Augmenting Cyber Physical Systems through Data Collection & Machine learning

Haritha Deepthi[1*], Siddiq Moideen[2*] and
Jacquline Dorothy[3†]

[1*]Department of Information Technology, PSG Polytechnic College, Coimbatore, 641004, Tamil Nadu, India
[2*]Department of Computer Engineer, PSG Polytechnic College, Coimbatore, 641004, Tamil Nadu, India
[3†]Department of Basic Science, PSG Polytechnic College, Coimbatore, 641004, Tamil Nadu, India

E-mail(s) Corresponding authors: harithadeepthibtech@gmail.com; siddiqmoideen07@gmail.com; Contributing author(s): jacqulinedorothy@gmail.com;

Abstract

Industry 4.0, destined to an astounding breakthrough in the field of Production & Manufacturing through leading-edge Cyber Physical Systems, Monitoring systems & Automation. The next big Industrial Revolution focuses on digitalizing the industries which is popularly known as Digital Twinning, any changes to the Digital Twin reflects the Real Physical World. Cyber Physical Systems are the key players in the 4th industrial Revolution. Cyber Physical Systems are amalgamation of Theory of Cybernetics & Mechatronics where Physical Plant i.e., Full-Fledged Hardware component Controlled/Monitored by Computational platform i.e., Computer-Based Algorithms moderated by Network Fabrics & Sensors. CPS integrates the dynamics of physical processes, software & networking. Where Components of Physical and Computational elements are deeply intertwined. The Backdoors of the system aren't robust enough to tackle modern-day Cyber Threats. Digital Twinning gives us an upper-hand in both Security and Production perspectives. Our paper focus on enhancing the production & security of the CPS through Machine learning approach, analysing the digital asserts statistically to set a favourable pay.

Keywords: Industry 4.0, Cyber Physical Systems, Digital Twin, Game Theory, Machine Learning, Simulated Systems, Security Mechanisms

1 Introduction to CPS

Cyber Physical Systems are multidisciplinary systems that conduct feedback control on widely distributed embedded computing systems through combination of communication, computation and control technologies. Modern CPS are able to realize the real-time, dynamic, safe and reliable collaboration with physical systems represented via embedded system. They are Integral mixture of existing network systems and traditional embedded systems. Where, Physical system data modules collect data by distributed field devices in CPS system, then pass data to the information processing layer as per the complete given tasks and demands of services by information processing technologies such as statistical signal processing, feedback control, data security processing and data uncertainty management. CPS interact with physical system through networks, the end system of CPS is normally traditional centralized tightly coupled embedded computing system, which contains a large number of physical systems composed of intelligent wired/wireless actuators & sensors.

The potential benefits of the convergence of 3C technologies for developing next-generation engineered systems that can be called Cyber Physical Systems are wide ranging and highly transformative via efficient computation, distributed sensing, high-level decision-making algorithms, control over wireless / wired communication networks formal verification technologies and multi-objective optimization; engineered cyber physical systems are in many societal critical domains such as construction, energy, transportation, and medical systems. Scientists and Engineers in this field have deep understanding of system and branches of mechatronics, biology, computer science and chemistry. Physical systems & Technical systems are developed and designed to be more & more reliable, efficient, smart, robust and secure.

With such high notions, the scope of CPS and integration of Cloud computing is about to bring the next big Industrial Revolution, Industry 4.0. The complete factory can be made into digital twin in the cyber space. Any changes in the cyber twin will reflect in the physical world which could boost the production & efficiency in the field of Avionics, Distributed robotics, Energy conservation, Process control, Smart structures, Defense systems, Critical infrastructure control, Assisted living, Environmental control, medical systems, Manufacturing and Traffic safety & control.

2 Industry 4.0

Industry 4.0, the next stage in the organization and control of the entire value stream along the lifecycle of a product. This cycle is based on increasingly individualized customer wishes and ranges from the idea, the order, development, production, and delivery to the end customer through to recycling and related services; which necessitates the consideration of many other issues that may occur in the upcoming new era, including standardization, safety and security, resource efficiency, new social infrastructure, work organization and work design, training, regulatory framework, etc. Technologies of nine aspects that power the transformation of the current industrial production to that of Industry 4.0 have been identified, which more or less have something to do with CPS. Fundamental here is the availability of all relevant information in real-time through the networking of all instances involved in value creation as well as the ability to derive the best possible value stream from data at all times. Connecting people, objects and systems leads to the creation of dynamic, self-organized, cross-organizational, real-time optimized value networks, which can be optimized according to a range of criteria such as costs, availability and consumption of resources. Industry 4.0 is a collective term for technologies and concepts of value chain organization.

Within the modular structured smart factories of Industry 4.0, CPS monitor physical processes, create a virtual copy of the physical world and make decentralized decisions. Over the IoT, CPS communicate and cooperate with each other and humans in realtime. Via the IoS, both internal and cross-organizational services are offered and utilized by participants of the value chain. In a smart factory, products and machines communicate with each other, cooperatively driving production. Smart products can refer to objects, devices, and machines that are equipped with sensors, controlled by software and connected to the Internet. Industry 4.0 will give rise to novel CPS platforms geared toward supporting collaborative industrial business processes and the associated business networks.

CPS provide a critical support to the vertical and horizontal system integration. The combination of CPS and the industrial IoT enables the creation of the IoT and IoS. Moreover, the widespread application of CPS means the generation of industrial big data, which requires cloud technology and big data analytics for storage and analysis. The virtual world of CPS consists of a great variety of models of the production facilities, for which simulation can play an important role. Augment reality technology is required for operators to interact with CPS. Additive manufacturing and robots are essential parts of the CPS-based manufacturing systems of Industry 4.0. Within a smart factory, all the physical production elements in the physical world have a cyber twin in the virtual world. The physical and virtual worlds as well as the physical assets and cyber twins in them are seamlessly connected to achieve the global optimization of production within a smart factory. Moreover, within a value network, multiple factories are horizontally integrated, i.e., the physical assets and the cyber twins are, respectively, integrated to enable optimized decision-making across the value network. The integration through the value network will give rise to CPS platforms, within which things, services, “data,” and “people” are connected over the Internet.

3 Cloud Manufacturing

Cloud Manufacturing is a new networked manufacturing paradigm that organizes manufacturing resources over networks according to client’s needs and requirements to provide a variety of on demand manufacturing services via networks & cloud manufacturing service platforms and a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable manufacturing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. In the cloud manufacturing mode, providers can supply their requirements to the platform for requesting services ranging from product design, manufacturing, testing, management, and all other stages of a product life-cycle to their manufacturing resources, which will be transformed into services and then pooled into the cloud manufacturing platform which is an advanced manufacturing business model that focuses on issues that are directly related to manufacturing and pays less attention to issues like urban production and demographic change. A cloud manufacturing platform has a multilayer architecture, including resource layer, virtual resource layer, global service layer, application layer, and interface layer.

The implementation of different layers requires different technologies. IoT, virtualization, and servitization technologies are needed to sense manufacturing resources and transform physical resources into virtual resources in the virtual resource layer. The core technologies for global service layer are cloud computing, service-related technologies and semantic Web technology. In the interface, it is the human–machine interaction technology that plays an important role. Certainly, the high-performance computing technology and advanced manufacturing model and technology are also essential. Many other supporting technologies such as big data are also necessary for the complete implementation of a cloud manufacturing platform.

Industry 4.0 represents a highly digital and networked manufacturing paradigm that satisfies enterprises' digital manufacturing and collaboration requirements between them and business partners, while cloud manufacturing can effectively satisfy the sharing and collaboration requirements of enterprises in a convenient and agile way. Which relies on Process and Digital Twin Modeling Solutions with modern developments of communication protocols and connection architectures, numerous initiatives have attempted to model physical machines, processes, and results with computational methods.

The concept of Digital Twin modeling is defined as "virtual machine tools of physical machines for cyber-physical manufacturing by using sensory data and information fusion integration techniques". These "Digital Twins" of the physical world have been developed to different extents, each modeling and predicting different areas of the production process. Digital Twin initiatives have been implemented to address a wide variety of modeling scenarios. Simple solutions range from open-loop monitoring of process data while more complex implementations include AI-powered feedback to correct errors in a manufacturing process. In contrast to the planning of engineering changes, decisions in production control need to be taken in real-time with limited information. The concept of self-control in a decentralized production system is based on the ability of several elements of the system to act and decide autonomously. In contrast to that, in the centralized approach, planning is accomplished by a superordinated planning entity. Therefore, especially for decentralized production control with a multitude of decision-makers, a fast recognition of data patterns is necessary to adapt the behaviour and decision rules of the acting elements.

Many Digital Twin methods integrate models of physical systems with information and result from both computer simulations and CPS connected data. These two sources of information are combined to provide a more accurate method of predicting the system's behavior and production results. The concept by summarizing the technology needed for a complete Digital Twin to predict microstructure development, residual stresses, and part defects in additive manufacturing. Through the integration of temperature and velocity fields from both numerical simulation and experimental measurements.

Digital Twin model for directed energy deposition in additive manufacturing provides a more accurate cooling rate and temperature gradient prediction than traditional conduction calculations. Looking to the future of Digital Twin modeling where complete factories of CPSs can be connected and monitored provides exciting opportunities for process development. With more advanced capabilities to link and correlate manufacturing data from different sources, demonstrate a future application for Digital Twin modeling in partial and parallel disassembly sequence planning for products. Implementation of the near real-time analysis for product information, timing information, and upstream downstream events would provide beneficial flexibility and adaptability for assembly planning methods. These two types of platforms are different in the aim, core technologies, operational mode, and platform architecture.

3.1 Simulated Production System

A simulation is used to acquire production data of a virtual factory. All of the factory's components like the machine tools and the product workpieces are virtual as well. Each machine has a machine type and each product has a product type, so there are different types of virtual products and machines. The work plan, i.e., the order of operations required to produce a final product, is given externally and cannot be changed. Thus, the product type defines which operations need to be processed sequentially to finish the product, while the machine type defines the operation the machine is capable of the ability of machines of a certain type to perform an operation with specific requirements is encoded into the machine type.

This leads to different process times, even for operations with the same production technology. The required setup times for each operation are included in the resulting processing times. To finish the production of a product, all of its operations need to be processed in order, while each operation takes a certain time.

Optimal solution cannot be calculated in a feasible time, but heuristics can be evaluated efficiently to come close to an optimal solution of product distributions. This work uses the heuristic of always choosing the machine that will have processed the product's individual operation first. Other heuristics could be considered as well, but since the choice of heuristic is not important for the demonstration of the presented analysis tool, the described heuristic is chosen out of simplicity reasons. Another issue is the optimal arrangement of machines in the factory. This problem is also computationally very expensive and cannot be solved optimally in a feasible time for a larger number of machines. Therefore, the arrangement of machines in the presented example was chosen as demonstrated by simply distributing groups of identical machines within the factory.

The transportation times of products between the machines depend highly on the arrangement of those machines. Since the production batches in the used example are very large, the resulting transportation times are very small in comparison.

Simulation and evaluation software for products and material flows can be applied in order to analyse processes and their changes. To examine process chain and information flow consistency, the available data need to be visualized in an aggregated manner. The scalability of the data is required to allow the user to focus on single products or machines and to well defined time steps of special interest. An essential requirement is to guide the user to the most interesting features of the regarded system and to show critical issues. Therefore, comparative and interactive data highlighting integrated in the spatial context of the factory are needed. The amount and complexity of data can still be managed manually, a test field based on a multi-agent system to test several self-organization concepts against each other in a real sized but virtual environment.

3.2 Security Mechanisms

classified into the following four categories: asset-centric security, network-centric security, data-centric security, and user-centric security. Asset-centric security refers to the protection of hardware and software associated with the asset. Network-centric security covers security related to communication channels between two or more assets. Data-centric security refers to the protection of data throughout its lifetime (creation, transmission, storage, updating) in a PRP. Clearly, the lifetime of protection in data-centric security is longer than network-centric security and this depends on the design stage in a sPRP. User-centric security refers to the different security requirements from the stakeholders in a PRP.

The process of combining security mechanisms poses a complex challenge as it requires a proper combination of security mechanisms at every stage in a PRP. One of the possible ways in which different security mechanisms can be integrated into multiple independent layers while achieving multi-objective security.

Combination of security mechanisms may be necessary to simultaneously achieve multiple security attributes. The strength of security in such combinations depends on its weakest security mechanism. In order to overcome such issues, security experts suggest to offer system security in multiple independent layers, i.e., integrate different types of security mechanisms for each security objective. Designers need to be careful while combining security mechanisms, as it may lead to undesired redundancy, or security gaps that make the system vulnerable to attacks.

The major security objectives:

- Confidentiality refers to the protection of the asset from unauthorized stakeholders.
- Integrity refers to the protection of the asset from unauthorized modifications.
- Availability requires the asset to be accessible, operational when it is needed.
- Accountability refers to the identification of the responsible entities for the various actions. In this section, we briefly discuss the common security mechanisms in each category and identify the security objectives achieved by each.

3.2.1 Assert-Centric Security

Asset-centric security subsumes all aspects that relate directly to the security of a device or asset. The security mechanisms in this category are further classified into hardware security and software security.

Hardware Security focuses on establishing the validity of physical components. For traceability, counterfeit detection, or liability reasons, typically through leveraging randomness that is extrinsic or intrinsic to the part which can be defended through Physical Marking/Watermarking and Crypto-processors.

Software Security solutions are aimed at protecting software against attacks such as tampering or causing run-time misbehavior through invalid inputs that the software erroneously treats as legitimate. The latter exploit the existing software bugs and vulnerabilities which are inadvertently introduced during the software development cycle which can be defended through Antivirus Software, Tamper-resistant software and Digital watermarking

3.2.2 Network-Centric Security

Network-Centric Security aims at providing secure communication among different assets in the network. These approaches are particularly important given the rise of cloud-based design and manufacturing. The common security mechanisms are hit through **HTTPS**, **Secure Multicasting**, **Virtual private networks**, **Blockchains, Onion routing** and **Intrusion detection systems**.

3.2.3 Data-Centric Security

Main aim of data-centric security solutions is to protect data throughout its lifetime starting from its creation, transmission, storage, and updating which can be achieved through **Steganography** and **Encryption** which bags up Symmetric Encryption, Asymmetric encryption, Secret sharing, Secure multiparty computation, Cryptographic hash functions and Digital certification techniques.

3.2.4 User-Centric Security

Advancements in sensor technologies and communication methods have made it possible to monitor product use in real time. The users of the product could be either individuals or enterprises. Monitoring the product use could raise concerns related to privacy and data ownership. Which can be resolved through **Authorization, Authentication, Access control** and **Anonymity**

The security mechanisms help designers to come up with a security architecture that prevents an attack. There is another branch of security that deals with situations after an attack is discovered. Forensic analysis includes damage estimation, identifying the method of attack and its source.

3.3 Data Collection

Graph Database for Industrial Data Analysis, Due to their advantages including scalability, efficiency, and flexibility, GDBs are widely adopted in various industry-related applications and use cases such as network operations, fraud detection, and asset and data management. The authors have proposed a new object tracking approach for surveillance applications. The GDB approach is selected to contribute to the scalability of the proposed scheme and support the required connectivity analysis for the object tracking. Moreover, relationships in social networks have been modeled using a GDB for structural information mining and marketing. Conversely, GDBs are also deployed in business solutions for scenarios with multiple large data sources, which require distributed processing in decision-making for various problems such as fraud detection, trend prediction, and product recommendation.

GDBs perfectly improve the data management, include path finding with weighted and time-related path properties, mapping dependencies of various system components to capture potential weak points, and communications between various networked elements. Conversely, query languages are used to extract data, including traversing the database, comparing node properties, and subgraph matching.

The advantages of deploying GDB include having a more natural approach of data modeling and keeping data properties connected to nodes and relationships. Moreover, GDBs offer graphical and visualization interfaces to data and are able to keep the time-related information of events through various graph paths. Also, an extended list of applications and implementations of GDBs is presented in to show their use on enterprise data, social networks, and determining security and access rights. It was found that GDBs provide the much-needed structure for storing the data and incorporating a dynamic data model.

Heterogeneous data: The collected data from industrial wireless communications system is heterogeneous in different aspects as follows:

- Different sources: We collect network data at various network nodes in the system. Also, collected data using wireless sniffer describe the wireless physical environment. Data from the supervisor controller are also collected, which include the system states and the supervisory commands. Data from the robots are used to describe the physical actions taken.
- Different formats: The data include different file formats such as packet capture (PCAP) files, and data that come from different PLC and robot controllers are stored in the format of comma separated value (CSV) files. Another example is the time stamp format from different devices.
- Different rates: Data packets can be both periodic and event driven. Also, the robot state feedback is periodic with a different update rate than the update rate of the PLC state.

Entities are interrelated: This is the main requirement and challenge in this work where the goal is to obtain the direct one-to-one connection between physical actions and their corresponding entities including network activities, the physical wireless environment through sniffer reports, and the physical system state.

Various entity types: The data model will consider two types of system entities, namely, dynamic and static. The class of static entities covers testbed setup profiles, which contain testbed components, network interfaces, and their settings. These entities are normally predetermined or collected in the initialization of each measurement. The class of dynamic entities captures various system events such as machine status reports, network traffic, and information flows in the testbed. These entities are dynamically added into the data set whose quantities and properties are determined by the measured data.

Data model with multiple abstraction mechanisms: The considered data model and the corresponding queries should encompass multiple levels of abstraction including traffic data level, physical hardware level, physical environment level, physical actions level, and the interactions between these various levels. The network database system must allow for the categorization and labeling through these levels.

Time travel queries: The data model and the resulting database should allow for direct querying for temporal variations of the studied entities. Hence, temporal relationships between data packets and the corresponding physical actions should be stored and directly accessible.

Efficient path and relationship queries: Given the requirement of having interrelated nodes, the query language should allow for path and relationship queries to directly extract this information. These types of queries are used for calculating various system metrics and hence should be performed in an efficient manner.

4 Industrial Data Analytics

Industrial data analytics play an essential role in achieving the smart factory vision and improving decision-making in various industrial applications. Five main industrial data methodologies are generally studied including highly distributed data ingestion, data repository, large-scale data management, data analytics, and data governance. Industrial data processing offers valuable information about various sections of industrial applications including inefficiencies in industrial processes, costly failures and down-times, and effective maintenance decisions.

The industrial data analytics are generally deployed for improving factory operations through improving machinery utilization and predicting production demands, improving product quality by analysing market demands and reducing defective products, and enhancing supply chain efficiency by analysing risk factors and making accurate logistic plans and schedules.

4.1 Visualization

Helps the user in building a mental map of the production data. Since all views of the presented system always show data for the same time window or selection, a cognitive transition from one view to the others is straightforward. After the virtual manufacturing system is simulated once, the whole tool and its views work in real-time to provide flexibility of interactions to the user. The presented tool can be used to analyze virtual factories, provide user guidance for later optimization or comparison, and help in decision making.

The overall goal is the optimization of the production process with respect to a diversity of parameters. Still, this optimization cannot be done fully automatically because of its high computational complexity. Although the optimal solution is unknown to users, the presented tool can be used to iteratively improve factory settings. By that, users are enabled to approximate an optimal solution, thereby finding a sufficient solution and gain a certain confidence in their production process.

4.2 Game Theory Implementation

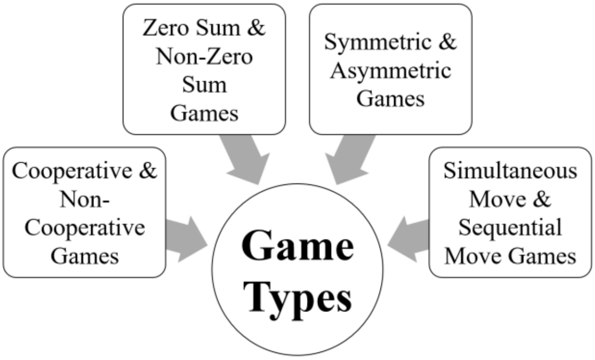
Game Theory is primarily a mathematical framework analyses the decision-making of a player based on how they expect other players to make a decision i.e., determining optimal rational choices given a set of circumstances which can be applied in many fields such as economics, politics, computer science, biology, philosophy & so on. Game theory depicts the game played between different players and the strategies of each player.

A game can be defined as interaction of different players according to a set of rules. Players may consist of individuals, machines, parties, companies or associations. The results of game theory depend upon behaviour of every other player present in the game and not only the current player.

Due to this reason, this approach is extremely scalable and versatile. The outcome of game theory also depends on the estimated payoff by each player before making decisions, which is a measure of the satisfaction obtained by each player by making that decision. Therefore, the players will perform actions and take decisions that would provide them the maximum payoff.

4.2.1 Game Types

In game theory, there are different types of games that help us analyse different problems. They are categorised in the basis of number of players involved, cooperation among players & symmetry of the game.



The prisoner's dilemma is one perfect example; how game analysed in game theory which shows why two completely rational individuals might not cooperate, even if it appears that it is in their best interests to do so.

Prisoner's dilemma is a situation where individual decision-makers always have an incentive to choose in a way that creates a less than optimal outcome for the individuals as a group.

Two members of a cartel named Robert & Walter were arrested and imprisoned. Each prisoner is in solitary confinement and they have no means of communicating with each other. The prosecutors lack sufficient evidence to convict the pair on the principal charge.

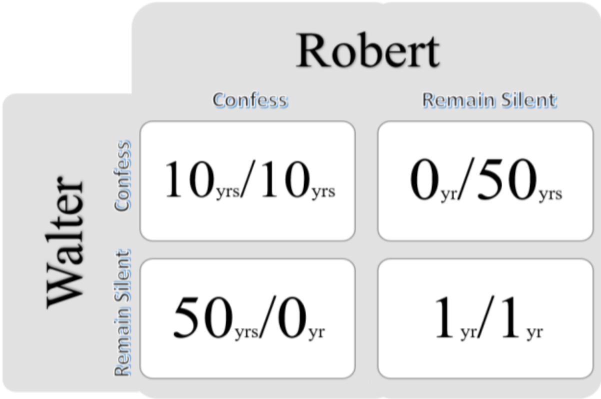


Figure 13: Decision Matrix

Prosecutors hope to get both of them sentenced on a lesser charge. Simultaneously, the prosecutors offer each prisoner a bargain. Each prisoner is given the opportunity either to betray the other by testifying that the other committed the crime or to cooperate with the other by remaining silent. The options offered are:

• If Robert & Walter, both betray each other i.e., if they both confess, each of them serves 10 years in prison.

• If Robert betrays Walter but Walter remains silent, Robert will be set free & Walter must serve 50 years in prison.

• If Walter betrays Robert but Walter remains silent, Walter will be set free & Robert must serve 50 years in prison.

• If Robert & Walter both remain silent, both of them will only serve 1 year in prison.

Note that it is implied that the prisoners will have no opportunity to reward or punish their partner other than the prison sentences they get and that their decision will not affect their reputation in the future.

We'd feel both remaining silent would be the best option. But the individuals won't opt for it; Both of them will betray each other i.e., they would confess. Because that's the human psychology, confession seems to the best option for both the parties. Individually the prisoner clings on luck that he would be set free, if the other prisoner does not confess & fears that what if I remain silent and the other confesses.

Because betraying a partner offers a greater reward than cooperating with them, all purely rational self-interested prisoners will betray the other, meaning the only possible outcome for two purely rational prisoners is for them to betray each other and that is Nash Equilibrium A.K.A. Optimal state for all the participants.

The prisoner's dilemma game can be used as a model for many real-world situations involving cooperative behaviour. The label "prisoner's dilemma" may be applied to situations not strictly matching the formal criteria of the classic or iterative games; for instance, those in which two entities could gain important benefits from cooperating or suffer from the failure to do so, but find it difficult or expensive but not necessarily impossible to coordinate their activities.

4.2 Machine Learning Approach

////To Be Filled////

Examples for engineering changes are the reconfiguration, addition, substitution, or removal of production equipment, e.g., machine tools, in a manufacturing system. They usually have extensive impacts on the manufacturing system due to the manifold interrelationships among production objects and hence need careful analysis and planning before implementation.

The methods of industrial data analytics can be split into different categories such as descriptive, diagnostic, predictive, and prescriptive analytics. Descriptive and diagnostic analytics are responsible for analysing historic data and the causes of events and behaviours. Predictive and prescriptive analytics require more processing power, anticipate the trends of data, and deploy the historical data in making decisions to achieve production goals. A platform for performing industrial big data analysis is presented where the performance requirements are introduced to achieve a cost-effective operation.

The data available in a cyber-physical production system can be used to make production systems more flexible as the transformability of the system to engineering changes on medium- or long-term perspective achieved by decentralized production control on a short-term view. It is obvious that the vast amount of data is not useable without refinement. scalability is a further required functionality that enables to select single hotspots and establish detailed comparisons between machines, products, or time steps. Embedding the data into the spatial context of the factory is needed to give the user a realistic and intuitive understanding of the factory and its performance.

Identify patterns on an aggregated data level to derive the system's sensitivity to changes of decision routines. As a consequence, aggregated views displaying the overall performance in a spatial context, and detailed views representing the perspective of single elements, are necessary to understand an entire system. To summarize, one major issue for data visualization is to be intuitively understandable. Therefore, an interactive guidance for the user is required, which makes it easy to find interesting features in a data set. To get a quick but comprehensive overview of the status of the production system, different perspectives on an aggregated level are needed. These have to be interlinked to navigate through the perspectives.