems and control methods for operational management support in agricultural production systems. *Environmental Modelling & Software*, 139, 105031. <https://doi.org/10.1016/j.envsoft.2021.105031>
- Van Ooteghem, R. J. (2010). 'Optimal control design for a solar greenhouse', in *IFAC Proceedings Volumes*, pp. 304–309.
- van Straten, G., van Willigenburg, G., van Henten, E., & van Ooteghem, R. (2010). *Optimal control of greenhouse cultivation* (0 edn). CRC Press. <https://doi.org/10.1201/b10321>
- Vanthoor, B. H. E., Stanghellini, C., van Henten, E. J., & de Visser, P. H. B. (2011). A methodology for model-based greenhouse design: Part 1, a greenhouse climate model for a broad range of designs and climates. *Biosystems Engineering*, 110(4), 363–377. <https://doi.org/10.1016/j.biosystemseng.2011.06.001>
- Verdouw, C., Tekinerdogan, B., Beulens, A., & Wolfert, S. (2021). Digital twins in smart farming. *Agricultural Systems*, 189, 103046. <https://doi.org/10.1016/j.agsy.2020.103046>
- Verdouw, C. N., & Kruize, J. W. (2017). 'Digital twins in farm management: Illustrations from the FIWARE accelerators SmartAgriFood and Fractals', In. *7th Asian-Australasian Conference on Precision Agriculture Digital*, Hamilton, New Zealand, p. 6.
- Vos, J., Evers, J. B., Buck-Sorlin, G. H., Andrieu, B., Chelle, M., & de Visser, P. H. B. (2010). Functional–structural plant modelling: A new versatile tool in crop science. *Journal of Experimental Botany*, 61(8), 2101–2115. <https://doi.org/10.1093/jxb/erp345>
- Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J. A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., ... Segal, E. (2015). Personalized nutrition by prediction of glycemic responses. *Cell*, 163(5), 1079–1094. <https://doi.org/10.1016/j.cell.2015.11.001>
- Zhang, W., Cao, X., Yao, Y., An, Z., Xiao, X., & Luo, D. (2021). 'Robust model-based reinforcement learning for autonomous greenhouse control', *arXiv:2108.11645 [cs]* [Preprint]. (Accessed: 19 February 2022) <http://arxiv.org/abs/2108.11645>