

Article

Not peer-reviewed version

Innovative Data Models: Transforming Material Process Design and Optimization

[Amir M. Horr](#)^{*}, Matthias Hartmann, [Fabio Haunreiter](#)

Posted Date: 29 May 2025

doi: 10.20944/preprints202505.2378.v1

Keywords: real-time modelling; data models; material processes; machine learning; data science



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Innovative Data Models: Transforming Material Process Design and Optimization

Amir M. Horr *, Matthias Hartmann and Fabio Haunreiter

LKR Light Metals Technologies Ranshofen, Austrian Institute of Technology, Vienna 1210, Austria

* Correspondence: amir.horr@ait.ac.at; Tel.: +43-(0)50550-6918

Abstract: As the use of data models and data science techniques in industrial processes grows exponentially, the question arises: to what extent can these techniques impact the future of manufacturing processes? This article examines the potential future impacts of these models based on an assessment of existing trends and practices. The drive towards digital-oriented manufacturing and cyber-based process optimization and control has brought many opportunities and challenges. On one hand, issues of data acquisition, handling, and quality for proper database building have become important subjects. On the other hand, the reliable utilization of this available data for optimization and control has inspired much research. This research work discusses the fundamental question of how far these models can help design and/or improve existing processes, highlighting their limitations and challenges. Furthermore, it reviews state-of-the-art practices and their successes and failures in material process applications, including casting, extrusion, and additive manufacturing (AM), and presents some quantitative indications.

Keywords: real-time modelling; data models; material processes; machine learning; data science

1. Introduction

The emerge of new data science techniques for material processes has pushed the process modelling and simulations concepts beyond the state of art, triggering fast and reliable design loops and efficient process controlling. This paper presents an overview over the generation of data solvers and interpolators, their rigorous validation concepts, and their performances for real-world process applications. The mathematical and computational features of these data models were briefly elaborated herein, and their limitations, performances and challenges are highlighted. The applications of these real-time models for popular material processes such as casting, extrusion, and AM have also been shown using real-world case studies. The goal here is to show the predictive power of these models along with their limitations and short comings for material process simulations.

With the emergence of numerical simulations in the last century, process modeling and optimization have undergone revolutionary changes, moving towards reducing design and optimization loops that were previously based on experimental trial and error. By incorporating more complexities into these simulations, multi-physical and multi-scale phenomena can be modeled, albeit at the expense of greater computational efforts and costs. Powerful computing facilities, including cluster and cloud computing, along with advanced algorithms, have been developed to meet the substantial computational demands of these sophisticated numerical models. Most manufacturing processes are inherently thermal events involving cooling and heating, and the characteristics of the final products are highly dependent on the evolutions of lower-scale phenomena, such as microstructure formations. Consequently, many numerical techniques developed for these processes are inherently capable of handling phase changes, bridging scales, managing thermal evolutions, and calculating mechanical properties [1].

Further challenges of these numerical models include the transient nature of these processes, which require consideration of startup and shutdown conditions, as well as changing conditions

during the processes [2]. Additionally, some manufacturing processes are generative in nature, where parts are built up over time (e.g., casting, AM). Typically, numerical simulations for these processes are set up using a predefined fixed domain size with a fixed discretization grid, where initial boundary, and external conditions can be enforced [3]. This can be particularly challenging, as generative processes may require dynamic domain sizes, meshes, and boundaries to accurately model the transient process. Consequently, some numerical modeling schemes tend to use steady-state, pseudo-steady-state, or even superposition of steady-state snapshots to construct a transient replica of the actual process.

The most critical issue with today's state-of-the-art numerical simulation schemes for transient processes is generating reliable and accurate results within a reasonable computational time while maintaining adequate detail and capturing the relevant physics and phases during the processes. Although powerful numerical techniques, such as finite element (FE) and computational fluid dynamics (CFD), have been repeatedly used to address the multi-physical, multi-phase, and multi-scale nature of material processes, achieving reasonable results still requires significant computational effort. Furthermore, for processes with multi-scale aspects (e.g., microstructure evolution modeling) and phase changes, capturing phenomena at lower scales and phase transitions, along with their associated thermal and stress histories, necessitates solver interactions and interfacing. This requirement limits the time efficiency and increases the computational cost of these interactive simulations, often necessitating fine discretization resolutions.

This research investigates the application of data science techniques in modeling material processes, encompassing both transient and steady-state conditions. It addresses critical issues such as data availability, data quality, the selection of data solvers and interpolators, snapshot processes, validation procedures, and the development of generative data models and databases. Real-world case studies are conducted to illustrate these concepts. Additionally, the study elaborates on ongoing research in process data acquisition and data generation, highlighting best practices for selecting the most suitable data solvers and interpolators, and discussing practical considerations.

2. Methodology- Data Models for Processes

The use of data science techniques in engineering process applications has grown rapidly in recent years, driven by the need to quickly predict process trends and understand complex process parameters more accurately and efficiently. Traditionally, analytical and empirical methods, along with physical testing, were employed for material process modeling. These methods later evolved into sophisticated numerical approaches. As computational power increased over the decades, the field of numerical process simulation began to incorporate more detailed mathematical models to represent various physical states, phases, and phenomena at different length scales. The implementation of data science techniques in material processes has started to revolutionize process modeling by addressing several critical functions. Firstly, they provide fast and real-time predictions under various process conditions, which can be used to develop real-time control and optimization schemes for digital twinning. Secondly, they help in understanding the evolution of fundamental properties of materials during processes, resulting in better quality control for products. Thirdly, these models can be combined with physics-based, analytical, and numerical approaches to rely more on fundamental material science principles, utilizing machine learning (ML) and data training to analyze large and complex process databases.

Over the past two decades, the rapid advancement of data science has led to the emergence of powerful data-driven and hybrid physical-data-driven modeling techniques in manufacturing. Several frameworks have been proposed to integrate these methods into process engineering, including Data-Driven Engineering Design (DDED), Data-Driven Process Systems Engineering (PSE), and Model-Based Systems Engineering (MBSE) [4–8]. In addition, hybrid approaches such as Data-Model Fusion (DMF) for smart manufacturing have been introduced by Tao et al. [9], while Dogan et al. [10] provide a comprehensive review of ML and data mining (DM) applications in manufacturing, covering both supervised and unsupervised learning techniques. Wang et al. [11]

discuss the role of big data analytics in manufacturing, including frameworks, technologies, applications, and challenges. Ghahramani et al. [12] explore AI-based process modeling and optimization, proposing evaluation strategies for intelligent manufacturing systems. Furthermore, Sofianidis et al. [13] highlight the use of explainable AI (XAI) to enhance transparency and trust in AI-driven production environments. The application of reduced-order data models for real-time process prediction has also been explored by Horr et al. [14–17], focusing on the integration of data-driven techniques for process optimization and control.

Traditional modeling approaches for metal manufacturing processes can be categorized into several classes including steady-state and transient modelling, multi-physical and multi-phase modelling, multi-scale and interactive modelling, and hybrid analytical–numerical modelling. The integration of data science is transforming how these models are developed and applied. The evolution of data modeling techniques can be summarized as follows:

1. Data Models for Steady-State and Pseudo-Static Processes - For processes that can be approximated as steady-state or pseudo-static, spatial variations in process parameters can be used to construct 1D, 2D, or 3D data models. These models may incorporate multi-physical, multi-phase, and multi-scale phenomena (e.g., microstructure evolution), with transient effects approximated using statistical or averaging techniques.
2. Time-Series Data Models for Transient and Temporal Processes - These models capture the evolution of materials, boundary conditions, and process parameters over time. When multi-scale effects are included, appropriate length and time scales must be defined. The data modeling techniques must be agile enough to handle both gradual and abrupt changes in material properties (e.g., melting, solidification), thermal inputs/outputs, and control parameters.
3. Hybrid Analytical–Data and Generative Models - These models combine physical understanding of the process with trained data-driven components. They are particularly effective for systems where the underlying physics is known, but complex interactions are better captured through data science. Hybrid models may use single or split databases, integrating analytical and learned data to enhance prediction accuracy and computational efficiency.

Figure 1 shows the framework for the data strategy for process modeling and its role in digitalization of manufacturing industries. The framework offers a structured approach for managing and leveraging process data as a strategic asset for optimization and control of manufacturing processes throughout the production chain. It shapes how industrial data is collected, stored, processed, accessed, and used to support active controlling and optimization. This data strategy helps the implementation of process digital twins, enhances predictive power through AI-based techniques, and facilitates operational efficiency, and drives innovation in smart manufacturing environments.

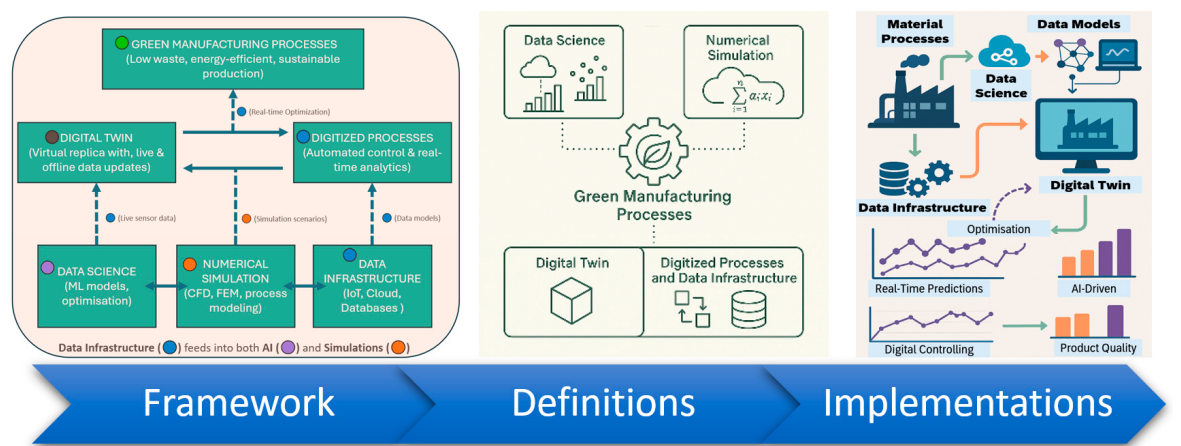


Figure 1. Framework for data strategy in manufacturing processes: foundational concepts and workflow implementation for developing process digital twins.

3. Steady State Processes – Data Strategy

Most manufacturing processes exhibit some degree of transient behavior, with process parameters and conditions often changing over time. Factors such as startup conditions, shifts in boundary conditions, abrupt disturbances, and planned ramp-up or ramp-down phases naturally introduce transient effects that can significantly influence the quality of the final product. However, in many widely adopted manufacturing processes, these transient phases are typically short-lived. Once stabilized, the process can operate under consistent conditions, making it suitable for steady-state modeling to support real-world predictive applications. Developing such models involves a structured methodology encompassing spatial data acquisition, strategic planning, data storage, model development, validation, and deployment [17,18]. For processes such as continuous casting, long wire extrusion, and the production of large-scale additive manufactured components with uniform geometry and process parameters, steady-state modeling approaches can be effectively applied. These approaches help streamline model development, reducing both time and effort while maintaining predictive accuracy.

3.1. Basic Development Plan

The development of data models for steady-state manufacturing processes can be structured into the following key stages:

- **Data Strategy:** A comprehensive data strategy begins with the structured collection of data from experiments, simulations, and validated literature sources. This includes capturing spatial variations within the modeling domain. For sensor-based and live data, a robust data acquisition system should be employed, incorporating filtering and translation tools to record both the spatial location of data sources and the corresponding filtered inputs. Prior to semantic data storage—ideally aligned with a predefined ontological framework—essential preprocessing steps must be performed. These include data cleaning, normalization, anomaly detection, and mapping to ensure consistency and quality.
- **Model Development:** In steady-state modeling, transient effects are either disregarded or represented through averaged or uniformly distributed approximations across the modeling domain. The choice of model type depends on the complexity and nature of the process and may include:
 - Pure data-driven models
 - Physics-informed or hybrid models (analytical, numerical, and data-driven)
 - Generative models for synthetic data generation

For complex manufacturing processes with high-dimensional data spaces, specialized combinations of solvers, interpolators, and machine learning algorithms are often required to achieve accurate predictions of product quality.

- **Real-Time Data Modeling:** Real-time predictive modeling can leverage various solver technologies, including:
 - Eigenvalue-based solvers
 - Regression and clustering algorithms
 - Support Vector Machines (SVMs)

To enhance prediction accuracy for new process parameter sets, advanced interpolation techniques such as Kriging, Radial Basis Functions (RBF), and Inverse Distance Weighting (IDW) can be employed [19,20]. In high-dimensional data scenarios, additional machine learning routines may be necessary to further train the model, reduce prediction error, and improve data fitting.

- **Model Training and Validation:** Training and validating data models for complex material processes—such as casting, extrusion, and AM—require careful consideration due to their multi-physical and multi-phase nature. In some cases, split databases are used to bridge different

length scales and to train parameter-specific datasets. Various training techniques can be applied, including:

- Neural networks
- Regression models
- Genetic Algorithm Symbolic Regression (GASR)
- Tree-based models

Model validation is conducted across the entire modeling domain, including normal, near-boundary, and extreme conditions. Performance metrics and accuracy indices are calculated, and normalized errors are assessed using predefined Design of Experiments (DOE) scenarios to ensure robustness and reliability.

These data models have been successfully applied to steady-state manufacturing processes across various domains, including thermal and mechanical systems, energy efficiency analysis, and microstructural evolution [14–17]. Figure 2 illustrates the foundational development plan for modeling steady-state manufacturing processes within the broader digitalization framework.

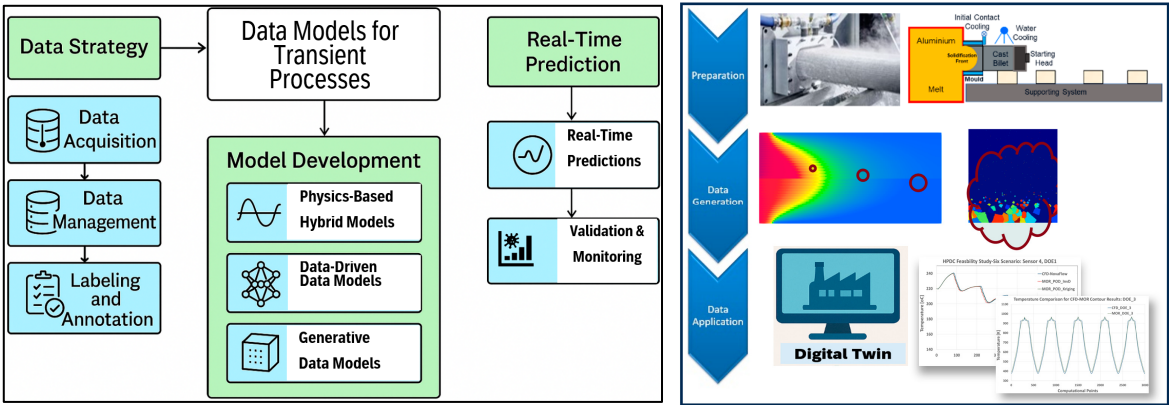


Figure 2. Basic foundational development plan for modeling steady-state manufacturing processes within broader digitalization framework.

3.2. Process Case Studies

To explore the application of real-time data models in manufacturing processes, two real-world case studies—casting and extrusion—are briefly discussed to illustrate the practical implementation of data modeling techniques. These studies focus on steady-state conditions, excluding the transient behaviors typically observed during process start-up. The evolution of thermal fields in both processes is analyzed, and spatial temperature variations at selected points are predicted.

Casting Process: Developing a digital twin or shadow of the casting process enables the evaluation of various process scenarios under different sets of parameters and boundary conditions. Casting is inherently a multi-physical, multi-phase, and multi-scale process, where macro-scale thermal fields significantly influence solidification behavior and microstructural evolution at smaller scales. Creating real-time data models for casting presents several challenges, particularly in data generation, handling, training, and interpretation across different length scales. To build the casting data model, appropriate solver technologies and data interpolation routines were employed. The following steps outline the methodology used in this case study:

1. **Simulation Model Calibration and Validation:** The CFD simulation model was calibrated and validated using experimental measurements.
2. **Definition of Prediction Objectives:** Key prediction goals—such as thermal and microstructural events—were defined and implemented for a laboratory-scale semi-continuous casting process.
3. **Scenario Development:** Variations in process parameters were considered to define realistic operational scenarios, forming a snapshot matrix.

4. **Simulation and Data Generation:** Using the calibrated CFD model, snapshot scenarios were simulated with open-source software to generate a small database for model training. Additional DOE scenarios were simulated for validation purposes.
5. **Data Model Construction:** Real-time data models were iteratively developed using a suitable combination of solvers and interpolators. Machine learning techniques were also applied to enhance model performance.
6. **Performance Evaluation:** The accuracy and reliability of the data models were assessed using DOE-based validation scenarios.
7. **Integration into Advisory Framework:** Finally, the data models were customized for integration into an existing web-based casting process advisory system.

For the numerical simulations of the snapshot matrix scenarios, fixed-domain fluid-thermal CFD simulations were employed to analyze thermal evolution and solidification behavior during the casting processes. The open-source solver directChillFoam [21], built on top of the OpenFOAM simulation framework [22–24], was used for melt flow modeling. The initial CFD validation setup was based on a previous project utilizing directChillFoam, where simulation results were benchmarked against experimental data from Vreeman et al. [25]. The casting material was defined as a binary alloy, Al–6 wt% Cu. Solidification modeling incorporated temperature-dependent melt fraction data (e.g., from CALPHAD routines), while solute redistribution followed the Lever rule [26]. For thermal modeling, heat transfer coefficients (HTCs) at the mold–melt interface (primary cooling) were locally averaged based on the solid fraction. Secondary cooling at the billet–water interface was modeled using tabulated HTC values derived from the correlation proposed in [27]. All snapshot scenarios were simulated with a runtime of 2,000 seconds—more than twice the time required to reach quasi-steady-state conditions. Time-averaged values of temperature and melt fraction over the final 1,000 seconds were recorded at all cell centers, along with their spatial coordinates (X and Z). Figure 3 presents the experimental setup, simulation domain, and melt pool size calculations for both horizontal and vertical die chill casting studies.

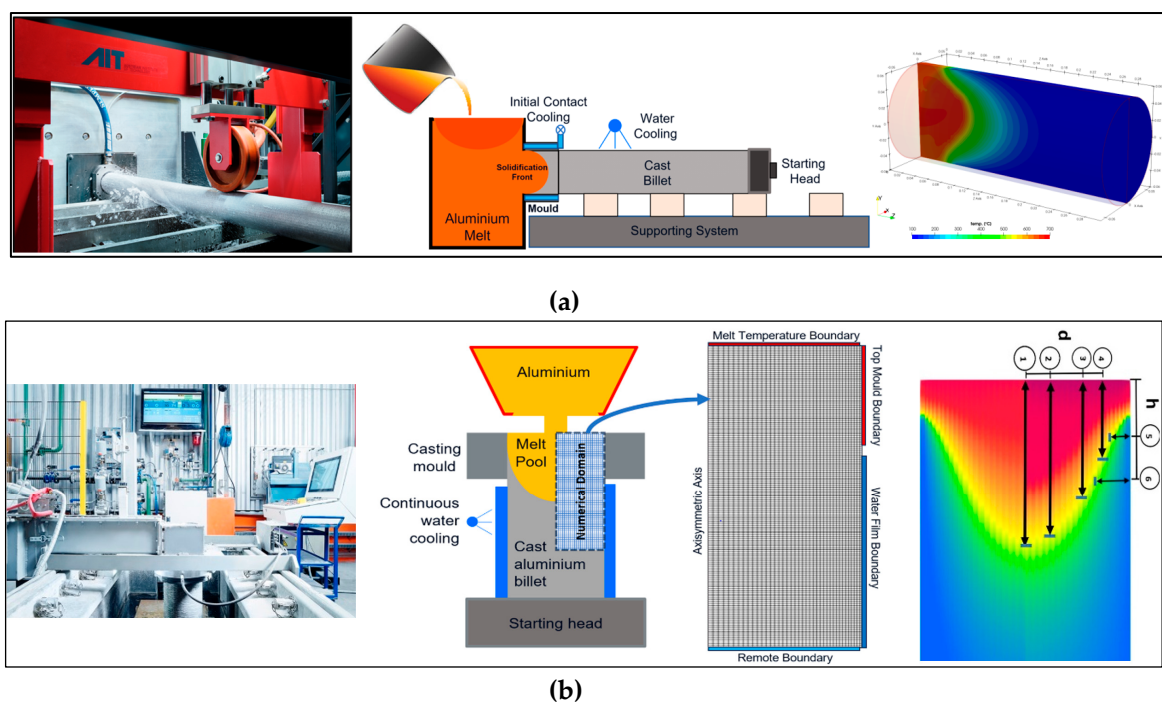


Figure 3. (a) Experimental and simulation setup for horizontal and; (b) vertical casting processes, including melt pool size estimation.

Extrusion Process: Using data science technologies, advanced data models can be developed to optimize and control metal extrusion processes. Extrusion is characterized by distinct features such

as large material deformation, plastic material flow, microstructural recrystallization, and grain restructuring. Therefore, any digital twin or digital shadow of the extrusion process must account for thermal evolution, large-scale deformation, and microstructure development to ensure accurate real-time predictions. In this research, a steady-state data modeling framework has been established using a multi-scale simulation and data integration approach. This framework incorporates geometric, material, boundary, and operational data. A specialized data structure was developed to manage information across different length scales, and a sequence of numerical solvers was employed to simulate phenomena at both macro and micro levels. The data modeling workflow, similar to the casting case study, followed these steps:

- **Definition of Prediction Objectives:** Macro- and micro-scale prediction targets were established across different length scales.
- **Parameter Variation Analysis:** Critical process parameters were identified, and their effects were analyzed through a range of process scenarios.
- **Balanced Sampling Strategy:** Techniques such as Latin Hypercube Sampling and Sobol sequences were used to evenly distribute parameter variations within the multi-dimensional design space, forming the final snapshot matrix.
- **Simulation and Data Generation:** A calibrated FE model was used to simulate the snapshot scenarios, and the resulting data were processed.
- **Multi-Scale Database Construction:** A split structured database was created to store macro- and micro-scale responses, organized using semantic rules.
- **Data Model Development:** Predictive models were generated for macro-scale thermal fields and micro-scale microstructural evolution.
- **Machine Learning Enhancement:** Additional training using machine learning techniques was conducted to improve model accuracy and robustness.
- **Model Validation:** Extensive validation studies were performed under normal, near-boundary, and extreme conditions using DOE scenarios.

To execute the scenarios defined in the snapshot matrix, numerical simulations of the extrusion process were performed using the thermo-mechanical solver HyperXtrude (HX). This solver utilizes an Arbitrary Lagrangian-Eulerian (ALE) hybrid technique to accurately capture large material deformations [28]. The primary process parameters—ram speed and initial billet temperature—were systematically varied, with aluminum 6060 selected as the billet material. The parameter ranges were set between 440 °C and 550 °C for the initial billet temperature, and 1 to 10 mm/s for the ram speed, based on a cylindrical billet with a 50 mm diameter. The computational domain was discretized using approximately 172,000 volumetric elements, modeled as a quarter section due to double symmetry conditions. Figure 4 presents the experimental setup, the snapshot matrix, and the FE simulation of the extrusion process.

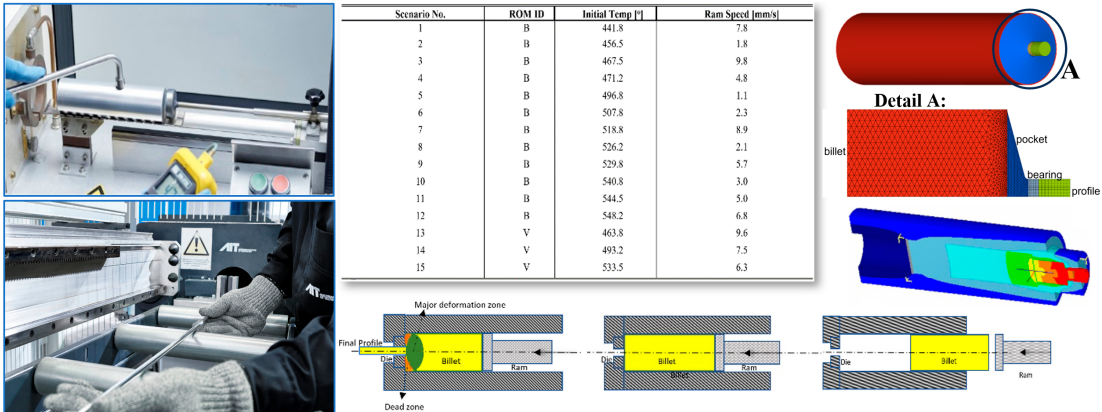


Figure 4. Experimental, simulation, and snapshot scenario matrix including basic and DOE scenarios for extrusion case study [17].

3.3. Analyses and Performance

The comparative analyses and performance studies for real-time data models can be performed to assess their performances for steady-state process modelling. However, there are challenges and difficulties related to multi-physical and multi-phase casting processes which highly depend on thermal field evolution during these processes. Likewise, for the extrusion process modelling the simulation of thermal field development during the start of the process and within the large deformation zone is necessary to have accurate predictions. Hence, data model needs to be adequately trained to properly reflect the thermal field including heat transfer, cooling, and temperature dependent material properties. Data from properly designed experimental work and verified numerical simulations can be employed to adjust the models for an accurate prediction of the thermal field.

Casting Processes: Casting processes can significantly benefit from the integration of data-driven models, particularly by analyzing variations in key process parameters such as initial melt temperature, water-cooling configurations, and casting speed. Leveraging such models can help mitigate the risks of hot tearing, cold cracking, and the formation of voids and other defects. Additionally, it is important to account for buoyancy effects induced by gravity in both numerical CFD simulations and data-driven models. This consideration necessitates extending the computational domain vertically—from the top to the bottom of the billet—to accurately capture thermal and flow dynamics. However, for vertical casting processes, geometric simplifications are possible. The domain can often be reduced to a quarter section or even a quasi-2D symmetric wedge, particularly when modeling round billet geometries, thereby reducing computational cost without compromising accuracy.

When calculating melt pool depth and temperature contours using data-driven models, a key challenge arises from the structure of the data produced by the CFD solver. The solver is optimized for minimal matrix sizes, which results in a computational node ordering that may not be suitable for data model training. Specifically, reading nodal temperatures in computational order—from the bottom of the billet to the top of the melt pool—can produce abrupt temperature gradients and discontinuities. These high-gradient data sequences are difficult for data models to interpret and generalize from effectively. To address this, additional post-processing was performed to transform the raw CFD output into a format more suitable for data modeling. A snake-shaped nodal reordering scheme was applied to smooth the data gradients and reduce sharp transitions. This reordering improves the continuity and consistency of the input data, enhancing the performance and reliability of the data models. Figure 5 illustrates the nodal reordering strategy for both vertical and horizontal casting processes, while Figure 6 presents a comparison between CFD-calculated and data model-estimated temperature values, along with the normalized error graph for the first 3,000 nodes in DOE scenario no. 3.

For comparison, the computational wall-clock time for each CFD simulation ranges from approximately 720 to 1,200 seconds using eight parallel computing cores. In contrast, the real-time data-driven solver requires only about 1.3 seconds on a single core to estimate the thermal responses. Upon examining the comparison graphs, it becomes evident that the highest normalized errors—ranging from approximately 6% to 8%—occur at locations with steep temperature gradients, particularly near the bottom of the melt pool. As the snake-like estimation path traverses from the solidified billet toward the melt pool base, the temperature variations become more pronounced. Consequently, even with the most optimized combination of data solver and interpolator, the real-time data model struggles to fully capture the rapid thermal transitions in this region.

Extrusion Process: Extrusion processes are highly adaptable manufacturing methods used to produce components with a wide range of shapes, sizes, and complex cross-sections through the use of pre-designed dies. Numerical simulations, along with fast real-time predictive models, play a crucial role in optimizing these processes by providing insights into the large thermo-mechanical deformations that occur during material flow. To develop effective data-driven models for extrusion, it is essential to construct a well-structured initial database that offers sufficient data points with

balanced density across the multi-dimensional parameter space. As outlined in previous sections, a snapshot matrix was generated using advanced sampling techniques to capture the variability of key process parameters within the operational limits of the extrusion machine.

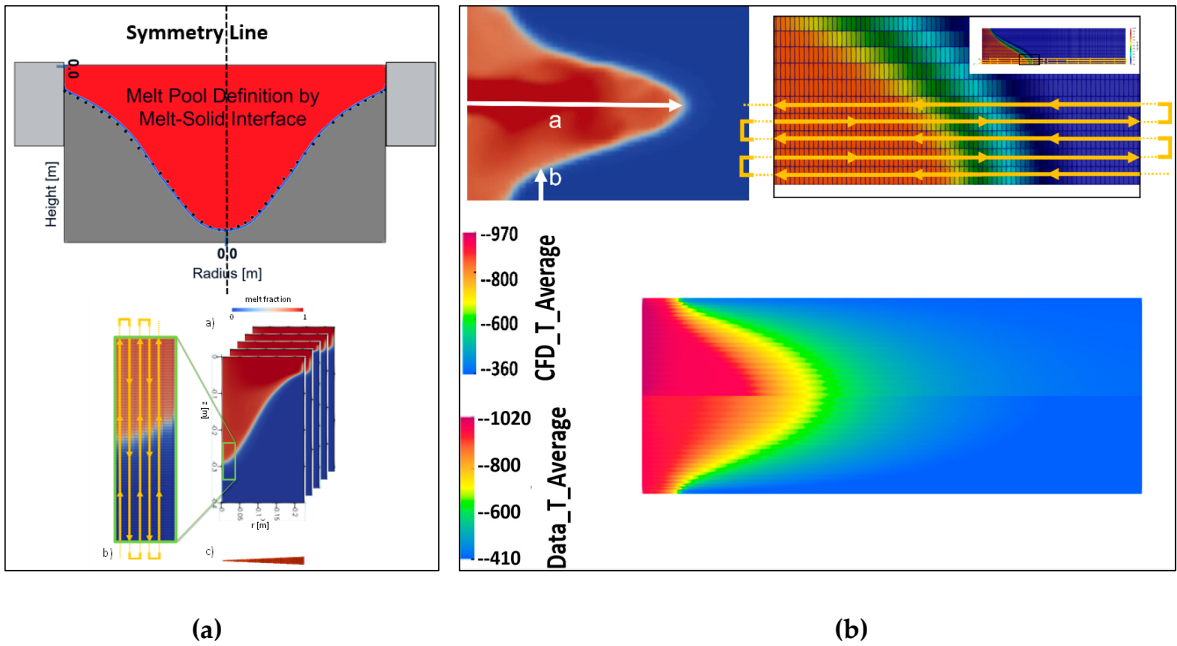


Figure 5. Model setup and nodal reordering strategy for both (a) vertical and; (b) horizontal casting processes [14].

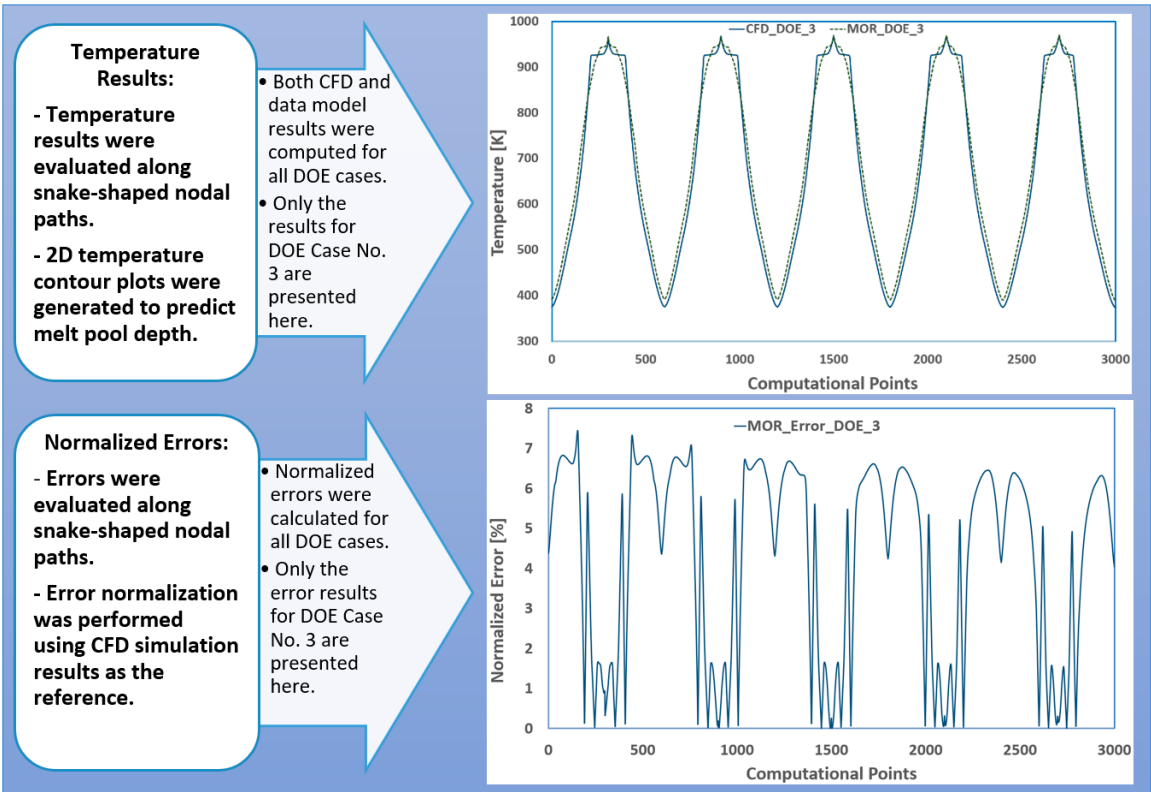


Figure 6. Comparison between CFD-calculated and data model-estimated temperature values, and normalized error graph for first 3,000 nodes in DOE scenario no. 3.

Similar to the casting case, validation and comparative analyses were conducted to evaluate the performance of data-driven models for extrusion processes. The objective was to assess the reliability and accuracy of real-time predictive models using data derived from the process scenarios defined in the snapshot matrix. For model development, widely adopted eigenvalue-based techniques—specifically Singular Value Decomposition (SVD)—were employed alongside regression analysis methods. Figure 7 presents a comparison of temperature graphs generated using both SVD and regression-based models, benchmarked against DOE results. To interpolate process data for validation cases, the inverse distance weighting (InvD) technique was applied, enabling localized estimation based on proximity within the parameter space.

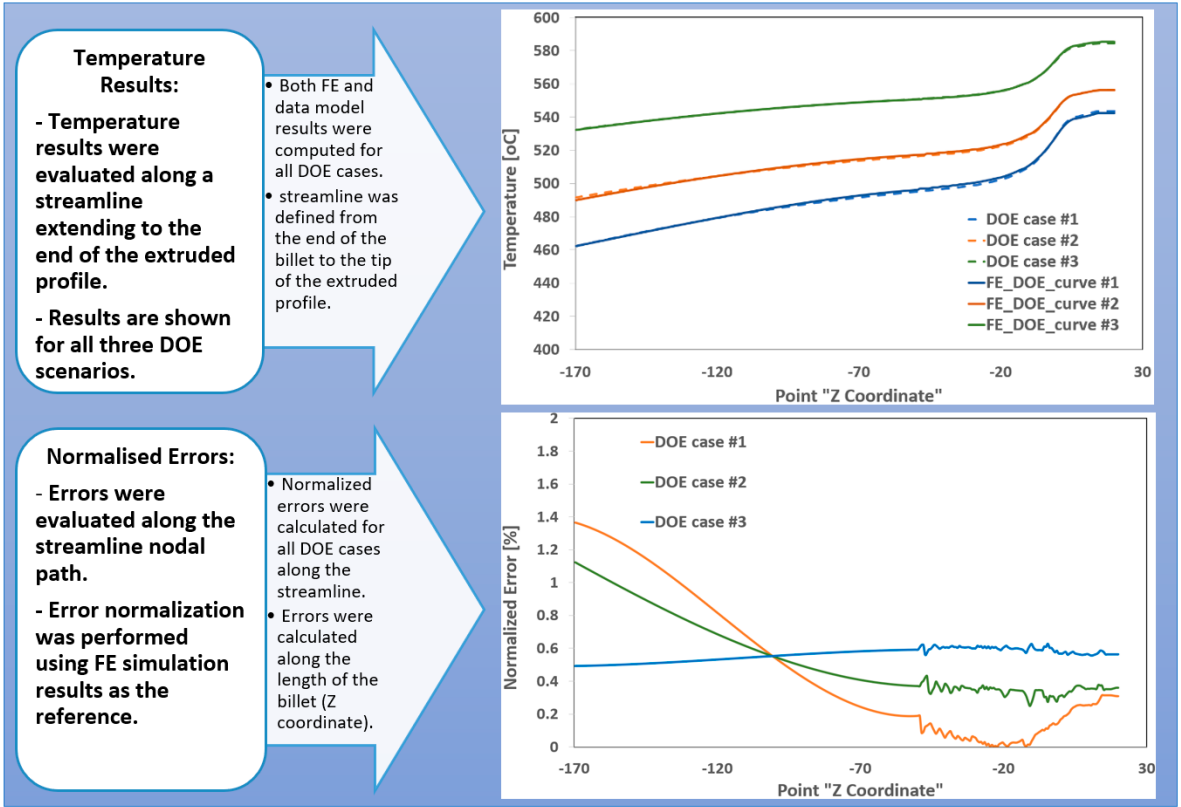


Figure 7. Comparative analysis of FE-calculated and data model-estimated temperature values, along with a normalized error graph along the streamline points for all DOE scenarios.

Steady-state FE thermal results for three validation scenarios were compared with predictions from the real-time data-driven models along the entire length of the billet and the final extruded profile. Normalized error percentages were calculated by comparing the data model predictions at each node along the billet with the corresponding FE simulation results. The analysis revealed that the highest normalized errors occurred at the billet boundaries—specifically at the initial billet entry and near the die exit—where significant material deformation takes place. These regions are characterized by complex thermo-mechanical interactions and rapid temperature fluctuations as the material undergoes large deformation while passing through the die. Such conditions present challenges for the data model, limiting its ability to accurately capture the thermal behavior in these zones

4. Transient Processes – Modelling Strategy

Most manufacturing processes exhibit a degree of transient and time-dependent behavior due to dynamic changes occurring throughout the process. These transient events may include startup and shutdown phases, variations in material properties and process parameters, changes in tooling and boundary conditions, abrupt thermal fluctuations between cycles (particularly in multi-cycle

operations), and environmental disturbances. To effectively model such processes, it is essential to track these temporal variations through time series analysis, enabling better process optimization and control. Therefore, any data modeling strategy must account for temporal dependencies, non-linear process dynamics, and the need for fast, real-time operation.

4.1. Basic Modelling Concepts

Historically, classical time series modeling techniques such as Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) have been employed to capture process dynamics [29]. Additionally, methods like Fourier transforms, Wavelet analysis, and state-space models have been used to decompose and analyze dynamic effects in manufacturing systems. However, with the advent of data science in engineering, the modeling of transient processes has undergone a significant transformation. Advanced AI-based techniques—such as Time Series Transformers, Temporal Convolutional Networks (TCNs), Recurrent Neural Networks (RNNs), and eigen-based data models—are now being adopted to capture complex temporal behaviors with greater accuracy and adaptability [19].

The eigen-based data solver technology employed in this research enables the decomposition of time series process data and projects it into a defined parameter space tailored for process optimization. This approach facilitates the identification of dominant patterns and correlations by aligning the data with the most informative dimensions of the process parameter space. The use of mathematical eigen-solvers in dynamic and time-dependent engineering systems has a long-standing tradition. Eigenmodes have historically been utilized to capture the intrinsic characteristics of such systems, particularly in the context of vibration analysis, structural dynamics, and control systems. To illustrate this concept, consider the general form of a dynamic governing equation of motion for an engineering system [15]:

$$[K][u] + [M][\ddot{u}] + [C][\dot{u}] = F[t] \quad (1)$$

where $[K]$, $[M]$, and $[C]$ can be considered as stiffness, mass, and damping matrices, respectively. The $[u]$ and $F[t]$ are the displacement and dynamic force vector. Using modal decomposition, the system's tempo-spatial characteristics can be expressed as:

$$u(x, t) = \sum_n \hat{u}_n(x, \omega_n) e^{i\omega_n t} \quad (2)$$

where ω_n and $\hat{u}_n$ are the system dynamic modes and Fourier coefficients representing the system's dynamic modes. This leads to a modal form of the governing equation in modal coordinate as:

$$[\hat{K}_f][u] = \hat{F}[t] \quad (3)$$

where $[\hat{K}_f]$ and $\hat{F}[t]$ are eigen-based dynamic stiffness matrix, and force spectral decomposed vector. This same concept can be extended to data matrices derived from dynamic system responses (e.g., from simulations or experiments). Instead of assembling physical system matrices, the data-driven approach uses eigen-decomposition to extract system characteristics, significantly reducing computational effort and enabling real-time predictions for new input scenarios. An alternative data-driven formulation of the system response can be expressed as [15]:

$$[C].[X(x, t)] = [Y(x, t)] \quad (4)$$

where $[C]$ is the system characteristics derived via eigen-based techniques, $[X(x, t)]$ represents the input process parameters, and $[Y(x, t)]$ is the predicted system response. For a dynamic process represented by a snapshot matrix S , constructed using balanced sampling techniques (e.g., Latin Hypercube Sampling, Sobol sequences), the transient system response can be approximated as:

$$[Y(x, t)] = \frac{1}{N} [S][S]^T \quad (5)$$

where $[S]$ is the estimated system characteristics (for sampled process scenarios). To further decompose the system's tempo-spatial behavior, data science techniques such as Proper Orthogonal Decomposition (POD) or SVD can be applied:

$$[C(x, t)] = U \Sigma V^T \quad (6)$$

where U represents spatial eigenvectors, Σ is the diagonal matrix of singular values (eigenvalues), and V^T captures the temporal dynamics. By projecting the scenarios in snapshot matrix onto the most influential eigenmodes (those with the highest energy in Σ), the system can be reconstructed

using a reduced-order model. This significantly reduces the size of the system matrices, enabling fast, real-time predictions for ongoing processes.

4.2. Case Study

Many manufacturing processes can be classified as transient due to their inherently dynamic nature. Notable examples include AM and High-Pressure Die Casting (HPDC), both of which involve rapid and complex changes during operation and control. This research investigates the application of data-driven models for the optimization and control of such transient processes, focusing specifically on these two representative manufacturing methods.

AM Process: AM processes involve the layer-by-layer fusion of materials using thermal energy sources to build components directly from digital models. These processes enable the fabrication of geometrically complex, lightweight, and customized parts that are often difficult or impossible to produce using conventional manufacturing methods. Additionally, AM supports distributed and remote production with minimal tooling requirements and significantly reduced lead times. AI techniques are increasingly being applied to optimize various aspects of AM, including topology optimization, toolpath generation, and build orientation. However, developing a digital twin or digital shadow for AM processes presents significant challenges due to their highly transient and dynamic nature. To achieve accurate real-time predictions, data models for AM must be capable of handling rapid variations in energy input power, deposition speed, thermal and displacement fields. These variations can lead to phenomena such as overheating, rapid cooling, thermal distortion, and residual stresses. Moreover, AM multi-scale modeling is essential to capture layer-by-layer microstructure evolution, defect and void formation, and mechanical property development. Such models must integrate data across spatial and temporal scales to provide reliable predictions and enable effective process control.

In this study, numerical simulations of AM processes were conducted using the thermo-mechanical solver LS-DYNA, which applies an iterative approach to solving thermal, stress, and material deformation fields [30]. Key process parameters—such as initial base temperature, power input, and deposition speed—were systematically varied to construct the snapshot matrix for model development. The material used was aluminum 6061, with a single-layer deposition performed on a thermally pre-conditioned substrate wall measuring 100 mm × 50 mm × 6 mm. Figure 8 presents both the experimental setup and the FE mesh used for the AM process, along with the corresponding snapshot matrix utilized in the model-building phase. For data-driven model development, the eigenvalue-based solver described in previous sections was employed to enable real-time predictions tailored to AM applications.

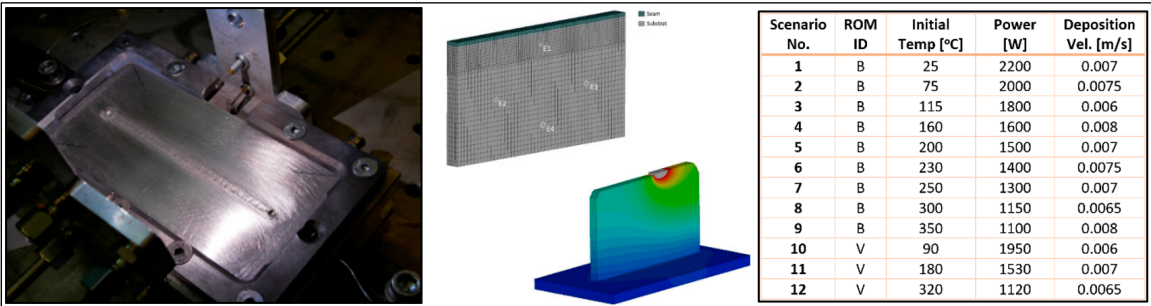


Figure 8. Experimental, simulation, and snapshot scenario matrix including basic and DOE scenarios for AM case study [16].

HPDC Process: The HPDC process is a highly efficient and precise metal casting method widely used for the cyclic mass production of geometrically complex components. In HPDC, molten metal is injected into a mold cavity under high pressure, enabling the rapid production of parts with excellent dimensional accuracy, surface finish, and mechanical properties. This process is extensively

applied in industries such as automotive, aerospace, and electronics. HPDC operates through repetitive cycles of metal injection, solidification, and cooling. In this research, the cyclic behavior of the HPDC process was simulated using real-time data models, where data solvers and interpolation techniques were employed to predict the effects of varying process parameters. The primary objective of these real-time models was to predict thermal evolution in different regions of the mold over multiple cycles using time-series solvers. A snapshot matrix was constructed to represent a series of scenarios with varying process parameters—such as initial melt temperature and maximum piston speed—based on a simplified initial setup. Numerical simulations for these scenarios were conducted using the NovaFlow&Solid simulation software [31].

The numerical domain used in this study encompasses the entire mold assembly, including cooling channels, gating system, and the injection chamber. Simulations were conducted over six consecutive casting cycles, incorporating both air and water cooling phases between each injection to replicate realistic thermal conditions. The casting material selected was AlSi9Cu3(Fe), while the mold was modeled using AISI H13 tool steel. One of the key input parameters was the piston speed profile, defined as a time-dependent curve to allow variation in maximum piston speed across different simulation scenarios. The second variable parameter was the initial melt temperature at the gating system, which was varied across the snapshot scenarios and DOEs. Figure 9 illustrates the experimental setup and the numerical CFD domain used for simulating the multi-cycle transient HPDC process, including all six filling cycles.

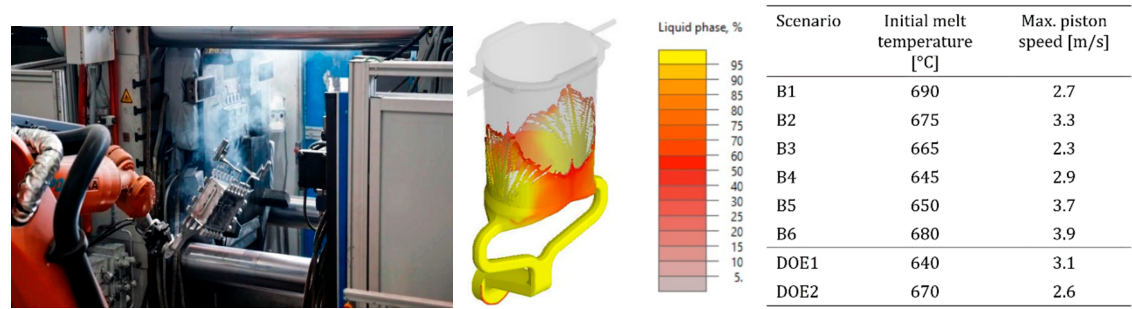


Figure 9. Experimental setup and numerical CFD domain for multi-cycle transient HPDC process, encompassing all six fundamental filling cycles and two DOE snapshot scenario matrices [14].

4.3. Analyses and Performance

To assess the performance of real-time models in transient and time-series process applications, comparative studies were conducted using validated FE and CFD simulation results. These studies encompassed both industrial AM and HPDC case studies. The objective was to evaluate the accuracy and reliability of real-time models by comparing their outputs against verified DOE-based FE and CFD results derived from process snapshot scenarios.

AM Process: In the AM case study focused on a time-dependent generative deposition process, a combination of solver-interpolator techniques was employed for model development. Specifically, the eigenvalue-based SVD solver was used in conjunction with InvD and Radial Basis Function (RBF) interpolators. These methods were further enhanced through machine learning-based training to accommodate the high dimensionality and complexity of the AM process data space. While there are no universally established guidelines for selecting the optimal solver-interpolator combination for transient data modeling, certain configurations demonstrate superior performance in terms of accuracy and reliability—particularly for highly dynamic processes such as AM. Figure 10 presents a comparison of temperature time histories between DOE simulation results and time-series data models, using the SVD_InvD technique at a nodal point on the wall during the AM process. Additionally, Figure 10 illustrates the corresponding normalized transient error graphs for the same DOE scenarios.

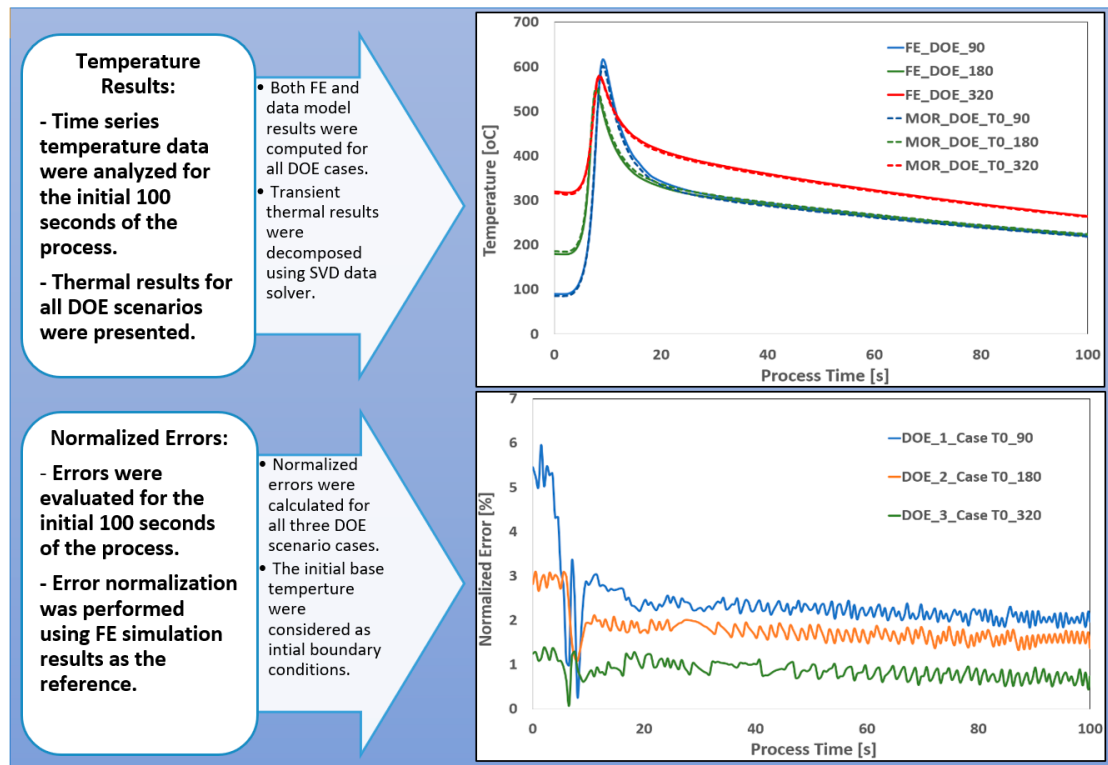


Figure 10. Comparative analysis of FE-calculated and data model-estimated temperature values, along with a normalized error history along the streamline for all DOE scenarios.

The results presented here demonstrate that the transient data models are capable of reasonably predicting the dynamic responses of AM processes. Despite the limited size of the snapshot matrix and process databases for the AM case study, the available data exhibit a balanced distribution across the multi-dimensional parameter space—such as the operational limits of manufacturing equipment. However, challenges remain in accurately capturing rapid thermal transients, particularly during high heating and cooling events (e.g., when the torch passes over a measurement point), and in aligning with initial process conditions (e.g., base temperature). These rapid fluctuations in data patterns—such as temperature and strain rate—can lead to significant prediction errors due to steep data gradients. As a result, further data enrichment and model training may be required to enhance the robustness and accuracy of these models for highly dynamic manufacturing processes.

HPDC Process: In the cyclic HPDC process for aluminum alloys, identifying optimal process parameters throughout the casting cycle remains a complex challenge. Traditionally, this task has relied heavily on extensive experimental investigations and numerous trial-and-error iterations. Although numerical CFD models are available to simulate the intricate interactions among various physical phenomena, their high computational cost significantly limits their practical application. As a result, only a limited number of process scenarios can be explored, constraining the size of the snapshot matrix and the resulting database used for transient data model development. To address this limitation, careful sampling techniques must be employed to select a small number of representative scenarios that capture the essential characteristics of the cyclic casting process. In this study, six primary scenarios and two validation DOE scenarios were selected. Due to computational constraints, CFD simulations were restricted to five casting cycles per scenario. Figure 11 presents the thermal results from a typical CFD cyclic simulation, highlighting the descending maximum and minimum temperature curves across six filling cycles. Figure 12 illustrates the estimated time-history temperature curves at two sensor locations, comparing the predictions from SVD-InvD and SVD-Kriging data models against CFD simulation results.

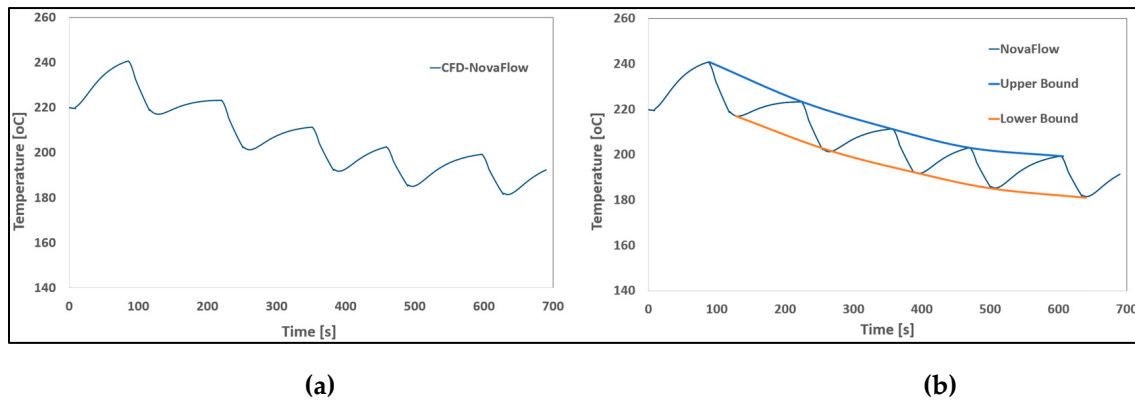


Figure 11. (a) Thermal cycle results from a typical CFD numerical simulation, and; (b) with descending maximum and minimum temperature curves across six filling cycles.

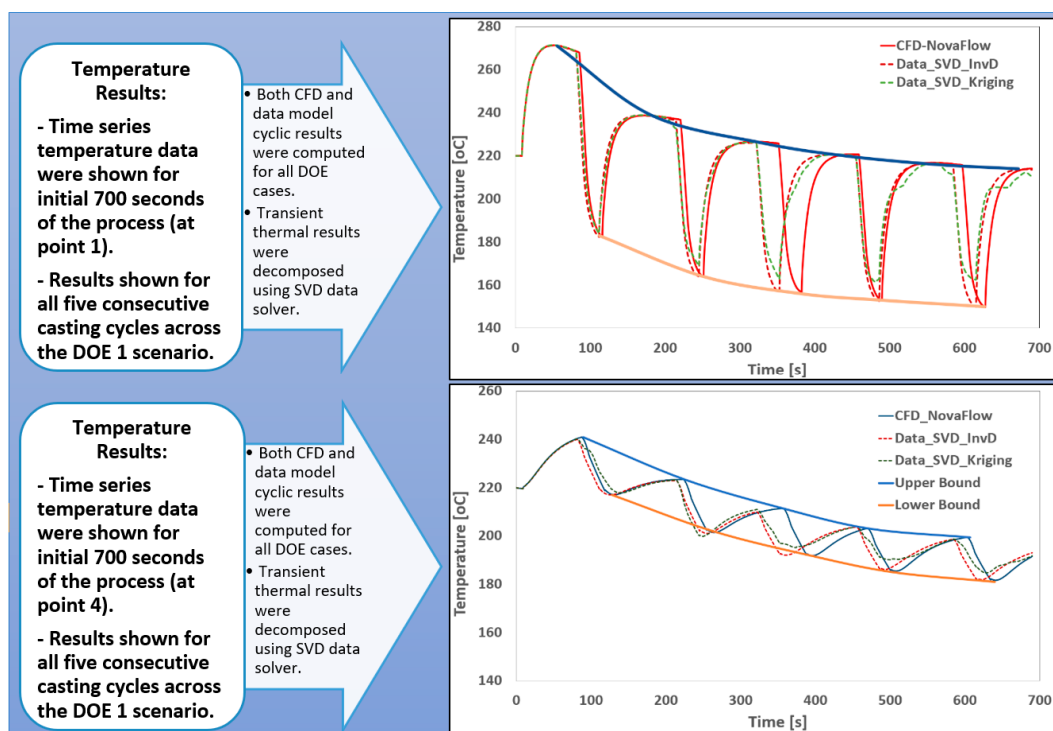


Figure 12. (a) Time-history temperature curves at two sensor locations, comparing the predictions from SVD-InvD and SVD-Kriging data models against CFD simulation results.

The total computational time required for the CFD simulations of six consecutive HPDC filling cycles was approximately 54,000 seconds. In contrast, the data-driven models were able to estimate the corresponding time-history responses in just 1.8 seconds. However, due to the limited size of the snapshot matrix—comprising only six basic scenarios with two varying process parameters—and the resulting small database, the data models encountered difficulties in accurately predicting the individual solidification times within each cycle. This limitation led to a lag in capturing the transition between cycles, ultimately causing a time shift in the overall predictions.

5. Cyclic and Generative Processes

Several widely used manufacturing processes operate in cyclic or generative modes, where parts are produced either in sequential cycles—as in HPDC—or through multi-component, multi-layer deposition, as seen in certain casing and AM processes. The performance and quality of these thermal history-dependent processes are highly influenced by the ability to control and adapt process parameters in response to evolving thermal and mechanical fields during production. In HPDC, the

inter-cycle heating and cooling phases contribute to the development of a thermal history, which is shaped by the characteristics of each cycle and the amount of thermal energy introduced. This thermal history affects die temperature, solidification behavior, and ultimately, part quality and tool life. In AM processes, it is often necessary to adjust the input power for each successive layer to maintain a consistent melt pool size. This is particularly important for achieving uniform melt pool size and inter-layer bonding, minimizing defects, and ensuring mechanical integrity across the build. To ensure optimal outcomes, digital twin or shadow frameworks with real-time monitoring and adaptive control strategies can be integrated, enabling dynamic adjustment of process parameters based on feedback from thermal and mechanical field data.

5.1. Generative Data Models – Basic Concepts

Conventional FE and CFD simulation techniques face limitations when applied to generative manufacturing processes, particularly due to the need for the numerical domain to evolve over time in order to accurately capture process history and transient events. This evolving domain must be synchronized with the material deposition rate and thermal energy input to reflect real-world process dynamics. While some adaptive meshing technologies exist—capable of modifying existing elements or adding new layers—these methods are often insufficient for modeling the generation of highly complex geometries. As a result, dynamic meshing and evolving domain techniques have been proposed to better align the numerical domain with the actual stages of production [32].

In parallel, data modeling strategies for these highly dynamic and generative processes present significant challenges. Creating a sufficiently rich initial dataset is often time-consuming and costly, especially since the optimal process settings are typically unknown at the outset. This necessitates numerous trial-and-error loops through experiments or simulations to fine-tune parameters across cycles or layers—an approach that is often impractical. To address this, a generative data modeling framework can be employed, enabling the development of predictive models from limited initial data. The proposed approach involves the following steps:

1. **Data Preprocessing:** Available experimental data or validated literature results are preprocessed and filtered to establish a data correlation framework.
2. **Simulation-Based Calibration:** Numerical simulations are conducted to replicate the experimental or literature scenarios for model calibration and validation.
3. **Correlation Analysis:** A correlation study is performed using verified simulation data to isolate and relate the effects of key process parameters through data sampling techniques.
4. **Initial Response Estimation:** Based on the correlation framework and additional data training, initial system responses are estimated and incorporated into the database.
5. **First-Generation Model Development:** A preliminary data model is created to predict process parameters across different cycles or layers.
6. **Validation and Feedback Loop:** Verified simulations are used to assess prediction accuracy. A feedback loop is established to refine the correlation framework.
7. **Database and Model Update:** The process response database is updated based on the refined correlations, and a new generation of the data model is developed.
8. **Iterative Refinement:** This cycle of error checking, correlation refinement, and database updating continues until the desired accuracy and reliability are achieved.

Figure 13 illustrates the schematic overview of the generative data modeling framework.

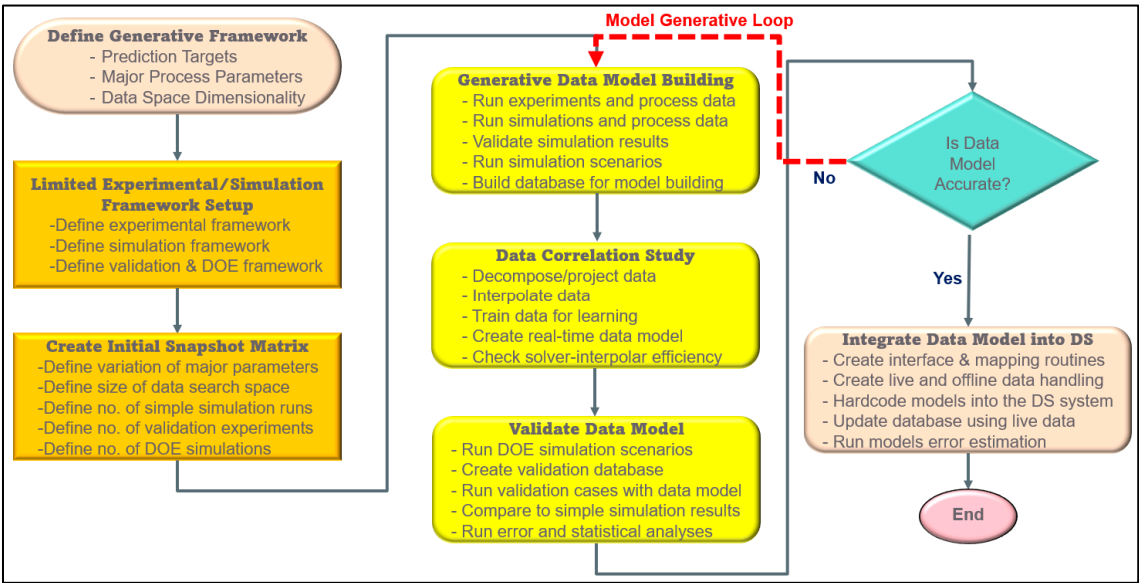


Figure 13. Schematic overview of generative data modeling framework, including key definitions and the iterative model generation loop.

5.2. Case Study – AM Power Prediction

Power prediction in multi-layered AM processes presents a significant challenge due to the strong influence of thermal history, which directly affects melt pool size and, consequently, the final part quality. In Wire-Arc Directed Energy Deposition (WA-DED), a generative approach is used to build components layer by layer, with energy supplied via a welding arc. To ensure high-quality outcomes, it is essential to accurately predict welding power, deposition speed, and thermal evolution throughout the process. These predictions depend on factors such as material properties, part geometry, deposition strategy, and machine limitations.

In this research, generative data modeling techniques were employed to develop real-time predictive models using an iterative combination of data solvers and interpolation methods. Specifically, eigen-based data solvers were integrated with the Kriging interpolation technique to disentangle complex data trends within a multi-dimensional parameter space. The modeling process began with a single experimental trial, guided by initial predictions from a calibrated numerical simulation. During the trial, power settings for the first 15 layers were dynamically adjusted to maintain a uniform melt pool size across all deposited layers. The experimental results validated the simulation predictions, confirming the model’s reliability for further data correlation studies. Subsequently, simplified scenarios based on data sampling techniques was simulated using constant power settings across all layers, with variations in initial melt pool sizes. These simulations generated a small response database, which was then analyzed to establish correlations between power input and melt pool size at different deposition layers.

Using this correlation study, initial estimations of thermal history effects were derived through a combination of data solvers and machine learning routines. Based on the resulting insights, the first-generation process data model was developed, enabling preliminary power predictions for achieving uniform melt pool sizes across all layers.

The predicted power sets were subsequently evaluated using FE numerical simulations to estimate the normalized prediction error. The results from these comparative studies were post-processed and fed back into the correlation framework, forming a closed-loop feedback system aimed at improving prediction accuracy. Based on this feedback, a second-generation generative data model was developed using the trained input data and an enriched response database. This updated model was then used to generate a new set of power predictions. These revised power settings were again validated through numerical simulations, and a new set of error margins was calculated. The iterative

feedback and model refinement loop continued until the model achieved the target error margin of approximately 3%.

Figure 14 presents the experimental setup, simulation environment, and initial power calculation results based on the initial snapshot matrix, using constant welding power across all deposition layers.

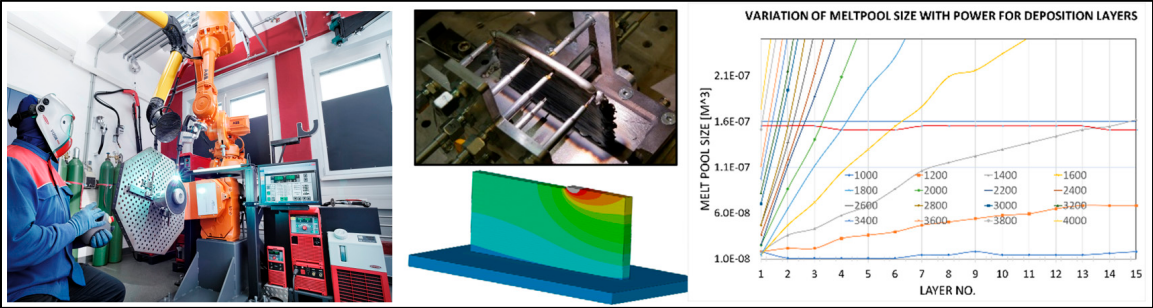


Figure 14. Experimental setup, simulation, and initial power calculation results using initial snapshot matrix and constant welding power for all deposition layers [33].

5.3. Analyses and Performance

The performance of generative data models at various stages of development can be evaluated using specific error indices and performance indicators. In the early stages, when the initial database is relatively small, the first generation of the data model primarily relies on correlation studies between process parameters. The initial snapshot matrix typically consists of simple, deterministic process scenarios designed to capture the fundamental relationships among key parameters, thereby forming the foundation for early data correlation analysis. These scenarios are often generated using calibrated numerical simulations, which produce tabular datasets reflecting the variation of individual process parameters. As the database expands through successive iterations of the data model, standard error evaluation methods can be applied to assess model accuracy and reliability. Figure 15 illustrates the initial variation of process parameters in relation to the resulting melt pool size, along with the corresponding data correlation matrix.

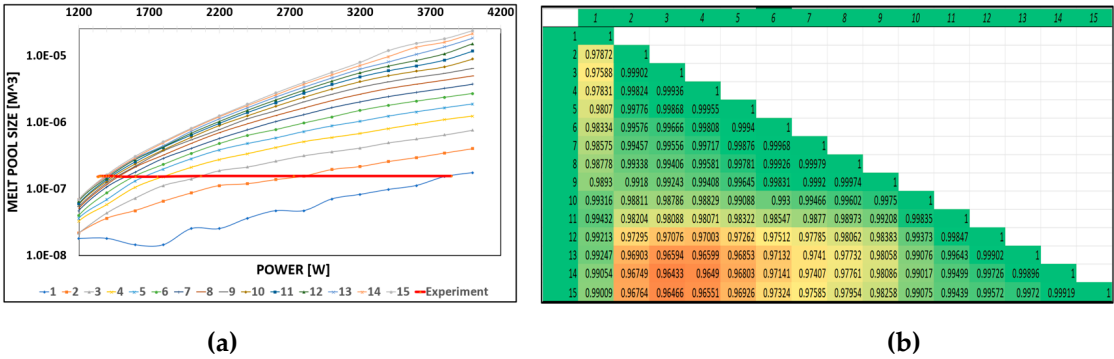


Figure 15. (a) Initial variation of welding power in relation to the resulting melt pool size, and; (b) the corresponding data correlation matrix.

As generative models progress through successive iterations, error indices remain essential benchmarks for evaluating improvements in accuracy and reliability. By analyzing error trends across different generations of model development, researchers can assess the impact of newly integrated data, refined solvers and interpolation methods, or enhanced machine learning configurations. Furthermore, normalized error analysis—such as evaluating deviations along specific lines, at critical points, or during key time steps—can inform targeted data enrichment and model tuning strategies. This iterative feedback loop enables the model to become increasingly robust and

predictive, particularly for complex and transient manufacturing processes like AM and HPDC. Figure 16 illustrates the real-time predicted power across different deposition layers aimed at maintaining a consistent and stable melt pool size using the third-generation data model. It also presents the normalized error percentages for three consecutive stages of data model development.

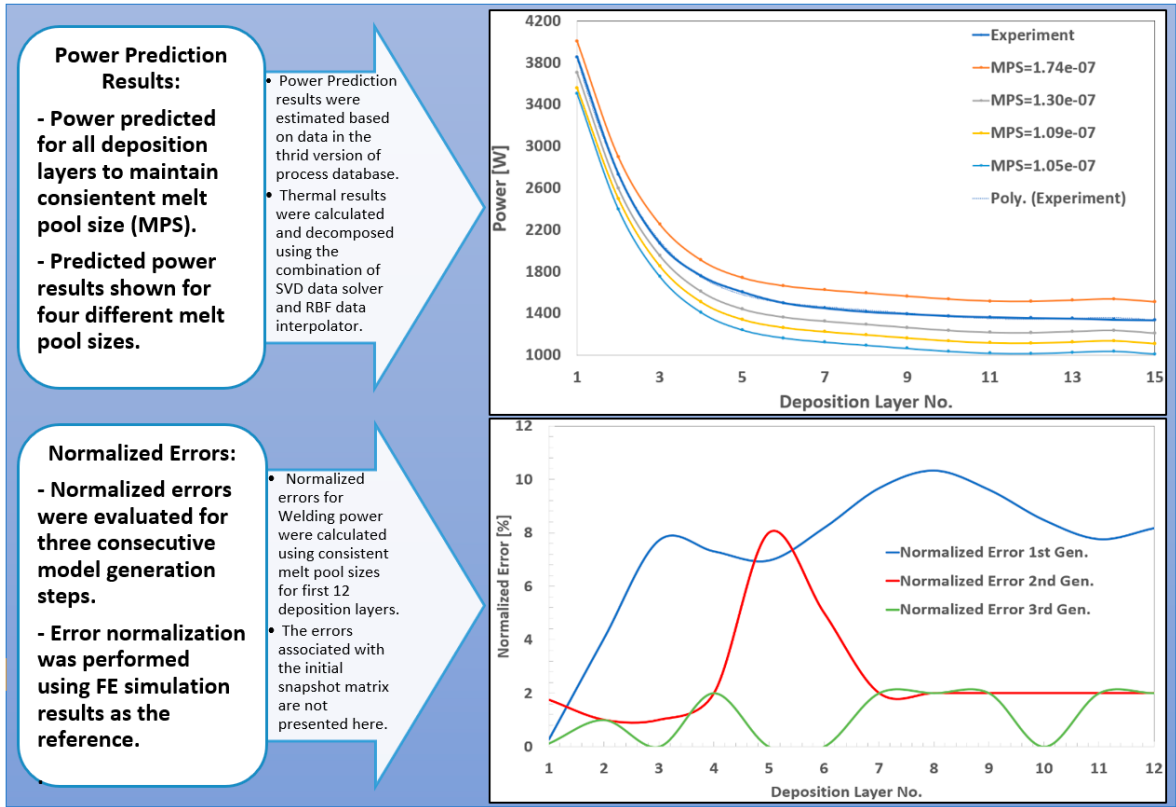


Figure 16. Predicted welding power across different deposition layers aimed at maintaining consistent melt pool size using third-generation data model and the normalized error percentages for three consecutive stages of data model development.

6. Results and Discussions

The performance, accuracy, reliability, and validation of real-time data models for manufacturing processes—including steady-state, transient, and generative types—must be rigorously evaluated to ensure their applicability to real-world industrial challenges. For such models to gain acceptance in practical settings, they must demonstrate robustness under varying process conditions and align with industry-specific requirements. In this research, a combination of data solvers, interpolation techniques, and ML routines was employed to develop efficient and accurate real-time data models. ML-based training and learning were conducted in one or two iterative rounds, both prior to and following the initial model generation, to optimize performance and enhance predictive capability.

6.1. Performance and Reliability

The results presented in the previous sections offer insight into the performance of data-driven models, demonstrating their potential to achieve reasonable accuracy in predicting steady-state, transient, and generative responses within manufacturing processes. Despite the high cost and time investment required to generate data for constructing comprehensive process databases, these limitations can be mitigated through the use of efficient sampling techniques and well-balanced data distributions across multi-dimensional parameter spaces. A significant challenge lies in the variability of data quality and resolution, which often stems from differences in process parameters,

material properties, and boundary conditions. These inconsistencies highlight the importance of a robust data strategy to ensure model reliability and generalizability. Another critical performance constraint is the expansive temporal and spatial scales inherent in multi-scale material processing. In particular, modeling microstructural evolution—essential for predicting final product quality—requires data models capable of capturing complex interactions across these scales with high fidelity.

In this research, a comprehensive performance evaluation was conducted to assess the reliability and effectiveness of data-driven models in simulating steady-state, transient, and generative behaviors in manufacturing processes. Figure 17 presents the comparative performance of various solver-interpolator combinations, including SVD_Kriging, SVD_RBF, SVD_InvD, and regression-based approaches, specifically applied to transient AM processes. The figure illustrates the maximum normalized errors and deviations from initial boundary conditions for each of the four solver-interpolator configurations. A preliminary analysis of these results reveals that certain combinations demonstrate superior performance in terms of predictive accuracy and data fitting capabilities, highlighting the importance of selecting appropriate modeling strategies for transient process simulations.

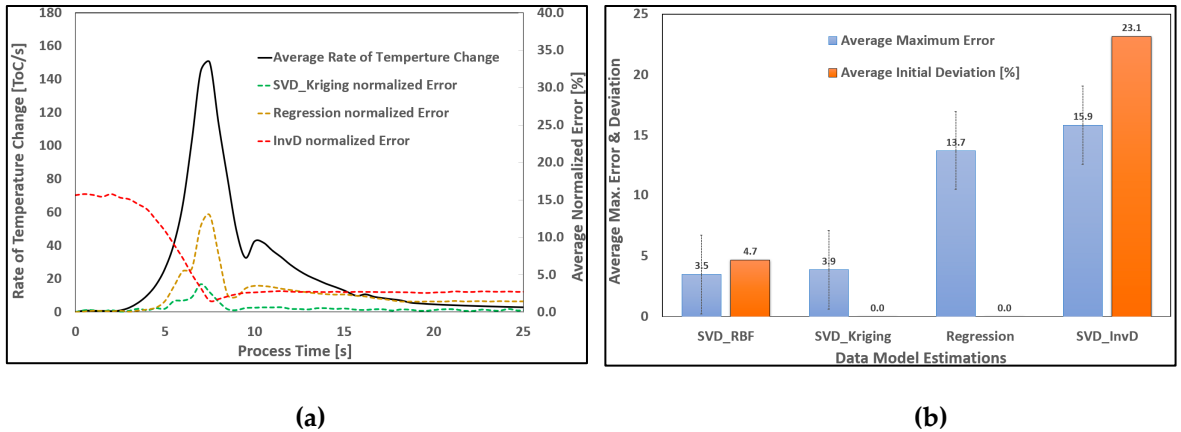


Figure 17. (a) Correlation study for rate-dependency and average maximum errors (over all DOEs) for different solver-interpolator combinations, and; **(b)** initial boundary data fitting and average maximum errors data real-time models.

As illustrated in the figures, the real-time data models exhibit a degree of temperature rate dependency—such as rapid heating or cooling during processing—as well as deviations from initial process conditions, including baseline temperature variations. Notably, certain solver-interpolator combinations, particularly SVD_Kriging, demonstrate superior performance in addressing both challenges, indicating their overall effectiveness compared to alternative techniques. In general, eigenvalue-based data solvers are well-suited for decomposing time-series data due to their ability to capture dominant temporal patterns. When these solvers are paired with advanced interpolation methods like Kriging, they can yield highly accurate predictive results. Conversely, for steady-state process modeling, regression-based approaches combined with spatial interpolators such as InvD and RBF tend to offer more effective spatial data decomposition and fitting.

6.2. Challenges and Shortcomings

As it has been shown in the previous section there are challenges with the use of data models for complex multi-physical, and multi-phase manufacturing processes. These challenges can be categorized in the following:

- 1- Data availability: As highlighted in the section 4.3, data acquisition and the construction of comprehensive process databases remain significant challenges in the development of reliable data-driven models. Sporadic, inconsistent, or incomplete process data often fail to adequately cover the full parameter space, leading to prediction inaccuracies, data gaps, and potential

biases. These limitations directly undermine the robustness and generalizability of the resulting models. Furthermore, acquiring high-fidelity data—particularly for multi-scale modeling, generative process prediction, and transient or high-speed operations—is both technically demanding and costly. These constraints significantly limit the volume, resolution, and diversity of data available for model training and validation.

- 2- Data fitting: Generalizing process data remains a persistent challenge in the development of robust data models. Many models struggle to accurately fit initial process boundary conditions, capture rapid changes in process variables (such as high heating or cooling rates), and handle abrupt nonlinear transitions in high-dimensional data spaces. These difficulties are particularly pronounced when training data is sparse, imbalanced, or lacks sufficient diversity. Overfitting to specific boundary conditions or high-gradient trends can significantly impair a model's ability to generalize to new or unseen scenarios, thereby limiting its practical application in dynamic environments such as digital twin or digital shadow frameworks. Compounding this issue is the data gradient—the variation in data quality, resolution, and contextual relevance across different process parameters and stages. For example, early-stage offline data (e.g., from validated simulations or literature) may be rich and well-structured, while real-time or online data streams are often noisy, incomplete, or difficult to interpret. This inconsistency complicates the seamless integration of data models into process advisory systems or digital twins, where both historical and live data must be harmonized to support real-time decision-making.
- 3- A significant challenge in developing data models for manufacturing processes lies in managing the vast temporal and spatial scales involved, particularly in the multi-scale modeling of material evolution. For instance, microstructural transformations in materials—such as those occurring during casting, extrusion, or AM—can unfold over milliseconds and micrometers, while the overall process may span hours and meters. Accurately modeling phenomena across these scales requires advanced multi-scale modeling frameworks, partitioned or hierarchical databases, data mapping and translation mechanisms, and sophisticated filtering techniques. These approaches are often computationally intensive and difficult to validate experimentally, posing practical limitations for real-time or large-scale deployment. Moreover, as manufacturing technologies evolve and new production methods emerge, static models and fixed-size databases quickly become obsolete. To remain relevant and effective, data models must be designed with adaptability in mind. This includes the ability to integrate dynamic, structured data streams, support regular self-updating mechanisms, and expand process databases over time. Such adaptability is essential for capturing evolving process conditions and maintaining model accuracy in rapidly changing industrial environments.

Figure 18 presents a data correlation analysis across three distinct process scenarios: a normal condition scenario, where process parameters are centered within the data search space; a boundary scenario, with parameters approaching the operational limits of the manufacturing system; and an extreme scenario, involving data points that extend slightly beyond the defined data search space. Despite the use of advanced sampling strategies and balanced scenario distributions across the parameter space, the results reveal notable differences in normalized error levels, particularly those associated with rate-dependent behaviors such as rapid heating or cooling. These discrepancies underscore the sensitivity of model performance to the positioning of process parameters within the data space. Furthermore, the Pearson correlation indices for two widely used interpolation techniques—RBF and Kriging—demonstrate that the choice of interpolation method significantly affects the model's ability to handle rate-dependent variations. While both techniques perform effectively under certain conditions, their predictive accuracy diverges when applied to scenarios characterized by high data gradients or near-boundary conditions, highlighting the importance of selecting appropriate interpolation strategies for robust data modeling in manufacturing processes.

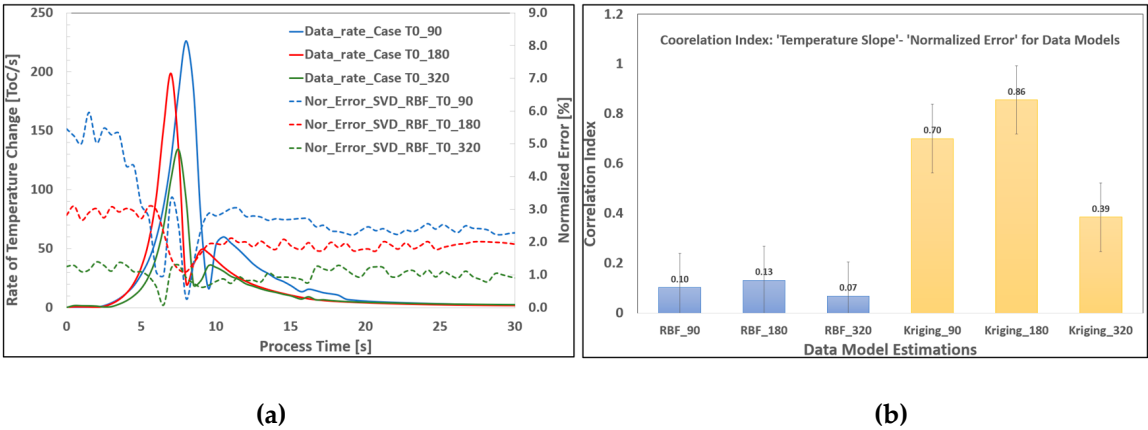


Figure 18. (a) Correlation study for rate-dependency and average maximum errors (over all DOEs) for different solver-interpolator combinations, and; **(b)** initial boundary data fitting and average maximum errors data real-time models.

6.3. Applications and Recommendations

To ensure the accuracy, reliability, and overall effectiveness of data models in manufacturing processes, several critical factors must be addressed. First, data availability and the construction of appropriately sized and structured databases are fundamental. High-quality, well-balanced datasets that comprehensively span the process parameter space are essential for building robust models. This requires the development of effective data filtering, mapping, and translation techniques for both offline and online data sources, ensuring coverage across a wide range of process scenarios—including normal, boundary, and extreme operating conditions. Second, rate dependency—such as rapid thermal, solidification, or mechanical changes—must be explicitly considered during both database development and model design. These dynamic behaviors significantly influence prediction accuracy, particularly in transient and generative manufacturing processes. Third, boundary and initial condition fitting is crucial. Models must be capable of accurately handling edge cases near operational limits without overfitting or compromising generalizability.

For transient, time-series, and generative process modeling, the use of time-resolved data and dynamic database architectures is essential to capture the evolving nature of manufacturing systems. In terms of modeling methodology, the selection of data solvers and interpolators plays a pivotal role. Eigenvalue-based solvers such as SVD are particularly effective for decomposing complex time-series data, especially when combined with advanced interpolation techniques like Kriging and appropriate machine learning methods to achieve high-fidelity predictions. For steady-state or spatial modeling, regression-based approaches paired with interpolators such as RBF or InvD offer efficient and interpretable solutions. Ultimately, data models should be designed with adaptability and scalability in mind. Incorporating mechanisms for continuous self-updating and the ability to expand databases over time is essential to accommodate evolving manufacturing technologies, changing process conditions, and the increasing availability of real-time data streams.

7. Concluding Remarks

In conclusion, the development of comprehensive databases and data-driven models for manufacturing processes holds substantial promise for advancing process sustainability, understanding, control, and optimization. However, the successful integration of these models into process advisory systems and digital twin frameworks requires careful consideration of several interdependent factors. This study emphasizes that the selection of appropriate data solvers, interpolators, and machine learning techniques must be closely aligned with the specific characteristics of the manufacturing process—whether it is steady-state, transient, or generative in nature. Eigenvalue-based solvers such as SVD, when combined with advanced interpolators like

Kriging or RBF, have demonstrated strong capabilities in capturing complex temporal and spatial dynamics, particularly in transient and high-gradient scenarios. Conversely, regression-based approaches continue to offer effective and interpretable solutions for steady-state modeling and spatial data decomposition. These insights underscore the importance of methodological alignment and data strategy in realizing the full potential of data-driven modeling in modern manufacturing environments.

The availability, distribution, and quality of process data are foundational to the success of data-driven modeling in manufacturing. High-resolution, well-balanced datasets that comprehensively span the full spectrum of process conditions—including normal, boundary, and extreme scenarios—are essential for building robust and generalizable models. Addressing data fitting challenges, particularly in the presence of sparse, noisy, or high-gradient data, requires the implementation of advanced data preprocessing techniques, including filtering, translation, and mapping strategies. Moreover, the inherent multi-physical, multi-phase, and multi-scale complexity of many manufacturing processes adds further layers of difficulty. Accurately capturing the interactions across different physical domains and scales—from microstructural evolution at the microscale to macroscopic process dynamics—demands the development of scalable, adaptive modeling frameworks. These frameworks must be capable of integrating heterogeneous data sources and supporting real-time updates, ensuring that models remain accurate and relevant in dynamic production environments.

Ultimately, the future of data modeling in manufacturing—and its integration into online advisory systems, digital twin platforms, and digital shadow frameworks—depends on the development of adaptive, self-updating, and interpretable models that can evolve in parallel with advancing manufacturing technologies. These models must not only deliver high predictive accuracy but also support seamless integration into broader digitalization ecosystems, including real-time control architectures and intelligent process advisory systems. By addressing the challenges outlined in this work and implementing the recommended strategies, the procedures for developing data models will become more robust, leading to the creation of more powerful, accurate, and resilient solutions. These advancements will play a critical role in enabling efficient, intelligent, and sustainable manufacturing processes—and will form the foundation of our next manuscript contribution.

Author Contributions: **AmirHorr:** conceptualization, methodology, writing-original draft preparation, software (datamodels), validation, data curation, writing-review, editing, visualization. **Matthias Hartmann:** conceptualization, methodology, supervision, project administration, funding acquisition, proof reading. **Fabio Haunreiter:** Software (FE simulations), investigation, data curation, validation, proof reading.

Funding: This research is financially supported by the Austrian Institute of Technology (AIT) under the UF2024 funding program, by the Austrian Research Promotion Agency (FFG) through the opt1mus project (FFG No. 899054), and by the European Commission under the Horizon Europe programme for the metaFacturing project (HORIZON-CL4-2022-RESILIENCE-01, Project ID: 101091635).

Data Availability Statement: The sharing of raw-data required to reproduce the case studies are considered upon request by readers.

Acknowledgments: The authors gratefully acknowledge the technical and financial support provided by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure, the Federal State of Upper Austria, and the Austrian Institute of Technology (AIT). Special thanks are extended to Hugo Drexler, Sindre Hovden, Rodrigo Gómez Vázquez, David Blacher, Dr. Siamak Rafiezadeh, and Dr. Johannes Kronsteiner for their valuable contributions to this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gooneie, A., Schuschnigg, S., Holzer, C. A Review of Multiscale Computational Methods in Polymeric Materials. *Polymers* 2017, 9, 16. <https://doi.org/10.3390/polym9010016>.
2. Horr A. M., Notes on New Physical & Hybrid Modelling Trends for Material Process Simulations; J. of Phy. conference Series, **2020**, Vol. 1603, 012008, <https://doi.org/10.1088/1742-6596/1603/1/012008>.
3. Liu, Cm., Gao, Hb., Li, Ly. et al. A review on metal additive manufacturing: modeling and application of numerical simulation for heat and mass transfer and microstructure evolution. *China Foundry* 18, 317–334 (2021). <https://doi.org/10.1007/s41230-021-1119-2>.
4. Bordas, A., Le Masson, P. & Weil, B. Model design in data science: engineering design to uncover design processes and anomalies. *Res Eng Design* 36, 1 (2025). <https://doi.org/10.1007/s00163-024-00442-w>.
5. Daria Vlah, Andrej Kastrin, Janez Povh, and Nikola Vukašinović. 2022. Data-driven engineering design: A systematic review using scientometric approach. *Adv. Eng. Inform.* 54, C (Oct 2022). <https://doi.org/10.1016/j.aei.2022.101774>.
6. Sumit K. Bishnu, Sabla Y. Alnouri, Dhabia M. Al-Mohannadi, Computational applications using data driven modeling in process Systems: A review, *Digital Chemical Engineering*, Vol. 8, 2023, 100111, <https://doi.org/10.1016/j.dche.2023.100111>.
7. Van de Berg D., Savage T., Petsagkourakis P., Zhang D., Shah N., Del Rio-Chanona E. A., Data-driven optimization for process systems engineering applications, *Chemical Engineering Science*, Vol. 248, Part B, 2022, 117135, <https://doi.org/10.1016/j.ces.2021.117135>.
8. Wilking F., Horber D., Goetz S., Wartzack S., Utilization of system models in model-based systems engineering: definition, classes and research directions based on a systematic literature review. *Design Science*. 2024;10:e6. <https://doi.org/10.1017/dsj.2024.3>.
9. Tao F., Li Y., Wei Y., Zhang C., Zuo Y., Data-model Fusion Methods and Applications toward Smart Manufacturing and Digital Engineering, *Journal of Engineering*, 2025, <https://doi.org/10.1016/j.eng.2024.12.034>.
10. Dogan A., Birant D., Machine learning and data mining in manufacturing, *Expert Systems with Applications*, Volume 166, 2021, 114060, <https://doi.org/10.1016/j.eswa.2020.114060>.
11. Wang J., Xu C., Zhang J., Zhong R., Big data analytics for intelligent manufacturing systems: A review, *Journal of Manufacturing Systems*, Volume 62, 2022, <https://doi.org/10.1016/j.jmsy.2021.03.005>.
12. Ghahramani M. H., Qiao Y., Zhou M., O'Hagan A., Sweeney J., AI-Based Modeling and Data-Driven Evaluation for Smart Manufacturing Processes, *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 1026-1037, July 2020, <https://doi.org/10.1109/JAS.2020.1003114>.
13. Sofianidis G., Rožanec J. M., Mladenčić D., Kyriazis D., A Review of Explainable Artificial Intelligence in Manufacturing, 2021, ArXiv, <https://doi.org/10.48550/arXiv.2107.02295>.
14. Horr, A., Gómez Vázquez, R. & Blacher D., Data Models for Casting Processes – Performances, Validations and Challenges, Sept 2024, In: IOP Conference Series: Materials Science and Engineering, Vol. 1315, <https://doi.org/10.1088/1757-899X/1315/1/012001>.
15. Horr, A., Blacher, D. & Gómez Vázquez, On Performance of Data Models and Machine Learning Routines for Simulations of Casting Processes, R., 8 Jan 2025, In: BHM Berg- und Hüttenmännische Monatshefte. 2025, <https://doi.org/10.1007/s00501-024-01537-6>.
16. Horr, A.M., Real-Time Modeling for Design and Control of Material Additive Manufacturing Processes. *Metals* 2024, 14, 1273. <https://doi.org/10.3390/met14111273>.
17. Horr, A.M., Drexler, H., Real-Time Models for Manufacturing Processes: How to Build Predictive Reduced Models. *Processes* 2025, 13, 252. <https://doi.org/10.3390/pr13010252>.
18. Wang J., Li Y., Gao R. X., Zhang F., Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability, *Journal of Manufacturing Systems*, Vol. 63, 2022, 381-391, <https://doi.org/10.1016/j.jmsy.2022.04.004>.
19. Brunton S. L., Kutz J. N., Data Driven Science & Engineering - Machine Learning, Dynamical Systems, and Control, Cambridge University Press, Cambridge, England, 2019, <https://doi.org/10.1017/9781108380690>.

20. Qin J., Hu F., Liu Y., Witherell P., Wang C. L., Rosen D. W., Simpson T. W., Lu Y., Tang Q., Research and application of machine learning for additive manufacturing, *Additive Manufacturing*, 2022, 52, 102691, <https://doi.org/10.1016/j.addma.2022.102691>.
21. Lebon B., directChillFoam: an OpenFOAM application for direct-chill casting. *Journal of Open Source Software*, 2023, 8(82), 4871, <https://doi.org/10.21105/joss.04871>.
22. The OpenFOAM Foundation: <https://openfoam.org>.
23. Bennon W.D., Incropera F.P., A continuum model for momentum, heat and species transport in binary solid-liquid phase change systems—I. Model formulation, *International Journal of Heat and Mass Transfer*, Volume 30, Issue 10, 1987, Pages 2161-2170, ISSN 0017-9310, [https://doi.org/10.1016/0017-9310\(87\)90094-9](https://doi.org/10.1016/0017-9310(87)90094-9).
24. Greenshields C., Weller H., Notes on Computational Fluid Dynamics: General Principles, CFD Direct Ltd, 2022, ISBN 978-1-3999-2078-0, 291 pages, <https://doc.cfd.direct/notes/cfd-general-principles/>.
25. Vreeman C. J., Schloz J.D., Krane M.J.M., Direct Chill Casting of Aluminium Alloys: Modelling and Experiments on Industrial Scale Ingots, *Journal of Heat Transfer* 124, 2002, 947-953. <https://doi.org/10.1115/1.1482089>.
26. directChillFoam Documentation: <https://blebon.com/directChillFoam/>.
27. Weckman D.C., Niessen, P., A numerical simulation of the D.C. continuous casting process including nucleate boiling heat transfer. *Metall Trans B* 13, 593–602 (1982). <https://doi.org/10.1007/BF02650017>.
28. Hoque S.E., Hovden S., Culic S., Nietsch J.A., Kronsteiner J., and Horwatitsch D., Modeling friction in hyperxtrude for hot forward extrusion simulation of AA6060 and AA6082 alloys, 25th International Conference on Material Forming (ESAFORM 2022), 2022, *Key Engineering Materials*, 926:416–425, 8, <https://doi.org/10.4028/p-qfc820>.
29. Rizkya I., Syahputri K., Sari R. M., Siregar I, Utaminigrum J., Autoregressive Integrated Moving Average (ARIMA) Model of Forecast Demand in Distribution Centre, *IOP Conference Series: Materials Science and Engineering*, Volume 598, 012071, <https://doi.org/10.1088/1757-899X/598/1/012071>.
30. Brötz, S.; Horr, A. M., Framework for progressive adaption of FE mesh to simulate generative manufacturing processes. *Manufacturing Letters*, <https://doi.org/10.1016/j.mfglet.2020.03.005>.
31. NovaFlow&Solid NovaCast Systems AB., 2022, <https://www.novacast.se/product/novaflowsolid/>.
32. Horr A. M., Computational Evolving Technique for Casting Process of Alloys, *Mathematical Problems in Engineering*, vol. 2019, Article ID 6164092, <https://doi.org/10.1155/2019/6164092>.
33. Horr, A. M., Real-time Modelling and ML Data Training for Digital Twinning of Additive Manufacturing Processes, *Berg Huettenmaenn Monatsh* 169, 48–56, 2024, <https://doi.org/10.1007/s00501-023-01416-6>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.