

Assignment 2: Implementation of a Digital Twin in the Manufacturing Industry

Thomas Howe

Student ID: 2238761

Advanced Industry 4.0

ACPA7000

25/09/2023

Introduction

This essay explores the combination of several Industry 4.0 (I4.0) technologies with a view to mapping manufacturing processes through advanced modelling and digital twinning. It will also discuss how implementing a digital twin requires the integration of artificial intelligence, industrial internet of things, model-assisted Design of Experiments, RFID tags and big data analytics.

Currently, manufacturing systems can be 'running blind' and potentially sub-optimal. The amount of data that a process is communicating may be used in any number of ways and to varying degrees. And there may be situations, where useful data is not being sensed or captured, let alone utilised. This is where industry 4.0 (I4.0) can start to be implemented. Digital twin technology is an I4.0 technology that has been developed recently as an advanced form of simulation with 2-way feedback. It also relies on synergy with further I4.0 technologies to operate at maximum efficiency.

Manufacturing companies are therefore investing in digitization to accelerate growth. The International Data Corporation (IDC), published data in 2021 (1) showing that the global manufacturing industry is spending 13.1 percent on digitisation and smart manufacturing, forecast to reach \$2.8 trillion by 2025. (See figure 1)

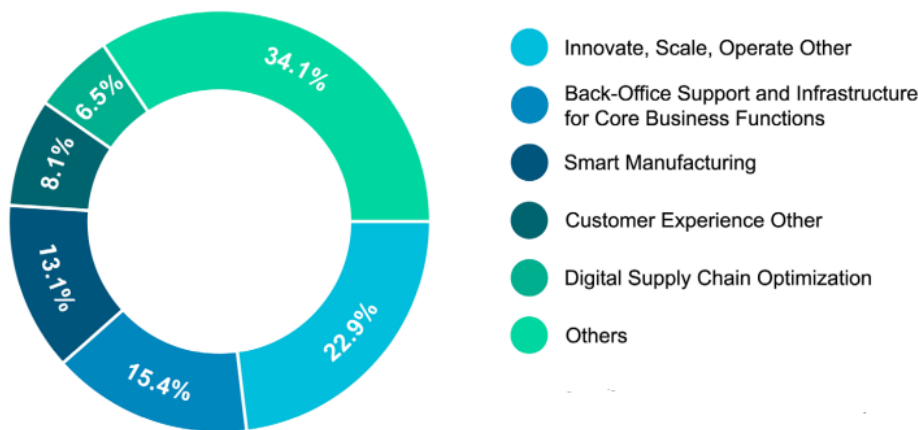


Figure 1. Spending trends of manufacturing global industry (from IDC digital transformation spending guide (1))

It is projected that digital twins can achieve substantial cost savings by reducing downtime and accurately predicting when maintenance needs to be carried out.



Figure 2. Outcomes enabled by digital twins (2)

Depending on the application, the industrial internet of things (IIOT), artificial intelligence (AI) and Machine learning (ML) will be used to get the most accurate possible model of the process. The IIOT is a system of sensors that are (usually) wirelessly linked to a central database or data warehouse. Sensors and human input generates data from the process which is often too numerous or complicated to be analysed by a human operator, or with simple equations. ML can be implemented to classify the data and start to learn trends and display relevant, critical data. The data is fed into a learning algorithm that starts to generate models, that can be thought of as a virtual process. This process can be purely digital or can have a graphic user interface (GUI) to aid visualisation.

Digital twins can spot trends and self-optimize: If normal variations in a process parameter affect the production rate, that parameter can be described as being a factor of production rate. Modelling may start to show that deliberate alterations in that factor can be used to deliberately influence the production rate. In traditional process optimisation, you may need to run a large number of experiments to establish this relationship. But the optimised model can be run in all kinds of conditions and scenarios to increase the likelihood that making those changes will indeed improve the throughput. Time and cost savings are self-evident.

Discussion

Defining the Digital twin

Digital twins are a system for modelling a process that is a digital representation of the physical process. According to ISO standard 23247-1:2021: (2) Digital twins monitor its observable manufacturing elements by constantly updating relevant operational and environmental data. Digital twins assist with detecting anomalies in manufacturing processes to achieve functional objectives such as real-time control, predictive maintenance, in-process adaptation, Big Data analytics, and machine learning. The novel factor of the digital twin (compared to traditional models) is that it is synchronous, with instant 2-way data transfer between the twin and the real process (3). It is useful for multifaceted processes that are too complex to be described in design equations.

This is a novel, state of the art version of traditional asynchronous simulation (digital modelling) because simulation doesn't have direct inputs and outputs that link to the physical process. A model is programmed with all data by the user, whereas a twin will take real time data from the digital version and transmit it to the physical process to control and augment its operation.

Also, by exception, a digital twin is not a digital shadow because in that case, data only flows from the physical object to the digital representation.

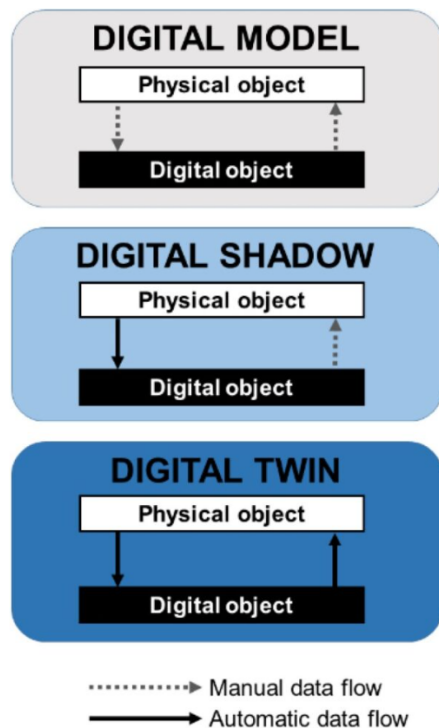


Figure 3. Difference in data flow between digital model, shadow and twin. (3)

As shown in Figure 3 above, digital modelling is the simplest phase of the digital twin and many software packages have existed for decades, that can model processes (e.g. Aspen Hysys/ Arena Discrete event simulation). These programmes can elucidate a variety of processes, from the optimal conditions for a chemical process to the flow of people through a shop. They can form a starting point for a digital twin, showing what needs to be measured and what equipment is required to do so. They can provide insights into superficial changes that can improve the process.

The digital shadow can then be created from the model, which takes data from the physical process through a range of sensors (and potentially some human input) to develop a more accurate digital version of the process.

The digital twin goes further, with full feedback and the ability to use the data generated from the digital version, to affect changes in the physical version. Iterative changes can then be made, with further feedback. The modelling component allows automatic alternate scenarios to be completed in the background and information used to optimise the real process. Giving 2 synchronous feedback loops: one simulated and one empirical.



Figure 4. Artistic rendering of a Digital twin of a jet engine (2)

Types of Digital twin

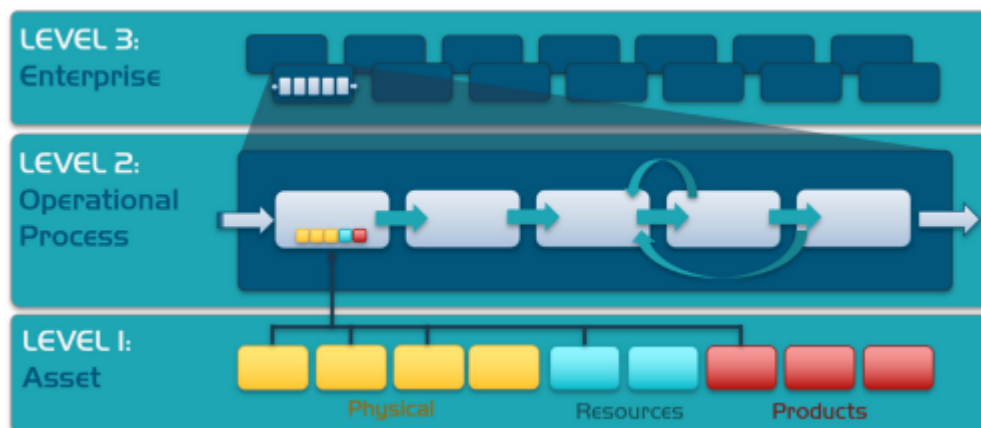


Figure 5. Types of digital twin (4)

Asset / Process digital twin

The stereotypical digital twin. It is primarily concerned with optimising Product life-cycle management at a machine level (4). It captures detailed data to visualise, simulate and analyse multiple facets of the asset. They mimic the machine (For example, a chemical reactor or production line) and allow visualisation of the real time data produced by sensors. They facilitate and enhance predictive maintenance, by analysing sensor data that precedes a breakdown and using it to diagnose further failures before they happen. Providing data for resourcing and procurement of critical spares. Managing resources, such as personnel, can also fall into asset DT, depending on the level of automation and human interaction.

Operational Process digital twin

Operational process integrates the overall operation of the production facility. They can be further subdivided into several modes.

Supervisory can be used for training and support. Ensure consistency of processes and reduce operational errors. And testing business scenarios and conceptual designs.

Diagnostic and control which monitor and diagnose problems in a dynamic way.

Predictive which use datalogs to model and predict. And can then be programmed with what if scenarios to generate scenarios for evaluation and facilitate better decision making.

Enterprise digital twins

A more abstract version of the operational DT, which incorporates information and modelling of people processes and technology. And their interdependencies. Working with a model of the business operations to translate the strategy of the business into a predictor for business outcomes and advice on choosing new initiatives.

Supply chain Digital twin

Supply chain is a process that can potential be modelled and twinned- sometimes called logistics 4.0. The following case study was investigated for an in depth summary of supply chain twinning: "Design and Development of Digital Twins: a Case Study in Supply Chains"

Currently, large scale manufacturing businesses would implement an enterprise resource planning (ERP) system to deal with increasingly complicated value chains. These can handle large amounts of data and multiple process streams. They are excellent tools for coordinating supplies, sharing data, and monitoring the flow of goods through their processes. However, the processes surrounding manufacturing operations are becoming ever-more complex, leading to new solutions being required. Mass customisation, for example, is an emerging technology, that requires processing flexibility and added complexity at every production unit. ERPs can be limited in terms of real time data and correspondingly slow responses to disruptions because they rely on human interaction for data transfer, simple decision making, analytics and optimisation. There are reports of them being targeted by cyberattacks.

Digital twins are capable of going beyond the ERP system, in terms of breadth of information, by existing as a real time model that incorporates all aspects of the financial and operational status (And has potential to model future events and recall events from its history). Furthermore, they can operate with more depth than ERP, with information on status and configuration of machines or assets, order status and incoming work. The decision-making aspects of the digital twin can schedule tasks in the most operationally efficient fashion, reducing lead times, increasing customer satisfaction and reducing load on the workforce/assets. The ability to anticipate workflow and react to disruptions is a further layer of control, that is normally a human task or decision, which often requires significant resource.

As previously mentioned, a Digital twin requires a range of sensors embedded in the system or process that it is modelling. Typically, reactors, process plants or so-called assets are equipped with a range of sensors that feedback information to a PLC or similar device. That means a large part of the work for a digital twin and the digitisation of the asset is already well underway. However, a supply chain or value chain is not typically so adorned. This implies a large up-front cost to digitalisation with a fairly long experimental commissioning phase. The gold standard in supply chain management is the pharmaceutical industry, and rightly so: Quality and Traceability are paramount with expensive chemicals. Medications are high value materials which are supplied to an extremely wide range of facilities. So they

must be well managed logistics and also cannot end up in the wrong place or misplaced. This industry is likely to be an early adopter and one to be emulated by others.

Radio frequency identification tag (RFID) is the major recently developed tool for monitoring the flow of materials through a value chain. They differ from traditional scanning. Barcodes or QR can only be scanned in a short range from one direction. But RFIDs are scannable from a distance and through other items, such that obscured and covered containers can be scanned.

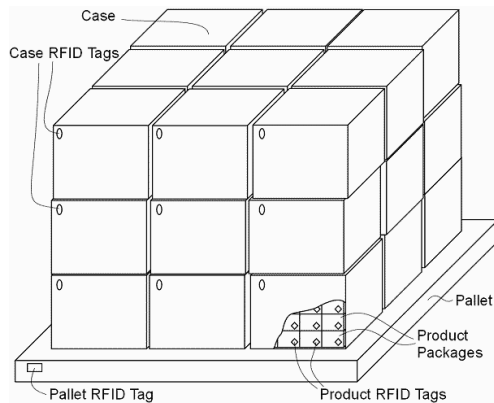


Figure 6. The possibility of reading RFID tags in locations where barcodes would be obscured, means the availability of an abundance of data on tracking, stock locations etc.

Also, movement of personnel can be modelled. Spaghetti diagrams are the traditional method for this and target the 'Muda' aspect of lean by reducing the waste of movement.

The recent, digital version of spaghetti diagrams, discrete-event simulation (DES) can map processes in considerably more detail. And, most importantly, provide scenario modelling. A well modelled process will give highly detailed information on resourcing, optimisation and reduction of latency and bottlenecks. For example, a workbench that is used sequentially by 4 operators may provide a bottleneck. Providing extra workbenches would allow concurrent operations. However, is this actually of benefit, if it exposes upstream bottlenecks or downstream inefficiencies? Could the bottleneck be scheduled out? This is where the software can be so much more advanced, by running scenarios to show what may actually happen if you get more benches or operators or shorten walkways, etc. All decision making can then be supported by relevant statistics and return on investments.

The digital twinning of the DES model can move beyond this, with real time event tracking, shop floor data and management of consumables (5). The multi-stakeholder environment of a warehouse/ shop floor environment requires fine control over many human and resource elements. The live, 2-way feedback of the digital twin allows real time scheduling and necessary rescheduling based on any process exceptions. Learning from these events and avoiding the events that lead to the exception.

Shifting demands and the agility of the digital twin

Manufacturing facilities have multiple needs. There is often a dichotomy between supplying a profitable product and constantly experimenting and optimising to increase throughput or reduce costs. The twin allows synchronous optimisation and production.

Some level of customisation is becoming normal in manufacturing, the days of mass production certainly aren't behind us, but the one-size-fits-all approach is no longer suitable for many customers and consumers.

Recent events have raised serious questions: What can be done when there is serious disruption? Such as during covid, with low staffing and supply and demand changes. Can process plants be made super agile with a higher degree of redundancy built in for unforeseen changes?

Implementation

The first step is inventory of assets and their connectivity. A semantic model is a good way to visualise this with nodes and connections representing the process. Typical examples of this include Process Flow Diagrams and asset hierarchy models.

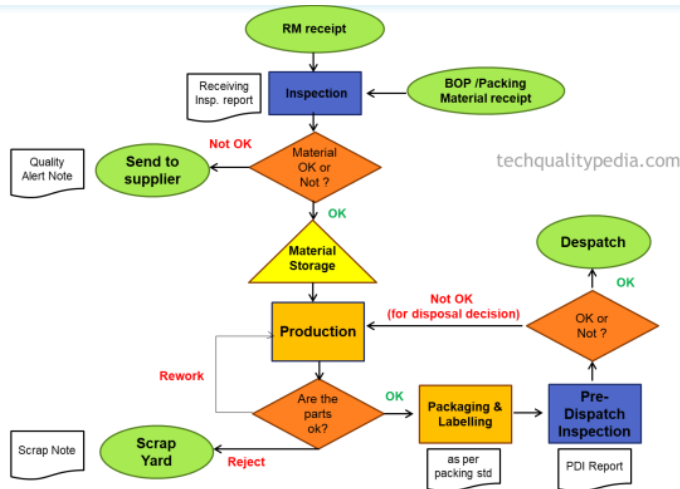


Figure 7. Typical process flow chart

The usefulness of any model relies on fidelity (6) which is a large task when you have a large, complicated system. Good modelling and process knowledge should allow the programming of a digital twin, that focuses on the key parameters, potentially simplifying the model and ignoring noise or unimportant input variables. Choosing these key aspects may require expert knowledge and a design of experiments that organises and emphasises the key input parameters and outputs. For example, the Pareto principle states that 80% of the consequences come from 20% of the causes and, if this is true for the given process, a large number of parameters could well be of little consequence. This quandary leads to the process of robustification, defined by Wikipedia as “the process of finding the inputs that contribute most to the random variability in the output and controlling them”. This can be done experimentally, by measuring the effect of the inputs on key parameters or analytically, using design equations or numerically, using simulation. As the amount of data rows, the model and the twin can start to take over data generation.

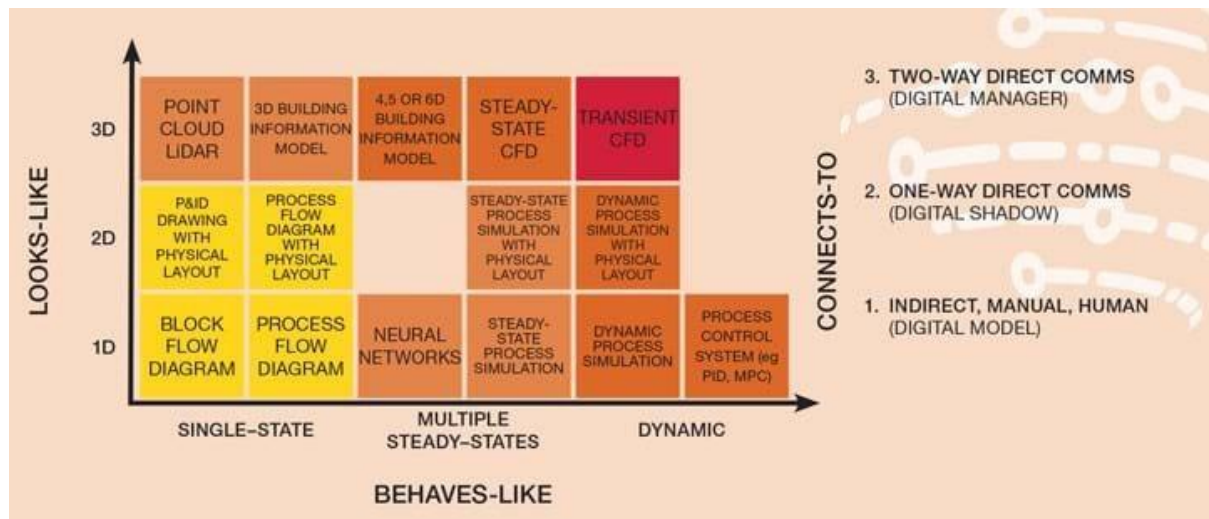


Figure 8. Digital model classification in the Chemical processing industry. (6)

Once the key parameters have been elucidated, the level of complexity of the model can be chosen. As shown in figure 8 above, there are many available levels of complexity and choosing the correct level of model and how dynamic it needs to be will prevent overspec'ing and correct cost-benefit.

Choosing the right model

Data management and programming are the next key steps to good models. A digital twin needs to be able to 'think' for itself. Data processing using AI is the I4.0 solution to this. Using data for interpolation or extrapolation and assessing previous actions for future outcomes can all be programmed.

The degree of operator interaction should be well defined: Incorporations with Artificial intelligence (data mining and deep analytics) and Machine learning isn't completely foolproof and monitoring/maintenance of fidelity will be required, by an engineer.

Types of Artificial intelligence

Artificial intelligence has become a catch-all for many different and distinct technologies. It is therefore important to define some of the AI available to a digital twin.

Datamining:

Merging of statistics and machine learning for extracting useable information (especially patterns and trends) from large datasets and transforming inputs into structured, relational datasets.

Artificial Neural Network (ANN)

Pathways of learning, trained by the weight of inputs: More frequent occurrences reinforce stronger pathways, so that reliable predictions can be made from future inputs.

Humble AI

There must inevitably be part of the digital twin that interacts and combines with the human operator to aid decision making. An example of interacting AI is named as humble AI, because it can

make simple decisions, but it is 'humble' enough to know its limitations and will defer to a human operator, when the decision falls outside a predetermined risk range.

Digital ghosts for cyber-attack detection.

An additional layer of security to the traditional firewalls and antivirus software. By scanning the operations of a network, system or asset, the 'ghost' learns what normal data transfers and interactions look like. In the event of an abnormal event, the ghost will flag potential threats to the human operator or shut down processes or operations to prevent further breaches. This technology was developed by General Electric for the continuation of power supply to critical systems. If a cyber-attack prevents the main system from operating properly, it reverts to operating data from the digital model. (E)

Cross communication AI.

The digital twin of an asset can be programmed to share its data with a similar or identical asset within the same company. This way, an asset can have knowledge of existing problems and potential fixes. Engineers working on breakdowns can have complete visibility of potential root causes, before they have even started diagnosing the problem.

Deep learning

Uses ANNs and further levels of learning to produce using several types of real world data

Types of data

Part of scoping a digital twin project is deciding what data is to be captured and this will lead to determining its uses and usefulness.

Structured

Well defined schemes, such as relational databases in a tabular format.

Semi structured

Has some structure, but is not stored in relational databases. HTML and XML are examples of this.

Unstructured

No defined format. Difficult to organise. Often contains text and multimedia

Metadata

Data about data: can be a small subset of data that describes a larger dataset.

Discussion of Machine learning types

Extracting insights of large datasets can be done more efficiently by using machine learning techniques.

Supervised uses sample input output pairings as training data for the algorithm. The algorithm can then infer a function and identify goals from the inputs. Care must be taken to give representative data samples. Classification and regression are the typical means to organise the data into meaningful outputs. Regression gives a fit; classification separates data into sets and subsets.

Unsupervised uses unlabelled datasets. It is a data driven process. Outputs are generative features, trends and structure, clustering, and association. This can reduce bias by potentially reducing the amount of chosen training data and focussing on more raw data.

Semi-supervised is a combination of the above. It can process both labelled and unlabelled data to provide better predictions than with labelled or unlabelled alone.

Reinforcement learning is based on reward or penalty and is trained to identify and seek rewards or identify and reduce risk. Which is quite abstract, but can be taken to mean moving towards or away from a defined target, such as a high throughput or a low number of defects.

Bias

Generally, it is accepted that all programmes will be biased by the programmers experience (or lack thereof) and possibly what outcomes are expected by the programmer. After all, is it possible to programme a scenario that hasn't been encountered before in reality, by the programmer? Further to the programme itself being skewed, the selection of training data given to an ML algorithm could also cause it to model things incorrectly.

Types of analysis

The main goal of analysis is to start to see trends in the data. This could take the form of a classification, which- in the manufacturing industry- could photograph a component and scan it for features that are abnormal, reducing the need for a human inspection of parts. Or a regression which would model the interactions of input parameters. For example, combining a higher temperature and a lower pressure in a reaction vessel may lead to a faster throughput of a chemical product.

Classification analysis

Supervised learning involving sorting information based on different data labels. It can be relatively simple, with binary classification or multiclass classification. Typically, manufacturing processes will be multilabel and contain several classes and labels. Classes will be hierarchical and labels can span multiple classes, so the algorithm is tasked with sorting mutually non-exclusive classes and labels. Some examples of algorithms are "Naïve Bayes" and "decision tree". In the manufacturing industry, this can be used to start to group together information about your product quality and cost and throughput and start to see trends that can be measures of success and routes to optimisation. A digital twin will start to flag these things automatically and potentially implement improvements with little or no interaction from human operators. An example for Chen et al (7) which used a camera linked to ML classification algorithms to develop a data-driven method for enabling automatic inspection of integrated circuit (IC) wire bonding defects.

Regression analysis

Different from classification because classification predicts distinct class labels. Whereas regression shows a continuous quantity, such as a trend or a relationship between variable x and y . This is familiar to most scientists, as it is a way of defining correlations and dependency of one variable upon another (8). The typical higher level algorithms are also quite well understood, for example polynomial or logarithmic. The difference with AI driven analysis, is the ability to spot trends that

may be invisible to human operators, due to their complexity or they are buried in very large datasets. A digital twin can 'see' these trends, work out relational variables and start to make predictions about optimal operating conditions. For example, in chemical manufacturing, there can be complicated rate equations, which change over time and with changes in feed stock. These multivariate problems can be hard to model and spotting trends can be laborious. Design of experiments could be suggested by the digital twin, or existing data could be interpreted for multivariate analysis (9)

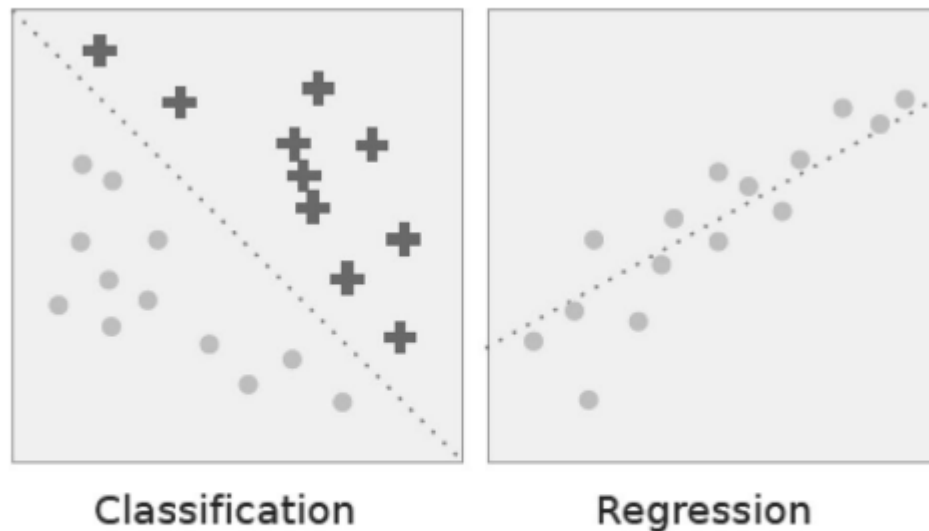


Figure 9. Classification vs regression. In classification the dotted line represents a boundary between 2 classes. Regression shows the relationship between 2 variables

Data driven virtual optimisation.

Using the digital twin to develop operational excellence, increased operational insight and deeper understanding of the capability of systems. Quite often these simulation software already exist in chemical process manufacturing. These are often models of the plant, with examples of software that include Hysis they can be deployed for operator training, plant optimisation and remote support. And sometimes referred to as a plant twin to differentiate from workflow or supply chain twins.

Design of experiments (DOE) is something that is typically carried out in the physical world (9). With parameters chosen and modified to learn something critical and useful about the process. As a physical process, it is normally carried out one factor at a time (OFAT), which is essentially the guiding principle of the scientific method: Change one input variable and measure the system response. Otherwise: change more than input variable and you will not know what has elicited the system response. However, large, complicated systems often have huge economic gains to be made by optimisation and changing one thing can be time-consuming, especially if there is interruption to normal production. Interrogating the wrong input variables can also be fruitless. Sometimes processes are running continuously and cannot be stopped or altered for experimenting. This is where a digital twin can aid with optimisation. A well-modelled system can be used to run virtual design of experiments, which highlight parameter optimisations without affecting the physical

process. This is known as model-assisted design of experiments (mDoE) (10). Data gathered in this way can then be used to propose physical DoEs.

It is a standard task for a process engineer to screen for important input variables by using full factorial DoEs and partial factorial DoEs to reduce the number of physical experiments, but with complex systems, there will come a point, where the mathematics becomes overly complicated. Shown as an example in Figure 10 that simple linear models can be relatively straightforward for a human to describe mathematically, whereas most people would struggle to fit data to a partial cubic model.

A correctly-programmed digital twin has the capability to fit complex algorithms by analysing multivariate, incomplete and noisy data (10). This deep level knowledge can further decrease the number of physical experiments and ultimately save time and costs. With the distinct possibility, that a true optimum process can be derived, which simpler models would never have elucidated.

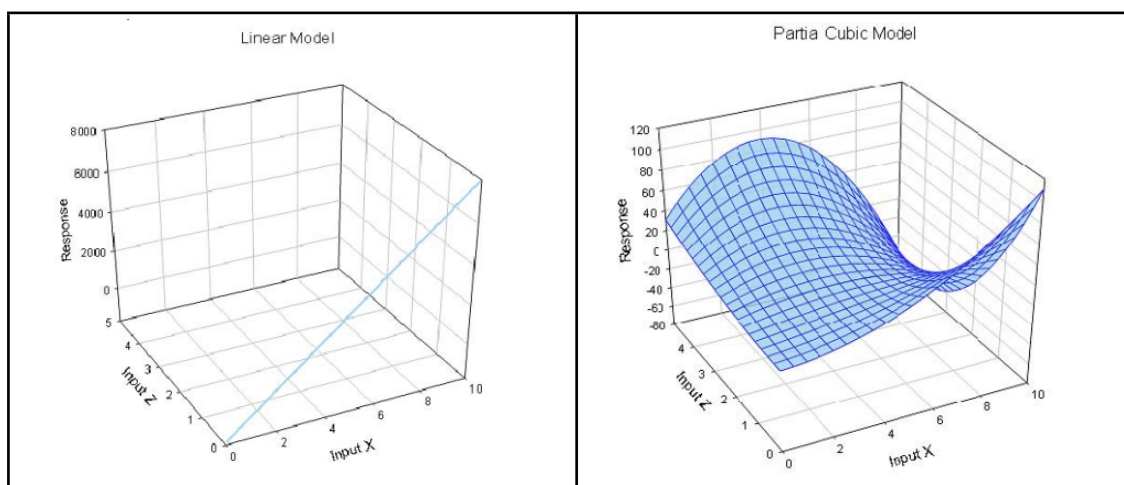


Figure 10. Extremes of model complexity (F)

The secondary (but maybe equally important) part of computer modelling is the ability to reduce errors (standard deviation) with the ability to run the virtual experiments ad infinitum. Outputs generated can form feedback loops with the machine learning algorithms and adapt after each iteration to trial further experiments that fill gaps in incomplete data. i.e. if changing parameter A is shown to increase throughput, ML asks what happens if we vary parameter A within a certain range.

Testing the model

Model verification is the next major step in the cycle. As with Plan, Do, Check, Act, the model must be checked to ensure it is a true representation of the physical system. This is where the physical system is changed in the fashion suggested by the modelling and the expected response is analysed. IIOT comes back into its own here, generating the output parameters to be analysed. Assuming everything operates as expected, the model can be accepted and if not, further modification to the model may well be required. It is important to check the model with various scenarios, before implementing the model in production of final product. These check points verify the models proximity to reality and should be carried out at near-optimum conditions and near suspicious behaviour (where the model may appear to be breaking down). Furthermore, check points are within the range of the experimental parameters that produced the model and within typical operational parameters. Again, all data can be fed back to ML for further learning. Hierarchically, this process could be aided by human operators at first, but once the process of iteratively improving through DoE has been learned by the digital twin, it could run continuous loops of virtual models followed by physical trials, followed by further modelling and trials.

The State of the Art

Global household names such as BP, Phillips, IBM, DHL and BP have some forms of digital twins that are in various stages of development and vary slightly in their definitions (11). However, there is still a gap between what is ideal and what is possible, usually due to cost and resources. Some digital twins do not possess artificial intelligence, for example. But the price of Industry 4.0 technology continues to fall, such that digital twins will be a normal part of process control in the near future.

Conclusion

Effectively creating a digital twin is a complicated and long process, with some business leaders likely to ask what is the necessity for such a project. It seems like a 'minefield': What is the best model? Can the model be true or close enough to reality to be reliable? The necessity to programme a suitable framework for the twin is likely to be quite complicated. Choosing the right data to be analysed and a machine learning algorithm that effectively manages data and gives the correct responses. Then relying on the twin to effectively communicate back to the asset or process in order to optimise and improve it. This has led to slow uptake, after the term digital twin was coined, more than a decade ago.

But the value comes from the ability to optimise on a near-constant timescale. The organisation of assets, production planning and value chain mapping can largely be taken out of the hands of humans, with predictive and preventative management of process exceptions. There is effectively always a plan B, C and D. Predictive maintenance gives control over the number and length of breakdowns. No longer would a part break to then find it is the only one on site, out of stock with the manufacturer and has a 10 week lead time: This would all be managed by the digital twin. In short, the complexity of processes is increasing, with the number of stakeholders and 'moving parts'. This drives up inefficiencies and the potential for processing exceptions. The level of communication and interconnectivity of a well-specced digital twin cannot be matched by humans, no matter how well-drilled the team is. Some level of twinning is therefore necessary, eventually, for all forward-thinking manufacturing facilities.

References

1. IDC (2021) IDC Spending Guide. Available at: <https://www.idc.com/getdoc.jsp?containerId=prUS49797222> (Accessed: 22/9/23).
2. (2021) Automation systems and integration — Digital twin framework for manufacturing — Part 1: Overview and general principles, ISO 23247-1:2021 Available at URL : <https://www.iso.org/standard/75066.html>
3. Shepard J (2022) What's a digital shadow and how does it relate to a digital twin? Available at: <https://www.analogictips.com/whats-a-digital-shadow-and-how-does-it-relate-to-a-digital-twin-faq/> (Accessed: 22/9/23).
4. Aitken, A, (2019) Industry 4.0: Demystifying Digital Twins. Lanner
5. Agalinos. K (2020) 'Discrete Event Simulation and Digital Twins: Review and Challenges for Logistics, Procedia Manufacturing, Volume 51, Pages 1636-1641
6. Tanner. J and Newbury C (2022) 'Digital Twins in the chemical process industries. Available at URL: <https://www.thechemicalengineer.com/features/digital-twins-in-the-chemical-process-industries/> (Accessed 25/9/23)

7. Chen J, Zhang Z & Wu F (2021) 'A data-driven method for enhancing the image-based automatic inspection of IC wire bonding defects', *International Journal of Production Research*, 59:16, 4779-4793, DOI: 10.1080/00207543.2020.1821928
8. Pedregosa F, et al. (2011) 'Scikit-learn: machine learning in python.' *J Mach Learn Res.* 2011;12:2825–30]]
9. Figard S, (2009)'The Basics of Experimental Design for Multivariate Analysis, Abbott Laboratories.
10. Kuchemüller KB, Pörtner R, Möller J. (2021) 'Digital Twins and Their Role in Model-Assisted Design of Experiments'. *Adv Biochem Eng Biotechnol*; 177:29-61. doi: 10.1007/10_2020_136. PMID: 32797268.
11. Jones D et al, (2020) 'Characterising the Digital Twin: A systematic literature review', *CIRP Journal of Manufacturing Science and Technology*, Volume 29, Pages 36-52, <https://doi.org/10.1016/j.cirpj.2020.02.002>.