Chapter 5
Digital Modelling and Simulation of Manufacturing Systems

Jack C Chaplin and Giovanna Martinez-Arellano

5.1 Introduction

Digital (adjective): involving or relating to the use of computer technology.

Simulation (noun): the production of a computer model of something, especially for the purpose of study.

- Oxford University Press

In the previous chapter, we introduced concepts of analysis and modelling manufacturing systems, focusing on productivity and capacity analysis, followed by queue modelling to improve throughout in a larger system. Both these techniques require the use of mathematics to evaluate data gathered from the real system. However, manual mathematical processing can rapidly get extremely complex for anything other than the simplest examples of manufacturing systems. Fortunately, computer-based digital modelling and simulation will allow you to glean the insights you require, while simplifying the process significantly.

Modelling tools and software help engineers be more productive in the process of conceiving, designing, modelling, evaluating and planning the implementation

J.C. Chaplin (✉), G. Martinez-Arellano
Institute for Advanced Manufacturing, University of Nottingham, Nottingham, UK
e-mail: jack.chaplin@nottingham.ac.uk, giovanna.martinez@nottingham.ac.uk

J.C. Chaplin et al (eds), *Digital Manufacturing for SMEs*
DOI: https://doi.org/10.17639/nfxp-ve75

of manufacturing systems. Modelling and simulating the manufacturing processes and the wider production line enables past experience to be utilised to predict the performance, and to detect issues and improve productivity in a virtual model before the expensive process of deploying a manufacturing process to the shop floor.

Modelling and simulation tools allow manufacturing companies to test changes and improvements to their manufacturing systems before physical implementation. In many cases decisions regarding manufacturing system changes are often based on experience and intuition rather than on quantitative prediction. Though this can often work, it relies on the availability of human expertise and is prone to error.

Creating modelling and simulation tools is a highly specialised field, requiring expertise in probability, statistics, and scientific computing. Fortunately, the wide range of commercial solutions means enterprises can take advantage of these tools off-the-shelf. Despite this, it is important to understand the underlying techniques used by simulations and models, as it will influence which one is best applied to which area. Often, multiple simulation packages may be required to give you coverage of both the more detailed areas of your system such as machine performance, as well as the larger more strategic areas like stock levels.

Taking a step backwards however, we will first discuss data capture in section 5.2 Data Capture and Analysis. As discussed in section 4.1, the purpose of analysis is to transform data into knowledge. Modelling and simulation are simply additional tools for analysis, so also exist to transform data into knowledge. However, your ability to accurately model your manufacturing systems relies on access to accurate data. Without accurate data, the insights generated by the models and simulationswill also not be accurate.

Section 5.2 Data Capture and Analysis discusses the entire data processing cycle from data collection, through analysis and to decision making. Following this, section 5.3 Modelling and Simulation Approaches discusses the range of computer-aided tools available for modelling and simulating manufacturing systems, the underlying priciples and assumptions, and example applications of their use.

5.2 Data Capture and Analysis

5.2.1 The Data Processing Cycle

Data (noun): facts and statistics collected together for reference or analysis.

- Oxford University Press

Analysis of data is often assumed to be taking collected data and running it through a mathematical formula or model to find an insight or result that helps you understand the subject of the analysis better. And in the first instance this is true. However, for more complex data analysis when systems are larger, changing, and producing much higher volumes of data, there are additional steps that need to be considered to ensure the analysis is accurate and useful.

The Data Processing Cycle is a series of steps to extract information from data in a structured and organised manner. It represents a series of steps from the collection of data to the disposal of it. However, the information gathered in the cycle may inform what new data is required, turning it from a process into a cycle. There are different life cycle models available in the literature, depending on the application. However, as shown in Figure 5.2-1 the general steps are:

1. *Data Collection:* Gathering the data from sensors or manufacturing processes and centralising it for processing.
2. *Data Pre-processing:* Initial cleaning and refinement of data to solve issues that might impede the analysis process.
3. *Data Analysis:* Use of a variety of techniques and methods to extract patterns and information from data and to find hidden information.
4. *Data Visualisation:* Presentation of data and information to decision makers to enable effective decision making.
5. *Data Interpretation:* Correctly understanding the information presented and making effective decisions.
6. *Data Storage/Disposal:* Either storing the data or information for future reference, or determining it to no longer be useful and disposing of it safely.

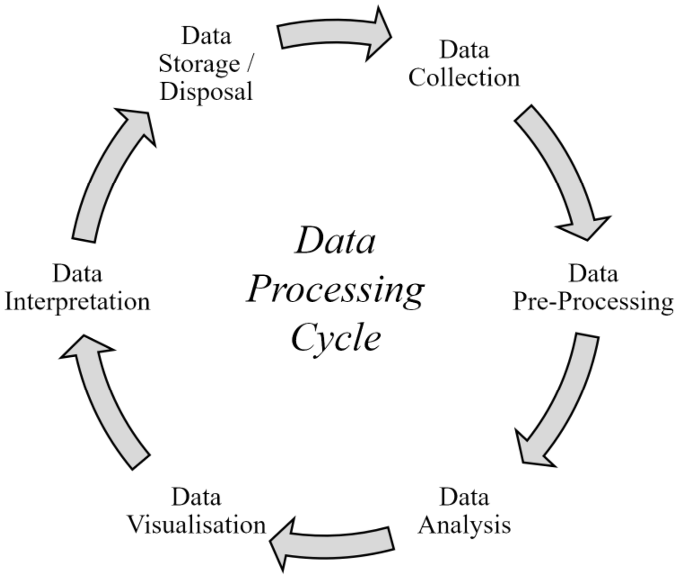

Figure 5.2-1 The data processing cycle.

We will discuss the stages in the data processing cycle in the next six sections.

5.2.2 Data Collection

Data collection is the first step in the data processing cycle and is one of the most critical to the success of the processing. The quality of the data that is acquired will impact on the quality of the analysis and decision-making. This has given rise to the common phrase “garbage in, garbage out” – a high quality data processing system will give bad outputs if given bad inputs.

Data gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Volume = 100), impossible data combinations (e.g. Product Finish = Unpainted Steel, Paint Colour = Green), missing values, etc. Analysing data that has not been carefully screened for such problems can produce misleading results. Typically, the data collected falls in one of the following types:

- *Structured data:* Data that has been organised into a formatted repository, typically a relational database. Usually organized as a table in rows and columns, and its elements mapped to specific fields, structured data is how databases store data. Fields have tight restrictions on the formats they accept such as numbers within a certain range, a choice from a list of options, or a simple Boolean (yes/no) value. Structured data is generally the easiest to work with, as the meaning of fields and numbers are well defined and organised.
- *Semi-structured data:* Data that does not reside in a structured table but does have some organisational properties. An email for example, has structured elements such as the recipient, sender, timestamp etc., but also unstructured elements such as the message body or attachment content. XML is another example of a semi-structured format with tags marking up the content, but the content of the tags being unstructured.
- *Unstructured data:* Data that is not organised in any pre-defined manner. Though the file type may have some structured elements the vast majority of the data does not conform to any standardised structure. Audio files, video files, most text formats, and this book chapter are all examples of unstructured data. Unstructured data is harder to work with, but 80% of all data gathered is unstructured.
- *Metadata:* Different to the above classifications, metadata means “data about data”. Any of the three above forms of data may include metadata. For example, a Word document contains metadata about who has edited the document, when it was last edited, the file size, and much more. This metadata can be used to aid in understanding what the data relates to and how best to process it.

The source of the data will influence the level of structure to the data. There are three types of data source generally considered in manufacturing systems:

- *Primary source:* Primary sources are sensors, transforming physical phenomena into electrical signals either binary, digital or analogue. Temperature, humidity, pressure, force, motion, acceleration, light and object presence are examples of things that can be detected and measured with sensors. Sensors may come built into manufacturing equipment, retrofitted into existing manufacturing equipment or may be stand-alone. In most cases, a manufacturing database will not receive data directly from sensors; the data will be first received and processed by an industrial controller, industrial PC, or other data acquisition device. These devices are *secondary* data sources.

- *Secondary source:* Devices receiving data from primary sources are secondary data sources. These can include devices such as programmable logic controllers (PLC), industrial robot controllers, CNC machine controllers, embedded computers, human-machine interface devices, and smart sensors with built in memory and processing capability. The advantage of taking data from secondary sources rather than primary, is that secondary sources have computational capability and storage memory. This allows converting the sensor data from raw electrical signals to more easily interpreted standards. Signals can also be compressed and pre-processed, requiring less storage space in the data warehouse. The data from multiple sensors connected to the same secondary source can also be combined and compared for additional insights that a single sensor may not be able to offer.
- *External source:* Not all data related to a product and its manufacture and service life may originate inside a single company. Process data from suppliers producing constituent parts can be centralised as the complete product comes together, aiding in understanding how the quality of supply can affect the quality outcomes of production. Modern use of external data sources remains heavily reliant on paper-based solutions, with parts and materials arriving with documentation. Highly fragmented and variable manufacturing data systems throughout the supply chain make automating acquisition of data from external sources a challenge.

Data gathered from primary sources such as sensors will likely be unstructured or semi-structured, whereas the computing power of secondary sources such as PLCs or embedded computers allows for a more structured approach to collected data. It is often easier to structure data closer to where it is gathered than to send it all to a single server for structuring and adding to a database.

The methods of collecting of data will depend on the manufacturing process you are looking to analyse, and whether the data is coming from primary or secondary sources. Data collection does not have to be automatic or in real time, periodic data collection and reporting remains useful, even when done manually. Automatic real-time data collection does have advantages, and these will be discussed in Chapter 6.

5.2.3 Data Pre-Processing

Noise (noun): irregular fluctuations that accompany a transmitted electrical signal but are not part of it and tend to obscure it.

- Oxford University Press

Conversion of data to information is more difficult when data is noisy or unreliable. It is also hindered when there is irrelevant and redundant information present. For this reason data pre-processing needs to be carried out to ensure the quality of the data and efficiency of the analysis process. Data pre-processing includes many sub-steps depending on the type of data, and these steps are discussed over the next seven subsections.

5.2.3.1 Data Cleaning (Missing Data)

Data may be incomplete, which could mean missing some values that should have been recorded, or missing certain attributes of interest during some measurement periods. This could be due to data not always being available (e.g. equipment malfunction), equipment not being set up correctly for some periods of measurement, or simply because the attribute which is now considered important was not thought to be worth measuring at the time.

For example a primary source sensor records vibration in a CNC machine. These readings are collected by a secondary source PLC, which adds the timestamp of the recording, and the spindle speed the CNC machine was using. Each measurement then has three attributes – vibration, timestamp, and spindle speed. This data is then used for analysis of the relationship between spindle speed and vibration.

A soldered connection was loose in the sensor, and for some periods of high vibration, the sensor recorded no data. As a result, the occasional random measurement in the data set has no value for vibration.

The CNC was serviced, and the PLC was incorrectly restored to an earlier configuration where the spindle speed was not being included in the measured data. As a result, a solid hour of measurements have no spindle speed attribute before the error was spotted and could be fixed.

Many analysis methods require homogeneous data – i.e. data where all values to be analysed have the same format and attributes. If attributes have been omitted due to failure or error, several methods can be used:

- *Ignore the entire measurement*: Best used only as a last resort when multiple attributes were missed or there’s no way of reconstructing the data.

- *Fill in missing attribute manually:* Only usable if a human worker can quickly and easily predict what the missing value would have been, such as a timestamp where measurements are taken periodically.
- *Use a global constant:* Define a "default" value for any missing attributes to automatically fill them in. Only usable for some types of data that has a single value most of the time.
- *Use an averaged value:* Use an average calculation to fill in the value, either the average of adjacent measurements, average over the whole data set, or some other calculation appropriate to the data used.

5.2.3.2 Data Cleaning (Noise)

Data may be noisy due to collection instruments being faulty or there being errors in data collection. However, the vast majority of noisy signals are unavoidable, and are a consistent reality in even the best electronic equipment. Noise is an unwanted random error or variance in a measured value, and hides the true value.

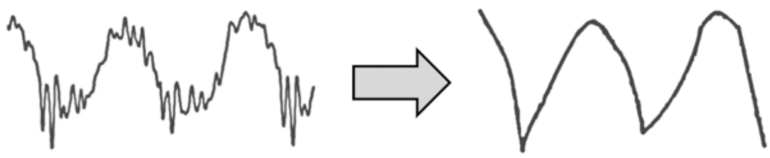

Figure 5.2-2 An example of a noise sensor measurement signal (left) and the same data after removing noise via smoothing (right).

Almost all electromagnetic devices and circuits will be subject to some degree of noise, and the study and characterisation of noise is a huge domain by itself. Noise can originate from the power supply of sensors, incorrectly grounded circuits, or the fundamental properties of electrons to name just a few. Several techniques to smooth the error can be applied such as:

- *Data binning:* These methods involve replacing values with intervals ("bins") based on the neighbouring values.
- *Clustering:* These techniques are used to detect unusual values or 'outliers' and remove them.
- *Regression:* Smoothing the data by fitting it to an automatically discovered function.
- *Filtering:* A variety of functions to remove certain frequencies or other components from the signal. For example, UK mains power has a 50Hz AC signal, and this often causes noise in sensor measurements. This component of the signal can be removed with a filter.

5.2.3.3 Data Cleaning (Inconsistency)

There may be inconsistencies in the data recorded for some processes, particularly when the data is being recorded manually. For example when manually recording information different operators may have inconsistent naming conventions for error types. Even automated data collection may be inconsistent when set up with inconsistent file names or data formats. Data may also be mistakenly duplicated, either due to manual error or a networking issue.

These issues should generally be corrected at the source, by defining and enforcing naming conventions and formats. Some automated tools do exist such as knowledge engineering tools, and they may be used to detect the violation of known data constraints based on previous knowledge and automatically highlight the inconsistencies.

5.2.3.4 Data Integration

Data may come from multiple sources including databases, spreadsheets, XML files etc. and you may not always be able to define the format and data representation standards used if they are set by the equipment you're using or come from external sources. Combining these can be difficult if data is labelled or represented differently. Redundancy is another issue. An attribute can be redundant if it can be derived from another one.

Though the most common solution to this problem is manual integration, using a data warehouse and the Extract, Transform, Load (ETL) method can simplify this process, transforming all the data into a single homogeneous format.

A data warehouse draws together many different data sources and provides a single interface to them all, allowing queries to look inside the many data sources without extra effort by the user. Databases are an example of a data source in a data warehouse. Data warehouses need to unify and integrate multiple data sources to enable them to be searched and compared. This is achieved with the ETL process.

A database is typically updated in real-time with the most recent values for data to enable decisions to be rapidly made. By comparison, a data warehouse will keep all historical values of data. A data warehouse is hence slower to access and analyse, but allows for a deeper insight into data than a database.

A *data lake* is another term occasionally used. Whereas a data warehouse uses ETL to harmonise all data into compatible formats, a data lake keeps data in its original form and only transforms it into other formats on demand. This process is sometimes called Extract, Load, Transform or ELT.

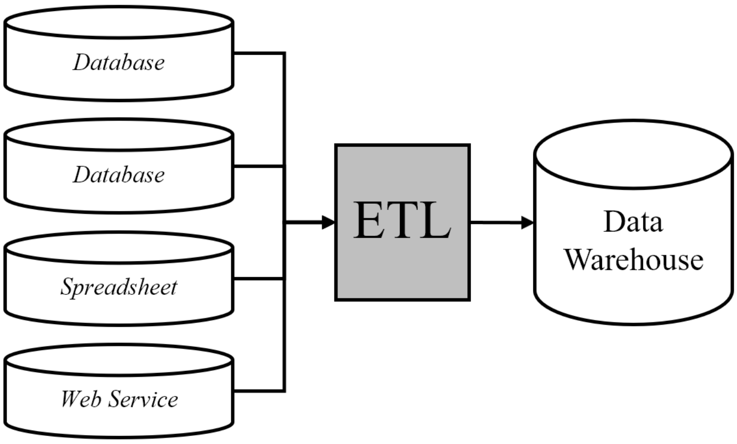

Figure 5.2-3 Data warehouses combine current and historical information from a wide variety of data sources using a method called ETL. ETL is an often complex process to regularly extract data, transform it into a common format, and load it into the warehouse. Though complex, this will save time in the long run when large amounts of data are stored.

5.2.3.5 Data Transformation

Whereas data integration converts an entire set of data from one format to another for use with different programs without changing the actual values of the data, data transformation changes the values in the data itself into a form more suitable for analysis. The transformations required will depend on the type of analysis, but can include:

- *Normalisation:* Changing the scale of values of the data to ensure compatibility. This often means using a specified range such as -1.0 to 1.0 or 0 to 1.0.
- *Aggregation:* Refers to creating summary or aggregation operations to attributes, for example calculating the monthly or annual amounts of daily sales rather than working with daily values.
- *Generalisation of the data:* Raw or ‘low level’ data is replaced by higher level concepts. For example, categorical attributes like street can be generalised to higher level concepts like city or county. Similarly, numerical values like age can be mapped to categories such as young, middle-aged and senior to move the focus from specific ages to categories.

For example, a company wants to compare the quality of parts manufactured with different machine centres on the shop floor, to see which machines might need maintenance or adjustment. There are a variety of machining centres of different *sizes and ages to be analysed. The workers have noticed that the newest machining centre produces more parts that fail quality checks that any other and are concerned about their investment.*

Critically, the only data the workers are looking at is the absolute number of parts which fail quality control. The new machine produces more parts overall than the other machining centres. If the data for each machining centre can be normalised into a percentage of failed parts rather than the absolute number, it becomes clear that the new machine has a lower percentage of rejected parts than any other, even if the absolute number is higher.

Note that a process also called normalisation is an important aspect of database design, but this process is distinct from the data normalisation discussed here.

5.2.3.6 Data Reduction

Reducing the volume or dimensions (i.e. the number of attributes) of the data. This technique is useful when data analysis on the complete data set is unfeasible or impractical. Data reduction techniques need to ensure the integrity of the original data is not compromised, and produces a reduced set that can still ensure quality knowledge extraction. Some strategies include:

- *Dimension reduction:* Redundant attributes may be removed, either manually or after an initial period of analysis.
- *Data compression:* Encoding mechanisms are used to reduce the size of the data set. Size can be a major problem in some applications, making it difficult or impossible to process given available computing power.
- *Numerosity reduction:* Data is replaced or estimated by alternative, smaller data representations, such using regression models or clustering techniques.
- *Discretisation and concept hierarchy generation:* Raw data values for attributes are replaced by ranges or higher conceptual levels. For example, replacing numerical values with concepts such as low, medium and high. This is a similar concept to generalisation for data transformation.

As an example, consider monitoring power quality disturbances. In order to effectively identify problem causes one needs to sample in the range of MHz [1]*. Compared to other applications where data is gathered at frequencies less than 25Hz, this represents 40,000 times more data. In this case techniques such as Wavelet Transform and Principal Component Analysis (PCA) can be applied to achieve a high reduction ratio, in some cases above 90%, depending on the technique used* [2]*.*

5.2.3.7 Data Pre-Processing Comments

Once the data has been pre-processed, the outcome of this process needs to be saved. Often a simple comma separated value (CSV) format is more than sufficient, and can be read by all spreadsheet and analysis applications. If dealing with a complicated output format, databases, XML files, or JSON files (amongst many other formats) can be used. These formats are established standards and almost all technologies that are available today can understand them easily.

As can be seen by the size of the data pre-processing section, it's a complex area and often overlooked by those less familiar with data analysis. However, as the table below shows, data pre-processing takes up the majority of professional data scientist's time, showing how important but also difficult the task is.

Task Name	Percentage of Time
Data Collection	19%
Data Pre-Processing	60%
Data Analysis	9%
Refining Analysis Algorithms	4%
Building Training Sets	3%
Other	5%

Table 5.2-1 The proportion of time data scientists spend on tasks [3]. The majority by far is data pre-processing. The same study that collected this data also showed that data pre-processing is the least enjoyable aspect of being a data scientist.

5.2.4 Data Analysis

After ensuring the data is cleaned and in the appropriate format and structure, it is ready for *analysis*. Modern data analysis blends traditional statistical data analysis with new and emerging computational methods. Analysis uses a huge number of techniques, and many are specific to the type of data being processed. Some of the largest categories are discussed below, starting from highly numerical analysis and moving to more textual analysis:

- *Statistical analysis:* The process of generating statistics from stored data and analysing the results to deduce or infer meaning about the underlying dataset or the reality it attempts to describe. Some of these statistics include Bayesian analysis, conditional probability, data classification, linear regression, resampling, shrinkage, tree-based analysis, to name just a few [4]. Though advanced data analysis methods such as machine learning can be very effective, for many domains well executed classical statistical analysis can be as or more effective. Statistical analysis may not be the most modern or trendy method of data analysis for numerical data, but techniques are well understood and refined, and results are often extremely good for all but the most difficult of circumstances.
- *Quantitative analysis:* Techniques that try to understand behaviour by using mathematical and statistical modelling and measurement, to create a model of

the process being measured. This model allows analysts to examine and test the past, current, and anticipated future events.

- *Qualitative analysis:* This uses subjective judgment based on unquantifiable information, such as management expertise. While quantitative analysis uses exact inputs, qualitative analysis deals with intangible, inexact concerns that belong to the social and experiential realm rather than the mathematical one. Quantitative and qualitative analysis are often used together in order to examine a company's operations and evaluate potential investments.
- *Semantic analysis:* This is the use of ontologies to analyse content in text-based sources. Ontologies are data models of the formal naming and definition of categories, properties and relation between concepts, allowing automated systems to understand the approximate meaning of text. It uses text analytics to measure the relatedness of different ontological concepts. It can be used for natural language processing i.e. automatically understanding human speech. A common example of use is understanding if a tweet or review uses positive or negative language about products.

Two additional techniques have emerged in recent years, and are used alone or in combination with the above techniques.

- *Data mining:* This is the process of discovering patterns in large data sets (also called "big data"). These patterns can then be seen as a summary of the input data and may be used in further analysis and predictions.
- *Machine learning:* This exploits patterns found in historical data to identify risks and opportunities. Machine Learning refers to the use of algorithms to learn automatically from the data, without using explicit instructions or relying on models.

5.2.4.1 Data Mining

Knowledge is a very valuable asset in manufacturing, as it enables a business to differentiate itself from competitors and to compete efficiently and effectively to the best of its ability. The advancements in information technology (IT), data acquisition systems, and storage technology as well as the developments in machine learning (ML) tools have led towards new ways of knowledge discovery in manufacturing processes. Data from almost all the processes of a manufacturing business such as product and process design, material planning and control, assembly, scheduling, maintenance, recycling, etc., are recorded. These data stores offer enormous potential as sources of new knowledge. However, data must be analysed and converted into actionable knowledge to be useful. In addition, the volume of collected data is becoming an issue with insights buried in large volumes of other data.

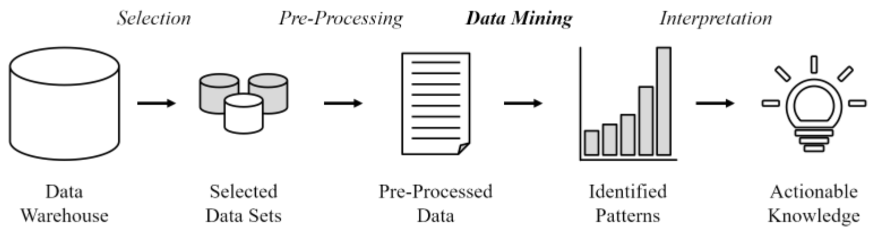

Figure 5.2-4 The data mining process. Sources of data are selected from the data warehouse, some pre-processing performed, and then patterns identified in the data. These patterns are then converted into actionable knowledge.

Data mining is an area of computational intelligence focusing on providing new systems, techniques, and theories for the discovery of hidden knowledge in large volumes of data. It is a blend of concepts and algorithms from statistics, artificial intelligence and data management. The term *big data* is also commonly used to refer to this area of data analysis.

The use of data mining in manufacturing began in the 1990s and is gaining traction in different aspects of the manufacturing process such as predictive maintenance, fault detection, design, production, quality assurance, scheduling and decision support systems. A typical data mining process is shown in Figure 5.2-4.

Data mining methodologies were introduced to provide a more holistic view to the knowledge discovery process, beyond the application of statistical or machine learning algorithms. Data mining looks at a single set of data to discover previously unknown patterns with no prior information, whereas machine learning uses past experience to find examples of known patterns in new data sets. The two domains cross over frequently, so the distinction is often blurry.

5.2.4.2 Machine Learning

Machine learning is a broad topic, but generally refers to algorithms that learn over time to give improved results. Machine learning as a term and as a concept dates back to 1959, but it is recent improvements in computational power that has brought it to the mainstream. Machine learning algorithms use a *training set* of data which has been labelled and uses that to automatically generated a predictive model. This model is then used on a new set of data and assigns a label automatically.

Machine learning is now in common use with applications including congestion prediction in satellite navigation apps, email spam filtering, fraud detection in banking transactions, and facial recognition systems. Anywhere in the manufacturing domain where there is a requirement to assign a class or type to a piece of data is a potential application of machine learning.

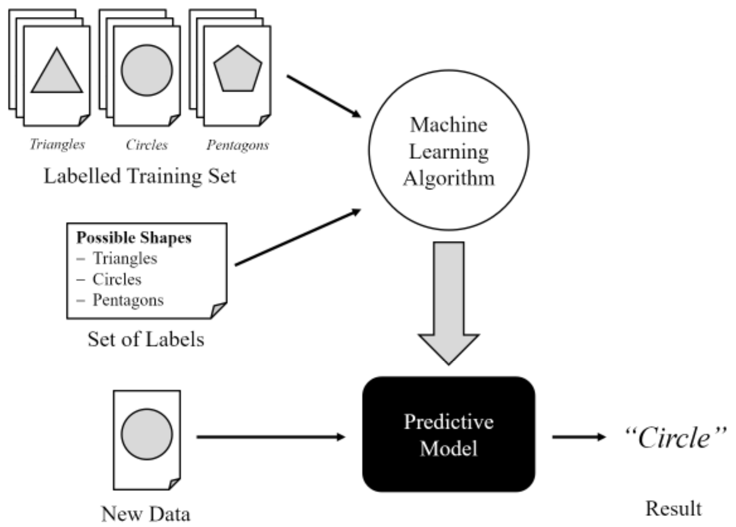

Figure 5.2-5 An overview of the machine learning process. A training set of labelled images of shapes (and a set of all possible labels) is used by the algorithm to generate a predictive model. Future unlabelled images of shapes will then automatically be classified, such as the pictured circle.

For example, a company has been using a vision system to identify defects in the paint finish of their products. The system can identify something out of the ordinary but cannot tell what type of fault has been detected without a worker to analyse the image. The worker is given a picture of the defect and assigns it a label. The label is the type of imperfections, and could be chip, abrasion, dent, scratch, bubbling, or more. Knowing the type is important as it helps identify the cause of the problem and stop it happening again. The company wishes to automate the labelling process as it is a time-consuming process and their workers could spend their time better elsewhere.

Here, the set of manually labelled images of imperfections becomes the training set for the machine learning algorithm. The algorithm creates a predictive model from this training set. Any future imperfections which are identified will be run through the algorithm, and the algorithm will automatically classify the imperfection type.

Machine learning is not without its drawbacks, however. Machine learning requires a large training set and access to a labelled set of data is often the largest hurdle in implementing a machine learning solution. The quality of the results is also highly dependent on the quality of the training set. Consider the example in Figure 5.2-5 – what if the new data were a picture of an octagon? The predictive model would likely classify it as a circle as it has no examples of octagons to use, and a circle would be the best fit from the available labels. Algorithms can state a level of certainty in their classification however, to flag issues like this. A final problem is that the predictive model is a "black box" – it was automatically generated and how a specific model works is extremely complex and almost always unknowable to a human. This means a model which has problems becomes very difficult to troubleshoot.

Both data mining and machine learning are made possible with extremely complex algorithms, but knowing exactly how they work is not necessary to start implementing them, as a great many software packages exist that handle the complexity. The main choice would be between free, open source packages such as Google's TensorFlow, Facebook's PyTorch, Keras, and Weka; or using proprietary integrated solutions such as those available in Amazon Web Services, Microsoft Azure, or Google Cloud.

5.2.4.3 Data Analytics

Also of note is the field of *Data Analytics*. Whereas analysis focuses primarily with data from the past and understand why events occur, analytics focuses on what is likely to occur in the future and how to respond to it. Data analytics is also sometimes called *Business Analytics* when applied to more general aspects of business performance. Analytics is typically defined as having three types:

- *Descriptive analytics:* What happened? This is effectively analysis, and focuses on what happened in the past.
- *Predictive analytics:* What will happen? Using predictive modelling such as machine learning and some statistical techniques to predict what is likely to occur in the future based on previous events.
- *Prescriptive analytics:* How can we make things happen? By understanding what events or parameters contributed to past events, automatically recommend changes to parameters to influence future events.

Many software packages or programming languages exist for data analysis and analytics, with free open source examples including KNIME (Konstanz Information Miner) for data analytics, R which is a dedicated data analysis programming language, and SciPy which is a data analysis library for Python. Many commercial software packages also exist, often described as *Business Intelligence* software.

5.2.5 Data Visualisation

Data analysis is often carried out in parallel with *data visualisation*, and indeed many data analysis and analytics tools will include data visualization capabilities. Their purpose is to show in a graphical and easily understood way, the trends and the structure of the information that has been discovered through the analysis process. Broad relationships and patterns can be brought out, as can emerging trends. Visualisation also helps to quickly narrow the search for information of interest. It can be static, in the form of charts and diagrams in reports, or live, such as manufacturing dashboards reporting live data from the shop floor.

Common examples of visualisation software include Microsoft Power BI, Dundas BI, and Tableau, but many companies also develop customised dashboards with visualisation libraries such as R Studio's Shiny or Grafana. Even spreadsheet software's reporting and graphing features can achieve these goals.

Data visualisation is an example of a data-driven decision support system. Decision support systems will be discussed more in Chapter 6.

5.2.6 Data Interpretation

Though automated tools exist to assist in making decisions the ultimate interpretation and decision making responsibility still typically lies with humans. The data processing cycle exists to facilitate the decision making process, but the decision itself should be in line with established company procedures.

It is important when a decision is being made to understand the type of problem that is actually being solved, as this will influence how much and what type of data is required to come to a good decision.

- *Unstructured problems:* These possess multiple solutions and there is no algorithm or formula that can lead to the optimal solution due to uncertainties in the problem, and there are few parameters that can be directly affected to solve the problem. Unstructured problems are often rare or novel, and do not have established responses. One example would be a decision around employee reward schemes where the results of the decision are more about employee satisfaction and company perception rather than the exact value of the reward. Unstructured problems are typically solved with informal intuitive decisions based on experience.
- *Structured problems:* By contrast, structured problems have very clear parameters and goals, and the results of a decision are possible to model and predict. Often there is a small number of criteria to be maximised, and algorithms exist to model the decision. Structured problems are often routine and well understood. An example would be the response to a common maintenance requirement for a piece of equipment or machine tool. The problem is common and well understood, and experience shows what the best solution is. The costs of replacement parts and downtime are all understood.
- *Semi-structured problems:* These are problems where only part of the problem can be optimally solved, requiring a combination of a standard solution and individual judgement. They are the grey area between structured and unstructured. An example would be creating a maintenance plan for a new piece of equipment. You know the cost of downtime and replacement parts, but you have no intuitive understanding of how frequently the machine will need maintenance performed until it has been in use for a considerable time. Failure modes are not fully understood.

The more structured a problem is, the easier it is to apply data and find an optimal solution. So far, the methods we have been discussing are best applied to structured and semi-structured problems. Unstructured problems are more challenging to take a data-driven approach too, but that does not mean it is impossible. Many unstructured problems can be reduced to semi-structured one upon closer analysis, and tools such as decision analysis models (DAMs) can start to decompose the problem into smaller, manageable components. DAMs are statistical tools and methods such as analytical hierarchy processes (AHP), decision tree analysis, multi-criteria decision analysis, and probabilistic forecasting to support decision making where multiple criteria need to be considered and there is no single optimal answer.

There are two common problems when it comes to making decisions assisted with analysed data:

- Decision makers make bad decisions because they do not have access to high quality data, and either are misinformed by bad data or instead use intuition or gut-feel.
- Data scientists attempt to make decisions based on data, when they do not have a detailed understanding of the real world in which the data originates.

The simple solution is to ensure data scientists and those with domain knowledge work together on the decision making process. Performing all the steps in the data analysis cycle will provide a solid, reliable set of analysed and visualised data with which to base decisions on, but this must also be combined with real-world experience.

5.2.7 Data Storage and Disposal

For the collected, processed, and analysed data this stage of the cycle represents a fork. Data can be stored and fed back into the data collection stage of the cycle for future use and analysis as new data is acquired. As the amount of data that can be generated for the production of a single manufactured item can be in the order of gigabytes per day, there is a need for a database that is capable of storing this data. Though the absolute cost of data storage continues to decrease (particularly when cloud-based solutions are considered), the amount of data that is being generated continues to increase, and the rate at which data can be sent and accessed is not increasing at the same speed.

Alternatively, data can be considered to have fulfilled its purpose, and be disposed of. Data can take up a large volume of storage space, and permanent storage is a costly endeavour. Though storage costs are decreasing, the volumes of data being gathered by companies are also increasing. An alternative is to dispose of the raw collected data, while keeping the analysed data. Analysed data typically requires a smaller amount of storage space, so you can keep the insights even if you dispose of the source of the insight.

When disposing of data, consider how well the data is erased. Simple file deletion may not remove data from the storage medium. This may not be a huge concern for gathered process data, but data which is commercially or personally sensitive should be securely deleted or it could be recovered by a malicious individual such as a corporate rival or disgruntled former employee. Consider also your responsibilities under the General Data Protection Regulation (GDPR) in the EU or equivalent local regulations – personal data must be gathered only when necessary, used for a clear and specified purpose, and disposed of securely when no longer required.

5.3 Modelling and Simulation Approaches

5.3.1 Introduction to Modelling and Simulation

A model is a mathematical representation (and often a simplification) of a system built using mathematics. A simulation is using the model to predict the behaviour of the system under specified parameters, by inputting values to the model and recording the results. Modelling and simulation tools allows manufacturing companies to test changes and improvements to their manufacturing systems before physical implementation. In many cases decisions regarding manufacturing system changes are often based on experience and intuition rather than on quantitative prediction. Though this can often work, it relies on the availability of human expertise and is prone to error.

The creation of a modelling or simulation package requires extremely specialised knowledge and expertise. Though these tools will hide many of these details of their implementation from the user, understanding the underlying model type will aid in understanding which package is best applicable to your needs. Modelling tools tend to specialise in specific areas, and often a combination of packages will be required to get good coverage of the problems you want to solve.

For manufacturing, these specialised areas can be broadly grouped into three domains, and it is often the case that you will need a separate package or solution for each domain at minimum:

- The *product domain* is related to the product that is to be manufactured, and can be classified into structure-oriented, geometry-oriented, feature-oriented, or knowledge-oriented models.
- The *process domain* represents the relations between events and activities in a manufacturing system, describing how a manufacturing system as a whole should act, without going into the specific details of each manufacturing machine and resource.
- The *resource domain* represents the operative instructions for a given manufacturing resource, such as a machining centre or robot. These instructions are typically specific to the resource being modelled, such as the G-code of a CNC machine.

The following section gives an overview to the most common modelling approaches used by softwarte packages, and details their applicability to help you evaluate a package for your needs.

5.3.2 Types of Modelling Approaches

5.3.2.1 Discrete Event Simulation

Discrete Event Simulation is a simulation methodology that has been widely adopted within industries to test manufacturing system changes virtually before implementing them physically. It allows for a high-level analysis of the system's performance by statistically and probabilistically reproducing the interactions of its components and resources.

The entire manufacturing facility can be modelled as a sequence of operations being performed on passive entities (e.g. components), as they pass through the processing sequences. Although the components are passive, they have attributes that affect the way they are handled and some of these attributes change as the component advances through the process. The simulation allows you to test changes to your production line and answers questions such as:

- How long will the cycle time be for this new product?
- What is the utilisation of equipment, do I need to invest in more of some resources or optimise the utilisation of others?
- Are my buffers likely to fill up and queues form in the process?
- What is the predicted error rate of this process, and is the process cost effective?

You may notice many of the questions above are similar to those answered in Chapter 4, and that's not a coincidence. DES is a statistical simulation rather than a physical one. It allows for the inclusion of existing data from your resources (e.g. what is the cycle time of a process, how often does a machine require maintenance, and how often does it fail) to predict future performance under different circumstances. It is often simple to implement within a company, provided you have historical data to base predictions on.

5.3.2.2 Agent-Based Modelling

Agent-based modelling is a tool that allows the modelling and simulation of complex systems by breaking the problem into simpler units and modelling those using intelligent agents. In this way, the modeller only needs to understand the behaviour of simple components to create the model, whereas techniques such as system dynamics require a better understanding of the system as a whole.

An *intelligent agent* (typically simply called an *agent*) is a piece of independent software is capable of exhibiting autonomous and intelligent behaviour [5]. An agent is an individual problem solver with some capability of sensing and acting upon is environment. They can be used both for the control of systems (where the environment is the physical system), or for modelling systems (where the environment is a model). We focus here on the application to modelling systems.

There are many types of agents, but they have some basic properties:

- An agent can observe its environment with sensors.
- An agent makes a decision based on the observations.
- An agent initiates and executes actions using actuators to change the environment.

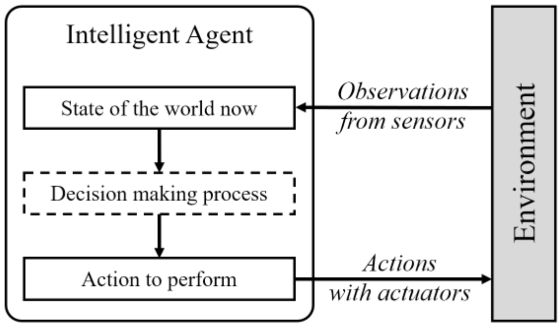

Figure 5.3-1 The basic function of an agent. The decision making process can vary from simple condition-action rules, to highly complex artificially intelligent learning processes.

Agent-based modelling breaks down complex problems by focussing on simpler components, with each agent representing the behaviour of a resource or worker in the system. The complexity then emerges from the interactions of these components, each represented by an agent. For modelling and simulation the environment, sensors, and actuators would all be virtual. However, agents are increasingly used as a control method for real physical systems, so the basic definition omits “virtual”.

In manufacturing, this tool has been typically used for modelling production planning and resource allocation, production scheduling and control, monitoring and diagnosis, production in networks and assembly and life-cycle management. Due to their distributed nature, agent-based systems provide modularity, robustness and autonomy, which offer an alternative way to design and control systems when compared to conventional approaches. An example of existing environment for the simulation of multi-agent systems in the manufacturing domain is MAST (Manufacturing Agent Simulation Tool) [6], developed by Rockwell Automation. This tool is particularly focused on dynamic product routing.

5.3.2.3 System Dynamics

System Dynamics is a technique for constructing models that focus on interdependencies, feedback effects, time dependencies, and causality in the object that is being represented. It allows for more complex modelling than DES, and is applied beyond manufacturing in domains such as finance, population growth, agriculture, and ecological behaviour. Feedback loops are one of the key considerations in system dynamics, and cause some of the least predicable behaviour in complex systems such as production lines and facilities, and can cause serious issues if not understood [7].

The system orientation of system dynamics makes it ideally suited for the analysis of future production dynamics, and can include operational and organisational aspects in addition to shop-floor elements. It allows studying how information flow affects the behaviour of productive systems, an aspect often not covered in other simulations, which only consider the flow of physical objects. A typical system dynamics model is built using the following elements:

- *Levels* (shown as rectangles) represent the quantity of some element of the system at a point in time, such as stock levels.
- *Rates* or *flow* variables (shown as valves) represent the rate of change levels in the system in an interval of time, such as utilisation of coolant.
- *Converters* (shown as circles) are additional equations or calculations that effect the system.
- *Connectors* (shown as arrows) are the information links in the system connecting other components together. The arrows are often given different weights to show the difference between information (thin arrows) and changing physical quantities (thick arrows).
- *System boundaries* (shown as clouds) are the edges of the system being modelled.

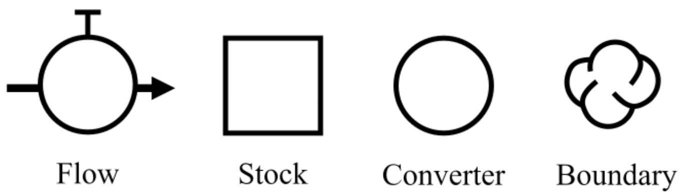

Figure 5.3-2 Systems Dynamics diagram components. Image rights: Author, adapted from [8].

Depending on the problem being modelled, system dynamics can be implemented in two phases:

1. *Qualitative system dynamics*: This phase involves the creation of cause and effect diagrams or system maps.
2. *Quantitative system dynamics*: This phase involves deriving the shape of relationships between all variables within the diagrams, the calibration of parameters, and the construction of simulation equations and experiments.

One of the most common diagrams in system dynamics is the levels and rates diagram (also known as the stock and flow diagram). Figure 5.3-3 shows an example.

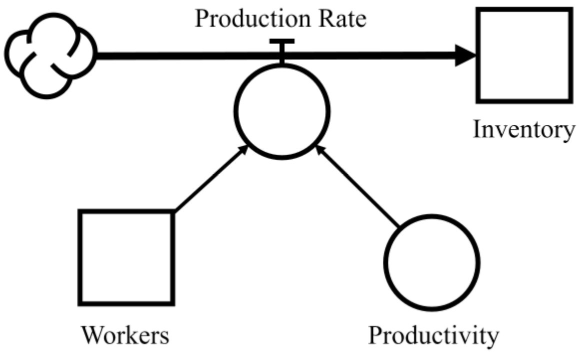

Figure 5.3-3 Example of a levels and rates diagram, which models a production line's production rate as being dependant on the quantity of workers available and the productivity of the system, resulting in finished products being added to the inventory. Image rights: Author, adapted from [8].

Once the system is defined with system dynamics, a detailed evaluation is used to find the ideal solution. Large amounts of data are required at this stage, making this evaluation a challenge for some manufacturing companies. In general, system dynamics is also a more complex simulation to create than DES, but offers considerably more possibilities for understanding highly complex systems.

5.3.2.4 Petri Nets

Petri nets are a graphical and mathematical modelling tool applicable to several systems. As a graphical tool, Petri nets have been used as a visual communication aid similar to flow charts or network diagrams. The classical Petri net is a directed graph with two node types called *places* and *transitions*. Places are represented by circles and transitions by rectangles. A petri net has states, represented by *tokens*. Places may contain zero or more tokens, which are represented by black dots. A transition is activated only if all the places that are inputs to the transitions have tokens, and a transition will consume the input tokens and move the tokens to the output places.

Figure 5.3-4 shows a Petri net that models a machine which processes jobs and has two states: free and busy. There are four places, *Storage* (1), *Available* (4), *Processing* (2) and *Complete* (3), and there are two transitions, process start and process finish. There are five tokens, 4 in place *storage* and 1 in place *available*. The tokens in *storage* represent parts to be processed, and the token in *available* represents the availability of the lathe for processing. The system proceeds in discrete time steps.

The transition representing the process can only activate if there is both a token in *storage* (representing the part to be processed) and a token in *available* (indicating that the process is not busy). The transition will consume these tokens, and place a token in *processing*. The process cannot start again as – although there are parts available in storage – the lathe is not available.

The transition representing the process finishing will consume an available token in *processing* and place a token in *complete* (representing a finished job) and in *available* (representing the lathe is available again). The Petri net is then ready to start a new job by taking a token from *storage*, and will repeat until all parts are processed.

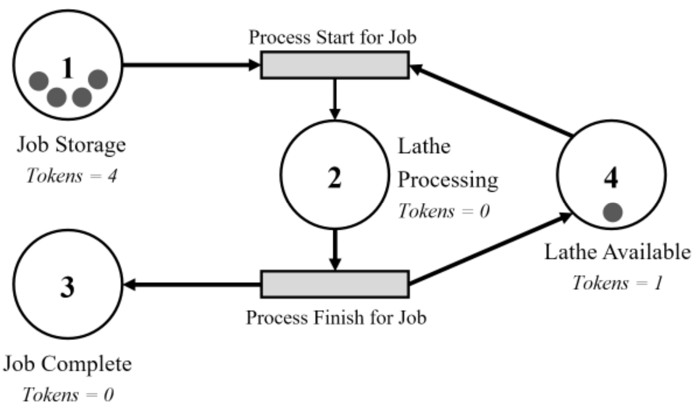

Figure 5.3-4 Petri net modelling a machine with two states: processing and available.

Petri nets are a powerful but difficult and verbose modelling method. They are often the underlying model for more user friendly modelling software, and many modelling tools will handle the lower-level modelling for you.

5.3.2.5 Monte Carlo Simulation

Monte Carlo simulation is a computerised analytical process that is used to evaluate and measure the risk associated with any given venture or project. It allows users to understand a given situation but also the impact of other possible scenarios.

Rather than using absolute values for a simulation (for example, an average run time for a milling process) it utilises a probability distribution of values.

An example of this probability distribution is the Normal distribution. During the simulation, samples are drawn randomly from the input ranges and the results recalculated over and over. The result is a range, or distribution, of possible outcome values and their associated probabilities of occurrence. Compared to deterministic methods, MCS offer the following advantages:

- Provides *probabilistic estimations of potential results*, rather than single values. For example, a conventional simulation might inform you of an estimated process time that is acceptable to you. By comparison, an MCS will inform you the odds of the process time falling outside of acceptable bands.
- Provides information of *which inputs contributed the most to a particular result*. Deviations in one process may have little to no impact on overall system performance, but a different process may have a huge effect with even small deviations.
- Allows the user to *understand the behaviour of inputs* related to a particular resulting scenario. What potential combination of variables are likely to cause your system to have problems? How can you protect yourself against this?

An example of the use of MCS in manufacturing companies is for minimising the number of disruptions caused by the supply chain. By using this type of simulation, variables such as sales demands, material costs, investments outlays, delivery and processing times, inventory uncertainty and disaster risk can be modelled, and both independent and interrelated variables considered. When the MCS runs, the user can consider thousands of scenarios and determine several outcomes such as which factors are most likely to impact supply chain; weather, material costs, overhead expenses, fluctuating prices, to name just a few.

5.3.2.6 Virtual Simulation

Many modelling methods now have three-dimensional rendering capacity, allowing the user to see the results of a simulation in a realistic and intuitively understandable way. Advanced in virtual reality and augmented reality technologies can even present the user the system in an immersive, interactive fashion.

However in some cases, creating a 3D model to get an intuitive idea of a design or system is sufficient to achieve the aims of the modelling process. Examples include computer-aided design (CAD) models of products to test for usability (though CAD offers a lot more also), and the creation of models of assembly stations or working areas for workers to get a feel for ergonomics and check for issues before the workstation is built.

5.3.2.7 Simulation Gaming

A model is a mathematical abstraction of a system, and the usual approach is to use the model to simulate the result of specified parameters, or to optimise the parameters for desired outcomes. Sometimes however, the model is used to test how humans will react to the situation. This is called simulation gaming and allows for testing human responses to unexpected or stressful situations such as incident management, or for testing the usability of interfaces and workstations.

5.3.2.8 Intelligent Simulation

Modelling and simulations can be extremely complex. These systems often have a lot of parameters and to explore all possibilities can require unfeasibly long run times, providing answers too late to be useful. *Intelligent simulation* is based on Artificial Intelligence (AI) techniques and typically supports other types of simulation techniques**.** For example, machine learning and intelligent sampling can be used for the optimisation and calibration tasks of an agent-based system, creating meta-models that can deliver dramatic speed increases for large-scale models. For digital twin and what-if analysis systems, AI components can be directly embedded in the simulation model to allow testing and forecasting.

Another application of Intelligent Simulation is the use of deep learning for the replacement of rule-based systems. PricewaterhouseCoopers, for example, currently supports a large car manufacturer to introduce autonomous vehicles for the public. Part of this work involves using deep reinforcement learning to determine optimal decision rules that allows the vehicles to maximise efficiency also satisfying customer trip demand. Similar approaches can be used to automated manufacturing lines.

5.3.2.9 Distributed Simulation

Where a model is so complex running a simulation would take an extremely long time, a distributed simulation may be used. Many underlying modelling methods can be used, the challenge here is to make model decomposable into smaller units which can then be distributed to multiple computing resources to execute the simulation in parallel.

Another use case for distributed simulation is where multiple entities (such as companies) wish to create a model and run a simulation together, but also wish to create their models using different modelling packages, or wish to keep elements of their model secret. Many of the standards in distributed simulation such as IEEE 1278 [9] have their origins in military wargaming, which required the distribution of simulation with elements kept secret.

5.3.3 Applications of Modelling

Modelling has a wide range of applications on manufacturing. Perhaps the most commonly used is CAD – Computer Aided Design. CAD allows for a model of a proposed design to be built, and virtually tested before prototypes are made. It allows for dimensional analysis and validation, as well as engineering analysis such as:

- *Mass properties analysis*, such as product volume, surface areas, weight, and centre of gravity.
- *Interference checking* in multi-component designs, checking components will not physically clash.
- *Tolerance analysis* to automatically determine what tolerances are required for correct product operation.
- *Finite element analysis* provides approximate solutions to tests such as stress-strain, heat transfer, or fluid flow which would otherwise require physical prototypes.
- *Kinematic and dynamic analysis* helps test the motion of multiple linked components and analyses their motion properties.

As well as this, quality of life services such as automated design reviews, automated drafting, and version control are typically integrated in CAD programs.

Within the context of manufacturing systems analysis, modelling has a large number of applications, with some listed below along with the underlying types of modelling typically used for these applications [10].

- *Assembly line balancing:* The design of assembly lines, and the process of balancing them. Balancing is ensuring that a line has sufficient resources (e.g. workers or equipment) to meet the required production rate, without excess spare capacity.

  Modelling Methods: Discrete Event Simulation.

- *Capacity planning:* Modelling external unpredictable elements of a business environment to ensure the business has sufficient capacity to deal with the fluctuations. For manufacturing this will include fluctuations in supply and demand, and ensuring the company has sufficient storage, buffers, and production capacity to cope, or identifying areas that could be changed or refined to meet the required capacity.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Monte Carlo Simulation, Petri-net Simulation.

- *Cellular manufacturing:* Optimising the layout of a cellular manufacturing set-up for increased production capacity and improved operator ergonomics. Also checking new scheduling schemes or process plans for potential operational issues.

  Modelling Methods: Virtual Simulation

- *Transportation management.* Supply chain modelling to evaluate the effectiveness of vehicle routing, truck loading, distribution centre utilisation, and incident management approaches.

  Modelling Methods: Discrete Event Simulation, Agent-based Simulation, Petri-Net Simulation, Other (Traffic simulation, a specialised set of modelling methods for transportation)

- *Forecasting.* Predicting future patterns and analysis of trends. In a manufacturing context this will generally mean market forecasting to predict the demand for a company's products under changing global situations.

  Modelling Methods: System Dynamics

- *Inventory management.* Evaluation of the cost and benefits of holding inventory, either to save space or insurance against supply disruption. Can also include elements of inventory utilisation policy (i.e. when to replenish stock, which stock to use first) and understanding at what level of stock should a replenishment order be placed.

  Modelling Methods: Discrete Event Simulation, Monte Carlo Simulation.

- *Just-in-time.* Design of just-in-time (JIT) manufacturing systems, which are systems where the receipt of parts from suppliers aligns extremely closely with when the parts are due to be used (i.e. the parts arrive "just in time"), resulting in low levels of inventory required and quicker cycle times. This requires a very carefully balanced production line, as even small disruptions to supply or process can cause significant system-wide delays.

  Modelling Methods: Discrete Event Simulation, Intelligent Simulation.

- *Process engineering - manufacturing.* Design of manufacturing processes including elements of process design, ramp-up optimisation, and performance measurement and optimisation. Can also be used to plan entire new facilities or to evaluate the potential impact of new equipment acquisitions.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Agent-based Simulation, Monte Carlo Simulation, Petri-net Simulation, Virtual Simulation, Intelligent Simulation.

- *Process engineering – Service.* Design of service processes such as logistics and distribution, waste handling, retail and the service industry, and financial services. Service processes share many similar concerns with manufacturing process, including scheduling, capacity, bottleneck analysis, and performance measurement and optimisation.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Distributed Simulation.

- *Production planning and inventory control.* A specialisation of process engineering – manufacturing, production planning and inventory control specifically focuses on batch size optimisation, bottleneck analysis, forecasting and scheduling.

  Modelling Methods: Discrete Event Simulation, Agent-based Simulation, Distributed Simulation.

- *Purchasing.* Optimisation of purchasing strategies for parts and supplies, including when to replenish stock, and understanding minimum purchase sizes and bulk discounts to keep sufficient stock with the minimum price and without excess inventory.

  Modelling Methods: Discrete Event Simulation.

- *Resource allocation.* Assigning equipment, workers, and time/overtime to tasks to improve process flows and productivity under varying circumstances. Can also include assignment of raw materials, tooling, and storage to processes.

  Modelling Methods: Discrete Event Simulation, Agent-based Simulation, Monte Carlo Simulation, Distributed Simulation, Intelligent Simulation.

- *Scheduling.* Modelling a production process to test potential job sequencing to optimise throughput and ensure orders are delivered on time. This is a broad subject area, including elements of throughput analysis and production capacity, resource allocation, workforce planning, and trade-off analysis between conflicting objectives such as delivery reliability and cost to manufacture.

  Modelling Methods: Discrete Event Simulation, Agent-based Simulation, Monte Carlo Simulation, Petri-net Simulation, Intelligent Simulation.

- *Strategy*. Modelling a business or system to simulate the effect of high-level policy change and proposed business strategies.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Agent-based Simulation, Monte Carlo Simulation, Simulation Gaming.

- *Supply chain management.* Modelling complex and interconnected supply chain systems to understand critical links, how inventory should be distributed, scheduling deliveries, and evaluating instability and robustness in a supply network.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Agent-based Simulation, Petri-net Simulation, Distribution Simulation, Simulation Gaming.

- *Workforce planning.* Modelling the impact of different shift patterns and staffing levels, the impact of training and cross-training (training employees in multiple jobs or processes so they can be more easily reassigned), and the impact of increased workforce versus investment in new equipment.

Modelling Methods: Discrete Event Simulation

- *Maintenance management.* Modelling proposed maintenance schemes and schedules against the predicted failures of equipment, to evaluate the trade-off between time and cost of the maintenance policy versus the likelihood and impact of a failure.

  Modelling Methods: Discrete Event Simulation, Monte Carlo Simulation, Virtual Simulation.

- *Knowledge management.* Modelling new product introduction and design, the learning curves of new processes or policies, as well as understanding the impact of organisation-level training.

  Modelling Methods: Discrete Event Simulation, System Dynamics.

- *Project management.* Modelling potential scenarios around project delivery and understanding which project management approach is most appropriate and understanding the propagated impact of delays or disruption.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Monte Carlo Simulation, Petri-net Simulation, Intelligent Simulation.

- *Organisational design.* Modelling changes to organisational structure and behaviour and analysing the effects on business outcomes.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Agent-based Simulation, Simulation Gaming.

- *Management training and education.* Modelling the outcomes of training and educating an organisation's management.

  Modelling Methods: Discrete Event Simulation, System Dynamics, Simulation Gaming, Distributed Simulation, Virtual Simulation.

- *Financial management.* Estimating the costs, risks, and outcomes of financial changes or capital acquisitions.

  Modelling Methods: Discrete Event Simulation, Monte Carlo Simulation.

- *Quality management.* A wide-ranging domain that models the outcomes of different quality paradigms on measurable outcomes in a manufacturing enterprise. Quality paradigms include continuous improvement, six sigma, total quality management, lean, and many more.

  Modelling Methods: Discrete Event Simulation, System Dynamics

5.3.4 Evaluating Modelling Tools

It can be seen that there exists a great many domains in which modelling and simulation can be applied in a manufacturing enterprise. Many software packages will cover multiple domains in a single package, or may be targeted to a specific sector. The range of modelling tools is constantly changing, and though some examples are listed below, it is important to evaluate a package against your own needs and to understand how well it fits within your business. Some criteria for evaluation include [8]:

- Handling heterogeneous information related to the design of products, processes and resources, and their operation during the manufacturing process. If a model cannot interpret the variety of information provided to it, the model cannot function.
- Integration of knowledge and information from different tools and techniques working at different levels of detail. Models and simulations work at different levels of accuracy and detail in accordance with their needs and requirements. Combining information from multiple sources and other models is required to make meaningful conclusions.
- Maintenance of the virtual representation of the manufacturing system so that it can be constantly synchronised with the physical counterpart. A model is only useful if it accurately models the physical world, so it is critical a tool makes it easy to update and change the model as the physical equivalent changes.
- Enable shop-floor engineers and technicians to make useful conclusions without the need for modelling and simulation specialists, to ensure the information is available to those who need it.
- Decrease of investment and operating costs, saving the company money if they take advantage of the model instead of adopting a “try and see” approach to process planning and design.

Modelling and simulations are highly effective tools for predicting the outcome of a manufacturing process, production line, or supply chain. However, most techniques and packages are best suited to specific areas, and have their own strengths and weaknesses. The best simulations and models are comprised of multiple smaller simulations and models sharing information to provide a more accurate level of prediction. To enable this approach, effective platforms support interoperability between digital factory tools. Not all platforms do however, and one should be cautious adopting or developing a tool that does not support the following [8]:

- A *common and standard data model* for representing entities related to production systems, resources, processes and products. Use of standards makes sharing data, or translating the data between different formats considerably simpler.
- A *shared data storage* accessible by different digital factory tools to retrieve and contribute with data. Standard data models are important, but the data needs to be stored somewhere that is easily accessible. Cloud data storage is increasingly common for this, allowing simple data access from anywhere.

- A *middleware* able to access shared data and correctly interpret/convert these according to the data model employed. It is less common for simulation models to communicate directly. Instead, middleware software acts as an intermediary.

Manufacturing simulation software must strike a balance between three key attributes:

- *Effectiveness*: How accurate are the results of the simulation, and how well does the software achieve the goals of the user?
- *Efficiency*: How long does it take the user to set up the simulation?
- *Ease of Use*: How easy is it to use the software? How steep is the learning curve?

An ideal simulation software will have excellent outcomes in all three aspects, but in practise different software packages will prioritise different attributes. In addition, simulation software usually specialises in different areas, such as line balancing, scheduling, supply chain modelling, inventory modelling, and many more. Choosing a package is about understanding what you are trying to model, and what attributes you want to prioritise. Other features you may want to consider when selecting a simulation package include:

- *Confidence:* Are there examples of successful implementations of this software package?
- *Cost:* How expensive is the package? Does support cost extra? Is it a one-time cost or a subscription?
- *Model runtime:* How long does the software take to produce an answer?
- *Reporting:* How are the results of the simulation reported, and is it a form that is useful to your business?
- *Support:* Is there a help desk you can ask for help? Is training available?

5.3.5 Example Manufacturing Modelling and Simulation Tools

The following is not an exhaustive list of modelling and simulation tools but serves as an example of the range of tools available. Most packages will cover several related domains described in section 5.3.3 so a single piece of software can solve multiple challenges in a common area. It's important to note that the field is constantly changing, and before committing time and money to a solution an enterprise should research currently available tools, and evaluate them with the questions and criteria described in section 5.3.4.

- *ARENA* is a simulation software that provides a fast, easy and intuitive way of building of a manufacturing process. It is based on drag-and-drop elements and structures with 2D and 3D visualisations. It also provides a dashboard for manufacturing optimisation, identify process bottlenecks, improve logistics and evaluate potential process changes.

- *AutoMod* is a 3D simulator able to model, analyse and emulate large and complex manufacturing, distribution and material handling systems by providing a simulation language. It has been largely employed in automation sector related to industries in automotive, airport, postal lines, warehousing and material handling.
- *Dassault Systèmes DELMIA* is a complete digital manufacturing and simulation software package, enabling users to analyse simulated and live production performance and document results for decision making. It features a collaborative 3D digital factory environment for process flow simulation and analysis, accuracy and profitability. DELMIA offers a flexible, object-based, discrete event simulation environment combined with visualisation and a colossal range of features and options.
- *Frepple* (Free Production Planning) is an open source supply chain planning software. It focuses on production planning constrained by machine/operator capacities, material availability and lead times. In addition, it provides inventory planning and demand forecasting.
- *Predator* offers stand alone or fully integrated manufacturing solutions for lean manufacturing and automation. Included in the suite are a variety of modelling and simulation tools primarily focussed on CNC machining. Predator Virtual CNC is an example of a resource domain simulation allowing for the verification and optimisation of CNC G-code offline before beginning the physical process.
- *FlexSim 3D* is a simulation software designed for modelling processes, including manufacturing, packaging, warehousing, material handling, etc. It imports relevant processing objects produced with CAD-based physical layouts. FlexSim allows end users to test all options in order to find the best combinations of operational characteristics to optimise performance and reduce costs. In addition, it provides support through accurate 3D animation and statistical reporting to both run "what-if" scenarios and make informed decisions.
- *Simio* provides a system composed by intelligent objects representing physical components such as forklifts and conveyors. This manufacturing solution encompass a set of industries like discrete manufacturing, automotive, consumer packaged goods, metals and plastics. The application areas are design of green field production plants, process improvement using Six Sigma and Lean Manufacturing, production planning and scheduling.
- *Tecnomatix* is Siemens' suite of manufacturing simulation technologies. It closely ties in with their automation products, allowing the results of simulations to be applied directly to (for example) PLC controllers, a process called "virtual commissioning".
- *Lanner Witness* is a process simulation package utilising both discrete and continuous event modelling, making it applicable to a wide range of modelling tasks. It also features an extensive library of 2D and 3D models of manufacturing resources, enabling both simulation and visualisation of proposed manufacturing lines.

5.4 Conclusions

Modelling tools are extremely valuable in the manufacturing domain. Manufacturing systems are extremely complex and extremely valuable, so understanding how they work, what the effects of proposed changes will be, and mitigating against risk and uncertainty are in term valuable pursuits. Use of these tools enables more complex calculations and simulations to be undertaken than performing the analysis by hand with mathematical formulae.

However, a manufacturing system is a dynamic, evolving, changing thing, and any model of a system must be accurate to give useful results. Even with modelling tools, creating an accurate useful model is time consuming and complex, and there is a risk that the final model may no longer accurately reflect the manufacturing system. There is also the issue of obtaining the data required to make the model accurate in the first place.

A *Digital Twin* is a live, automatically updating model of a manufacturing system which can perform simulations. Rather than being created offline by a worker, it is linked directly to the physical twin, and uses live data gathered from the real-world system to keep the model accurate and updated. Creating a digital twin is more involved that an offline model, but the cost of keeping it updated is significantly reduced. The next chapter will discuss digital twins, and how they can be implemented.

5.5 References

[1] S. Santoso, E. J. Powers and W. M. Grady, "Power quality disturbance data compression using wavelet transform methods," *IEEE Transactions on Power Delivery*, vol. 12, no. 3, pp. 1250-1257, 1997.

[2] M. K. Jeong, J. C. Lu, X. Huo, B. Vidakovic and D. Chen, "Wavelet-based data reduction techniques for process fault detection.," *Technometrics*, vol. 48, no. 1, pp. 26-40, 2006.

[3] G. Press, “Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says,” Forbes, 23 March 2016. [Online]. Available: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/.

[4] NIST/SEMATECH, “Engineering Statistics Handbook,” 30 October 2013. [Online]. Available: https://www.itl.nist.gov/div898/handbook/. [Accessed 2 11 2020].

[5] L. Monostori, J. Váncza and S. R. Kumara, “Agent-based systems for manufacturing,” *CIRP annals*, vol. 55, no. 2, pp. 697-720, 2006.

[6] P. Vrba and V. Marcík, “Simulation in agent-based control systems: MAST case study.,” *IFAC Proceedings*, vol. 38, no. 1, pp. 145-152, 2005.

[7] A. Oyarbide, T. S. Baines, J. M. Kay and J. Ladbrook, “Manufacturing systems modelling using system dynamics: forming a dedicated modelling tool.,” *Journal of Advanced Manufacturing Systems*, vol. 2, no. 1, pp. 71-87, 2003.

[8] T. Tolio, M. Sacco, W. Terkaj and M. Urgo, “Virtual factory: An integrated framework for manufacturing systems design and analysis.,” *Procedia CIRP*, vol. 7, pp. 25-30, 2013.

[9] Institute of Electrical and Electronics Engineers Standards Association (IEEE SA), *IEEE 1278.1-2012 - IEEE Standard for Distributed Interactive Simulation--Application Protocols*, 2012.

[10] M. Jahangirian, T. Eldabi, A. Naseer, L. K. Stergioulas and T. Young, “Simulation in manufacturing and business: A review,” *European Journal of Operational Research*, vol. 203, no. 1, pp. 1-13, 2010.

“The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein”

Digital Manufacturing Training System for SMEs (Digit-T)
Project ref: 2017-1-UK01-KA202-036807