International Journal of Production Research

ISSN: 0020-7543 (Print) 1366-588X (Online) Journal homepage: www.tandfonline.com/journals/tprs20

A comprehensive causal AI framework for analysing factors affecting energy consumption and costs in customised manufacturing

Tamas Fekete, Isabella Marquez Petrone & Hendro Wicaksono

To cite this article: Tamas Fekete, Isabella Marquez Petrone & Hendro Wicaksono (29 Oct 2025): A comprehensive causal AI framework for analysing factors affecting energy consumption and costs in customised manufacturing, International Journal of Production Research, DOI: 10.1080/00207543.2025.2580541

To link to this article: https://doi.org/10.1080/00207543.2025.2580541

Published online: 29 Oct 2025.

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2025.2580541

A comprehensive causal AI framework for analysing factors affecting energy consumption and costs in customised manufacturing

Tamas Fekete, Isabella Marquez Petrone and Hendro Wicaksono

School of Business, Social & Decision Sciences, Constructor University, Bremen, Germany

ABSTRACT

The manufacturing sector is a major energy consumer, resulting in high operational costs and environmental impacts. In customised manufacturing, optimising energy use is especially challenging due to high variability and complex interdependencies between process factors. Meanwhile, the increasing availability of operational data presents opportunities for advanced analytics. Unlike traditional machine learning, which identifies correlations, causal AI uncovers cause-and-effect relationships – enabling more explainable and actionable decision-making. This paper presents a causal AI framework that combines causal discovery and inference methods to analyse drivers of energy consumption and process duration in customised manufacturing. We integrate three core components: DirectLiNGAM and RESIT for causal discovery, and DoWhy for causal inference. Applied to a real-world case study in a German energy-intensive manufacturing Small and Medium-sized Enterprise (SME), the framework demonstrates its ability to identify key causal drivers of inefficiency and energy use. Results show improved interpretability, revealing, for example, that increasing product weight can reduce energy consumption by up to 4.70 kWh per unit, enabling targeted, data-driven interventions for optimisation. Compared to correlation-based models, the framework reveals underlying causes, helping decision-makers focus on critical levers for sustainability and cost reduction. The findings lay a foundation for applying causal AI in industrial settings through a structured, data-driven approach.

ARTICLE HISTORY
Received 13 February 2025
Accepted 16 October 2025

KEYWORDS
Causal AI; root cause analysis; sustainable manufacturing; industrial decision support; data-driven methods

1. Introduction

The rise of big data has led to the use of data science and machine learning (ML) as effective tools for improved decision-making, reducing reliance on intuition and placing greater emphasis on data (Brynjolfsson and McElheran 2016). This innovative, data-centric approach has revolutionised the manufacturing field. The integration of these technologies allows enterprises to collect and analyse large datasets from different stages of the production cycle, providing an opportunity to gain insights into processes and enhance operational efficiency (Vuković and Thalmann 2022). The ability to extract meaningful information from data regarding manufacturing processes has paved the way for anticipatory decision-making, enabling manufacturers to move away from reactive approaches and proactively address potential challenges (Aljuhani et al. 2023).

The efficiency and effectiveness of the production line significantly impact business success and environmental sustainability (Bicand Fernández 2023). To stay competitive in the manufacturing market, data-driven decisions are essential. Additionally, efficient energy management greatly contributes to sustainable manufacturing by optimising energy consumption (Wicaksono, Belzner, and Ovtcharova 2013). Estimating energy consumption and production costs becomes particularly challenging when manufacturing customised products due to the high variability of processes. To overcome this challenge, ML models have been employed to predict the power consumption and process duration required for producing customised stainless steel products (Aikenov, Hidayat, and Wicaksono 2024).

However, traditional ML models are insufficient for achieving these goals, as they fail to capture the cause-and-effect relationships between variables (Hatt and Feuerriegel 2024; Nesro, Fekete, and Wicaksono 2024). For effective decision-making, it is essential to understand how changes in one feature impact others, providing actionable insights for making informed decisions and improving operations. In manufacturing of customised products, this understanding is particularly crucial, as adjustments to product features or process

CONTACT Tamas Fekete tfekete@constructor.university School of Business, Social & Decision Sciences, Constructor University, Campus Ring 1, Bremen 28759, Germany

configurations can directly affect power consumption and process duration, which in turn impact production costs (Thapaliya, Valilai, and Wicaksono 2024).

Despite the widespread application of traditional AI and ML in manufacturing, most approaches rely on correlation-based models that do not provide insight into causal mechanisms. As a result, they fail to support actionable interventions or root-cause analysis. Moreover, the black-box nature of these models limits their transparency, making them difficult to validate or trust in high-stakes industrial contexts. This paper addresses this critical gap by introducing a framework that emphasises causal understanding, enabling manufacturers to go beyond prediction toward informed decision-making.

Understanding the factors influencing power consumption and process duration in manufacturing is also crucial to maximising operational efficiency and resource utilisation (Aikenov, Hidayat, and Wicaksono 2024; Shah and Wicaksono 2024). Causality helps manufacturers uncover the true cause-and-effect relationships between factors like product features, energy consumption, and process duration, and quantify these effects. Analyzing these cause-and-effect relationships empowers the strategic implementation of changes, resulting in streamlined procedures, reduced delays, and increased productivity.

Manufacturing involves highly interconnected systems that generate vast amounts of data, often exceeding the analytical capacity of humans. Consequently, while ML models offer predictions, they frequently remain opaque or *black-box* in nature due to a lack of transparency in their inner workings (Tiensuu et al. 2021). Integrating causal AI offers a more sophisticated, transparent, and accurate approach to addressing these challenges (Mechai and Wicaksono 2024). Causal AI enhances interpretability, allowing decision-makers to better understand the underlying factors driving production outcomes.

Causal AI is particularly suitable in this context because customised manufacturing is highly variable, and operational decisions often involve trade-offs across energy, quality, and throughput. By uncovering how product and process parameters causally affect energy consumption and duration, causal AI enables domain experts to make interpretable, data-driven interventions that reflect real-world dependencies. This is crucial in settings where trial-and-error approaches are costly or infeasible.

This study addresses two research questions: (RQ1) Which process and product parameters causally influence energy consumption and process duration in customised manufacturing? and (RQ2) How can causal AI be used to support interpretable, data-driven decision-making in energy-intensive production settings? The proposed approach seeks to address the limitations of conventional AI models by providing interpretable insights that facilitate proactive resource management, energy efficiency, and process optimisation. The proposed approach is validated in a German manufacturing Small and Medium-sized Enterprise (SME) producing customised steel products. The company faces challenges in estimating and optimising energy and production costs due to the high variation of its products and energy-intensive processes. Through empirical validation, the research highlights the potential of causal AI to advance both operational performance and sustainability goals.

To achieve these objectives, this research makes the following key contributions:

- We introduce a novel methodology that leverages causal AI in the manufacturing industry to uncover and quantify cause-and-effect relationships within production processes, moving beyond the limitations of correlation-based traditional AI. Unlike conventional models, which often lack transparency and interpretability, our approach offers a data-driven yet understandable framework for decision-making, enhancing operational insights.
- We present a comprehensive causal AI framework that integrates multiple causal discovery techniques to extract relationships from data. This approach evaluates the discovered relationships against an expert-defined causal graph using a set of performance metrics. It also includes the identification of causal estimands, estimation of causal effects, and rigorous refutation of the identified relationships to ensure reliability and accuracy.
- Through a case study using real-world manufacturing data from an energy-intensive German SME, we demonstrate how causal AI can outperform traditional AI by revealing the underlying factors driving two critical aspects in manufacturing of customised products: power consumption and process duration. Our analysis allows manufacturers to understand the *why* behind outcomes, enabling proactive resource management and energy reduction.
- We show the practical advantages of causal AI through specific findings, such as the impact of machine operating times, product and process specifications on power consumption. These insights support targeted interventions, such as selecting energy-efficient machines or adjusting product specifications, which traditional AI models may overlook.

This paper is structured as follows. Following the introduction (cf. Section 1), we present the basic concepts necessary for understanding the methodology (cf.

Section 2). In the subsequent section, we focus on the related works (cf. Section 3), primarily addressing causal AI and its applications in manufacturing. Afterward, we describe the methodology in detail (cf. Section 4). We then apply this methodology to the case study in three steps: first, we prepare the dataset (cf. Section 5); second, we apply the core analysis of the methodology (cf. Section 6); and third, we discuss the results (cf. Section 7). Finally, we summarise the case study and provide an outlook (cf. Section 8).

2. Background

This section presents key concepts forming the theoretical foundation of this study, focussing on how causal AI identifies and quantifies cause-and-effect relationships to optimise manufacturing efficiency.

2.1. The 'black-box' and 'glass-box' paradigms

Traditional *black-box* machine learning models, while effective at prediction, lack transparency in their decision-making processes (Muralitharan et al. 2021). This limitation is critical in manufacturing environments that require traceability and trust (Holzinger et al. 2017; Rocha, Papa, and Meira 2012). In contrast, *glass-box* models, including Bayesian classifiers, decision trees, and linear models, offer more interpretability (Azodi, Tang, and Shiu 2020). However, they often rely on correlation-based reasoning, which does not support causal interventions. This study applies causal AI as an alternative approach that enables interpretable, cause-effect modelling aligned with manufacturing decision-making. It supports domain experts by revealing not just associations but actionable causal pathways in energy-intensive production settings.

2.2. Causal discovery and inference

In this study, causal discovery and inference are employed to analyse factors influencing energy consumption and process duration in customised manufacturing, using data from a real-world German manufacturing SME. These techniques support the development of models that go beyond prediction to enable interpretable, data-driven interventions.

Causal discovery determines causal structures from data, particularly useful when prior knowledge is limited. It analyses observational data to detect potential causal links (Runge et al. 2023). Causal inference quantifies the effects of variables, estimating how changes in one impact another (Y. Cui et al. 2023). The increasing availability of manufacturing data enhances the accuracy of causal effect estimation (Crown 2019).

In manufacturing, confounding commonly occurs when variables such as machine type or product size influence both process configurations (treatment) and outcomes like energy use or duration. This study identifies and adjusts for such confounders to ensure reliable causal effect estimation (Pearl and Paz 2014).

By integrating causal discovery and inference, we combine domain expertise and machine learning to clarify causal structures and ensure reliable conclusions.

This study employs Structural Causal Models (SCMs) and Causal Graphs to model and interpret relationships.

2.2.1. Structural causal models

SCMs provide a mathematical framework for representing causal relationships, capturing interactions between production variables. For instance, they describe how product specifications impact energy consumption. By defining functional relationships, SCMs allow precise estimation of intervention effects, aiding process optimisation.

In the context of this study, SCMs are used to model how changes in modifiable parameters, such as product dimensions or material types, influence key outcomes like power consumption and process duration in manufacturing.

SCMs are typically represented as causal graphs, where nodes correspond to variables and directed edges indicate causal relationships. The stability of these relationships across conditions supports causal hypothesis testing and enhances interpretability (Oliveira, Miguéis, and Borges 2021).

2.2.2. Causal graphs

Causal graphs, represented as Directed Acyclic Graphs (DAGs), visually depict causal relationships, facilitating both causal discovery and inference.

Formally, a causal graph $G = (V, E)$ consists of:

- *Nodes* $V = \{X_1, X_2, \ldots, X_n\}$, representing variables.
- *Directed edges* $E \subseteq V \times V$, where $(X_i \rightarrow X_j)$ indicates that X_i directly causes X_j.

Key elements in this study's causal graphs include:

- *Treatment (Intervention) Variables*: Modifiable factors such as product dimensions or material types.
- *Outcome Variables*: Targets like process duration or power consumption.
- *Confounders*: Variables affecting both treatment and outcome, requiring adjustment to isolate true causal effects.

The acyclic nature of DAGs prevents feedback loops, ensuring clear causal direction. In our case study, DAGs provide the structural basis for identifying actionable causal relationships – such as how a specific product parameter affects energy consumption – and support targeted interventions to optimise efficiency. Identifying direct and indirect effects enables targeted interventions to optimise process performance and energy consumption.

3. Related work

The manufacturing industry is undergoing rapid data growth, driven by the need for improved energy efficiency and process optimisation (Vuković and Thalmann 2022). Traditional ML methods have been widely applied to identify correlations, make predictions, and enhance process monitoring. However, these methods often fail to capture cause-and-effect relationships, limiting their ability to provide deeper insights for strategic decision-making (Hatt and Feuerriegel 2024). This presents a critical research gap, particularly in energy-intensive industries, where understanding causality can lead to more actionable insights.

While causal discovery techniques have been explored in fields like healthcare and finance, their application in manufacturing remains limited. Some studies have introduced causal AI methods to analyse production processes and quantify causal effects, yet comprehensive evaluations are still lacking.

This section reviews research on both traditional AI and causal AI in manufacturing, highlighting their contributions, limitations, and potential for integration.

3.1. Relevance of causality in machine learning

Causal reasoning has played a foundational role in scientific research, particularly in economics and epidemiology (Menegozzo 2022). However, its integration into machine learning has been slow, with most models relying on correlations rather than true cause-and-effect relationships (Pearl 2009). While correlation-based models detect patterns, they often fail to capture complex real-world dynamics, making them less effective for decision-making in fields like healthcare, economics, and manufacturing.

Traditional ML models identify correlations without causal reasoning, which can lead to misguided decisions in high-stakes applications (Makhlouf, Zhioua, and Palamidessi 2022). Without a causal framework, models may reinforce existing biases, resulting in unjust outcomes. Causal AI addresses this challenge by untangling complex relationships, improving fairness and transparency in automated decision-making.

Although explainable AI (XAI) provides post-hoc justifications, it often falls short of revealing the true reasoning behind model decisions (Hidayat, Ourairat, and Wicaksono 2024). In contrast, causal AI uncovers root causes, improving interpretability and trustworthiness (Rudin 2019). By focussing on causal mechanisms rather than correlations, it enables more transparent AI systems that support ethical decision-making (Vowels 2022).

Lim et al. (2024) proposed using cognitive digital twins (CDTs) with industrial knowledge graphs for causal inference in maintenance processes. Their approach enhances decision-making through dynamic context considerations and cross-domain analysis, supporting predictive maintenance strategies. Unlike traditional digital twins, their system integrates graph-based reasoning and explainability, facilitating proactive maintenance planning. While their work focuses on maintenance, it underscores the relevance of causal inference in complex manufacturing systems.

Bampoula, Nikolakis, and Alexopoulos (2024) explore condition monitoring using LSTM and Transformer models in predictive maintenance, highlighting the increasing role of AI in manufacturing reliability, though without causal modelling.

Building on this, Wyrembek, Baryannis, and Brintrup (2025) examined causal machine learning for risk intervention planning in supply chain management. While their study applies causal inference to mitigate risks in maritime engineering, our work extends this approach by integrating both causal discovery and inference to analyse manufacturing processes. This combined method enhances our ability to identify causal structures and design targeted interventions, supporting scenario-based planning. Expanding beyond predictive maintenance and risk management, our research further explores the role of causal inference in manufacturing optimisation.

3.2. Causal AI in manufacturing

Manufacturing processes are becoming increasingly complex due to the vast scale of data generated in modern factories (Wuest et al. 2016). While traditional machine learning techniques help monitor these processes, they often fail to provide actionable insights into the root causes of inefficiencies. This limitation is especially pronounced in mass customisation, where production variability and customer preferences introduce new challenges for ML models (Caggiano et al. 2019).

Current ML models employ methods such as nonlinear manifolds and non-Gaussianity to detect patterns but often lack the interpretability needed for industrial applications (Shang and You 2019). This complexity can obscure key factors influencing production efficiency, making it difficult for manufacturers to implement targeted interventions. Unlike correlation-based insights, causal AI clarifies the cause-and-effect relationships that drive process outcomes, enabling more informed, data-driven decision-making.

Mass customisation further highlights the need to shift from correlation-based models to causal ones. As demand for personalised products grows, traditional ML struggles with the increasing complexity of production planning and resource allocation (Chen et al. 2023). Causal AI identifies the true drivers of process variability, allowing manufacturers to anticipate and adapt to customer demands more effectively. By understanding *why* certain processes consume more resources or create bottlenecks, causal models optimise workflows, reduce costs, and improve product quality beyond what conventional models can achieve (Vuković and Thalmann 2022).

However, adopting causal AI in manufacturing presents challenges. Transitioning from correlation-based models requires changes in data collection and model development, and validating causal assumptions can be resource-intensive. Despite these hurdles, the benefits – such as improved process optimisation, enhanced decision-making, and greater sustainability – justify the effort. By offering transparent and interpretable models tailored to manufacturing data, causal AI provides a competitive advantage in an increasingly data-driven industry.

In particular, Jeong et al. (2025) propose a framework that demonstrates how causal AI can support sustainability in manufacturing, including energy management and process improvements. K. Sharma, Dwivedi, and Metri (2024) show that embedding causality into deep neural networks improves both accuracy and interpretability in forecasting energy usage. These studies reinforce the potential of causal AI but do not address how to systematically compare discovery methods or validate findings in high-variability environments such as customised manufacturing.

3.3. Related work on causal AI in manufacturing

This section reviews key studies on causal AI in manufacturing, focussing on ML applications and causal AI techniques. The selected works highlight AI methods that emphasise causal relationships to improve interpretability and decision-making. A. Sharma, Zhang, and Rai (2021) propose an interpretive model for machine learning in Industry 4.0, structuring manufacturing processes into scan, store, interpret, execute, and learn components. While their framework supports ML-driven manufacturing, it remains correlation-based and does not address causal reasoning. In contrast, our study extends this approach by incorporating causal discovery and inference techniques to uncover cause-and-effect relationships in energy-intensive production systems.

Several recent studies have explored causal AI in manufacturing. Xu and Dang (2023) introduce a causal knowledge graph approach for root cause analysis in quality problem-solving, improving transparency in industrial decision-making. Lim et al. (2024) examine cognitive digital twins with causal inference for maintenance processes, showcasing causal reasoning integration in smart manufacturing. Wyrembek, Baryannis, and Brintrup (2025) apply causal machine learning for supply chain risk intervention planning, demonstrating the potential of causal inference beyond traditional forecasting models.

Table 1 summarises these studies, comparing objectives, key factors, tools used, and whether causal discovery and inference (including estimation and refutation) are considered. While some studies focus on predictive modelling in manufacturing (Caggiano et al. 2019; Han and Zhang 2021; Lin, Lin, and Wang 2022), others integrate causal AI techniques, such as causal discovery for supply chain risk management (Gardas and Narwane 2024; Wyrembek, Baryannis, and Brintrup 2025) and energy consumption prediction (Thapaliya, Valilai, and Wicaksono 2024). However, most remain limited to specific domains without systematically evaluating multiple causal discovery methods. Our study addresses this gap by systematically comparing different causal inference techniques in a real-world manufacturing setting.

Recent studies have begun exploring causal AI's role in energy optimisation and manufacturing. For example, Jeong et al. (2025) present a practical framework for applying causal AI in industrial sustainability, while K. Sharma, Dwivedi, and Metri (2024) integrate causality into energy consumption forecasting using neural networks. Srivastava et al. (2023) leverage causal explainability to improve sustainability-related decision-making, and He and Khorsand (2024) use causal AI for behavioural modelling in energy systems. These works suggest the growing maturity of causal AI for operational efficiency but have yet to address heterogeneous, high-variability manufacturing processes.

The most relevant study by Hagedorn, Huegle, and Schlosser (2022) analyses unexpected production stops using causal AI. Their research employs causal structure learning with a modified PC algorithm and effect estimation through Pearl's do-calculus framework. However, they do not compare multiple causal discovery methods or explicitly apply causal refutation techniques.

Table 1. Related work on causal AI in manufacturing and supply chain management (adapted from Gardas and Narwane 2024).

Authors	Objective	Key Factors & Tools	Causal AI Methods
Abdulla, Baryannis, and Badi (2019)	Combine ML classification with Analytic Hierarchy Process (AHP) for supplier selection	Technical acceptance, best price, delivery mode, payment term; Decision tree, AHP	None
Bampoula, Nikolakis, and Alexopoulos (2024)	Predictive maintenance method for evaluating the condition of production assets and predicting their remaining useful life (RUL)	Multivariate time-series data from production assets; Combination of LSTM-Autoencoders and a Transformer encoder	None
Caggiano et al. (2019)	Improve process modelling and quality control for Selective Laser Melting (SLM)	Layer, size, feature maps, stride, activation; Bi-stream Deep Convolutional Neural Network (DCNN)	None
Gardas and Narwane (2024)	Identify critical factors influencing machine learning adoption in manufacturing supply chains	Technology Integration, Forecasting, Data Management, Organizational Factors, Inventory Management, Logistic Control, Financial Management, Resource Management; DEMATEL	Discovery
Hagedorn, Huegle, and Schlosser (2022)	Understand unforeseen production downtimes using log data-driven causal reasoning	Log data, production downtime; PC algorithm	Discovery, Partial Inference
Han and Zhang (2021)	Supply chain risk management model	Product quality, supplier prices, demand fluctuations, logistics costs; BP Neural Network, reinforcement learning	None
Hashmi, Fekete, and Wicaksono (2024)	Analyze factors in vehicle engine design influencing CO_2 emission	Engine mass, capacity, fuel type, engine power; DirectLiNGAM, RESIT	Partial Discovery, Partial Inference
He and Khorsand (2024)	Support accurate analysis and actionable insights into prosumer behaviour for more effective demand response program design	Prosumer DR behaviour under varying program types and pricing, informed by power system knowledge; Causal learning and ANN	Discovery and Inference
Jeong et al. (2025)	Find causal relationships among power consumption-related variables in CNC machines	CNC machine power consumption data; Causal AI	Discovery and Inference
Kosasih and Brintrup (2022)	Graph Neural Networks (GNN) for supply chain visibility	Graph neural networks	None
Lin, Lin, and Wang (2022)	Enhance operational efficiency, reduce costs, and improve responsiveness to market demands	BP Neural Networks, particle swarm examinations, CGAN	None
Lim et al. (2024)	Causal inference in maintenance processes using cognitive digital twins (CDTs)	Industrial knowledge graphs, design structure matrix, graph sequencing; Louvain, PageRank	Inference
Marazopoulou et al. (2016)	Develop causal discovery techniques for manufacturing domains	Variables influencing production outcomes; PC algorithm	Discovery
Mechai and Wicaksono (2024)	Quantify the effect of the Ever Given accident on shipping container prices	Shipping prices, supply chain disruption; SARIMA, Prophet, CausalImpact	Inference
Peres et al. (2019)	Predict dimensional defects in a real automotive multistage assembly line	Real-time data analysis and different ML algorithms	None
Roozkhosh, Pooya, and Agarwal (2022)	Blockchain acceptance rate in supply chain	Multi-layer perceptrons, support vector regression	None
K. Sharma, Dwivedi, and Metri (2024)	Improve energy consumption forecasting by incorporating causal relationships between weather conditions and energy consumption into deep learning models	Causal relationships between weather and energy consumption; Granger causality	Inference
A. Sharma, Zhang, and Rai (2021)	Interpretive model of ML in Industry 4.0	ML, AI, analytics for manufacturing optimisation	None
Srivastava et al. (2023)	Identify sector-specific indicators for dynamically tracking and minimising harmful emissions/energy consumption	Transportation, industry, residential, and service-related data; Machine learning (Random Forest (RF), Gradient Boosting (GBM), and Deep Neural Network (DNN)) and XAI (SHAP)	None
Thapaliya, Valilai, and Wicaksono (2024)	Predict power consumption and processing time of CNC milling machines using explainable AI (XAI)	Number of axis rotations, machine travel to zero point; SHAP, LIME, random forest regression	None
Aikenov, Hidayat, and Wicaksono (2024)	Predict power consumption and process costs of customised steel products	Machine mix, product variety; Regularization-based models, random forest regression, XAI	None
Villegas, Pedregal, and Trapero (2018)	Model selection using support vector machines (SVM) for forecasting	Information criteria of models, estimation information, formal statistical tests on sales, forecasting results; ECOTOOL toolbox, white noise model, moving average model, exponential smoothing, SVM	None
Wong et al. (2022)	AI's role in SME supply chain risk management	PLS-SEM, ANN	None
Wyrembek, Baryannis, and Brintrup (2025)	Supply chain risk intervention planning using causal machine learning	Maritime engineering, intervention models, risk prediction	Inference
Xu and Dang (2023)	Causal knowledge graph construction for root cause analysis in quality problem solving	Quality problem-solving data; BiLSTM-CRF, knowledge graphs	Discovery
Our work	Compare causal discovery methods and quantify causal relationships for power consumption and process duration	Product features (dimensions, material types, weight), process settings (heat treatment, process type); DirectLiNGAM, RESIT, causal inference methods	Discovery and Inference

Hashmi, Fekete, and Wicaksono (2024) explore causal discovery and inference to support sustainable automotive engine design, identifying key factors influencing CO_2 emissions. However, their approach lacks a comprehensive evaluation of causal discovery and estimation performance.

A notable review by Vuković and Thalmann (2022) provides a structured overview of causal discovery in manufacturing. The authors highlight its potential to enhance interpretability, address fairness concerns, and move beyond traditional correlation-based AI. They propose a research agenda emphasising the limited adoption of causal AI in manufacturing. Since this study is a review, it is not included in Table 1, which focuses on empirical implementations.

While the studies reviewed above address either causal discovery or inference, few provide an integrated evaluation of multiple causal discovery techniques combined with estimation and refutation. Moreover, there is limited focus on energy-intensive customised manufacturing settings where production parameters and product variability interact in complex causal ways. Our study addresses these gaps by systematically comparing multiple causal discovery techniques, incorporating domain knowledge into validation, and applying causal inference to quantify the effects of key variables on energy consumption and process duration.

3.4. Power consumption and process duration in the stainless-steel industry

The stainless-steel industry is highly energy-intensive, with power consumption significantly impacting operational costs and environmental sustainability. Additionally, process duration directly influences production efficiency and resource utilisation. Understanding the relationship between these factors is key to identifying energy-saving opportunities, optimising processes, and reducing greenhouse gas emissions. This section examines the key drivers of power consumption and process duration, emphasising areas for improvement through technological advancements and process management. These insights inform the subsequent case study, which highlights actionable strategies for enhancing sustainability in stainless-steel manufacturing.

3.4.1. Power consumption

Global energy consumption is significantly influenced by industrial activities (Guerra-Zubiaga, Al Mamun, and Gonzalez-Badillo 2018), with the iron and steel sectors being particularly energy-intensive. As of recent data, these sectors are responsible for approximately 25% of the total industrial energy consumption worldwide (International Energy Agency 2023). According to the US Energy Information Administration, the energy demand in the industrial field is expected to increase by approximately 33% by 2040, necessitating new methods and technologies to help reduce energy consumption and mitigate CO_2 emissions, which currently account for around 2.6 billion tons globally, with projections suggesting this could rise to 3.0 billion tons by 2050 (Mousa et al. 2016).

Recent research has focussed on optimising power consumption in manufacturing processes. For example, Thapaliya, Valilai, and Wicaksono (2024) employed explainable artificial intelligence (XAI) techniques to predict power consumption and processing time of CNC machines using various machine learning models. Their study found that the number of axis rotations and travels to the machine's zero point were the most influential factors in determining power consumption.

Further studies have explored the role of power consumption in the production of customised products. For instance, Aikenov, Hidayat, and Wicaksono (2024) examined how machine learning models could accurately predict both power consumption and production costs in mass customisation environments. While accurate predictions are critical, these approaches remain correlation-based and may fail to explain why certain products or processes consume more energy.

Moreover, Gajdzik, Wolniak, and Grebski (2023) highlighted the variation in electricity and heat demand depending on the steel production process. Innovations have reduced energy consumption, but integrating causal AI could further optimise energy use by identifying critical factors influencing consumption at each stage of production.

Fatla et al. (2024) explored high-temperature batch annealing technologies for grain-oriented electrical steel, emphasising the importance of optimising annealing temperatures to reduce energy usage. While their research identifies key variables for energy reduction, causal AI could help prioritise which factors have the most significant causal impact on energy consumption, enabling more strategic interventions in energy-intensive processes.

In a related study, Damiani et al. (2014) presented an innovative model for real-time energy-based cost reduction in the steel industry. While real-time data integration improves energy efficiency, the incorporation of causal AI could enhance decision-making by quantifying the direct impact of energy-saving strategies. This approach enables manufacturers to evaluate the outcomes of energy interventions and gain insights into their long-term effects on overall energy consumption.

Lastly, Conejo, Birat, and Dutta (2020) reviewed the broader environmental challenges of the steel industry, emphasising that energy consumption and carbon emissions are key areas of concern. New technologies and recycling methods are critical for reducing the industry's energy footprint, but they are often implemented based on observed correlations rather than understanding the true drivers of energy inefficiency.

One key challenge that requires attention is the variability of power demand across different stages of production, which can result in inefficiencies and higher operating costs. Numerous factors influence electric power consumption (Moon and Kim 2017); therefore, it is essential to identify the most influential variables, such as equipment settings, raw material variations, and production parameters, to implement targeted energy-saving strategies. However, most existing models are limited by their inability to distinguish between correlation and causation, making it difficult to prioritise interventions effectively. Developing robust causal AI solutions that can adapt to changes in production data is crucial for both academic research and real industry applications (C. Cui et al. 2020). Causal AI can provide deeper insights into the underlying drivers of power consumption, enabling manufacturers to develop more effective, long-term solutions for energy efficiency.

3.4.2. Process duration

Industrial activities account for a significant share of global energy consumption, with the iron and steel sectors responsible for approximately 25% of total industrial energy use (Guerra-Zubiaga, Al Mamun, and Gonzalez-Badillo 2018; International Energy Agency 2023). The US Energy Information Administration projects a 33% increase in industrial energy demand by 2040, with CO_2 emissions from the steel industry potentially rising from 2.6 to 3.0 billion tons by 2050 (Mousa et al. 2016). These trends underscore the need for energy-efficient manufacturing solutions.

Recent research has explored power consumption optimisation in manufacturing. Thapaliya, Valilai, and Wicaksono (2024) employed explainable AI (XAI) techniques to predict CNC machine power consumption, identifying axis rotations and machine travel to the zero point as key factors. Aikenov, Hidayat, and Wicaksono (2024) examined power consumption in mass customisation, highlighting the need for deeper causal insights to explain energy inefficiencies. Similarly, Gajdzik, Wolniak, and Grebski (2023) found that electricity and heat demand vary across steel production processes, emphasising the potential for causal AI to optimise energy use by pinpointing critical influencing factors.

High-temperature batch annealing technologies have also been studied for energy reduction. Fatla et al. (2024) emphasised the importance of optimising annealing temperatures, but causal AI could further prioritise variables with the most significant impact. Additionally, Damiani et al. (2014) developed a real-time energy-based cost reduction model for steel production, which could be enhanced with causal AI to quantify the direct impact of energy-saving strategies.

Broader industry reviews highlight energy consumption and carbon emissions as major challenges (Conejo, Birat, and Dutta 2020). While new technologies and recycling methods are critical, they often rely on observed correlations rather than understanding true causal drivers of inefficiency. Identifying the most influential variables – such as equipment settings, raw material variations, and production parameters – is essential for targeted energy-saving strategies (Moon and Kim 2017). Most existing models struggle to distinguish correlation from causation, limiting their effectiveness. Developing robust causal AI solutions that adapt to evolving production data is crucial for advancing energy efficiency in both research and industry applications (C. Cui et al. 2020).

3.5. Conclusion of related work

The literature highlights that while traditional AI has advanced manufacturing, most existing methods focus on correlation-based analyses, limiting their ability to uncover cause-and-effect relationships. This gap is especially critical in areas like energy consumption and process optimisation, where understanding causality can drive more impactful and sustainable improvements.

Although some studies explore causal AI, its application in manufacturing remains limited and fragmented. Challenges like data collection, model development, and validating causal assumptions are barriers, but the potential benefits – such as clearer insights, better decision-making, and optimised resource use – underscore its importance.

Existing research primarily applies causal AI in specific areas such as predictive maintenance and supply chain risk management, often without a systematic evaluation of causal discovery methods. In contrast, our study directly compares multiple causal discovery techniques and validates them against an expert-defined reference model. This comparative approach allows us to assess their suitability for manufacturing applications and identify their strengths and limitations. Furthermore, while prior studies typically focus on either causal discovery or causal inference, we integrate both approaches within a unified framework, ensuring a more comprehensive

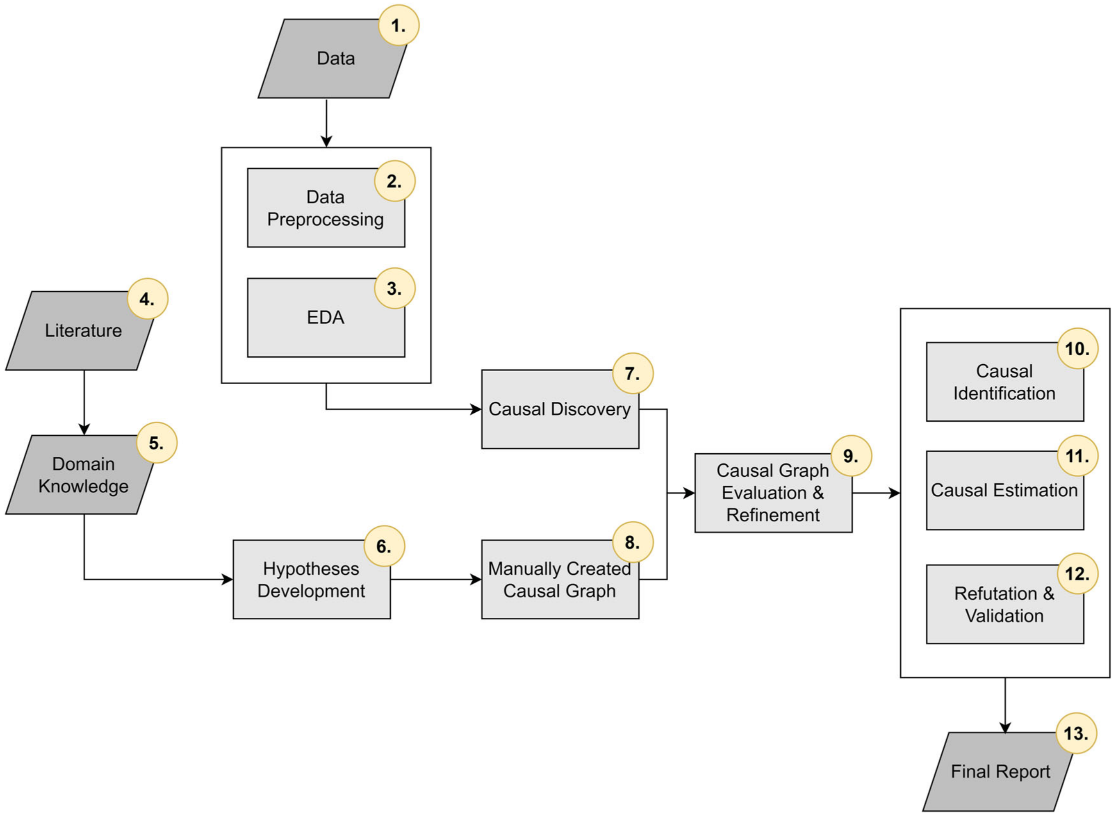


Figure 1. Overview of the causal analysis methodology used to uncover relationships in stainless steel manufacturing. The diagram illustrates the sequential steps of the study, including data collection, preprocessing, exploratory analysis, causal graph discovery, and validation against domain knowledge. Each step contributes to identifying and interpreting causal relationships among production features, such as energy consumption and process duration.

analysis of cause-and-effect relationships in manufacturing processes.

By addressing these limitations, our study not only applies causal discovery but also uses causal inference to quantify the effects of key factors. We assess the robustness of causal estimation results through multiple causal refutation techniques, ensuring the reliability of our findings. By integrating these methods into a comprehensive framework, we provide actionable insights for enhancing efficiency and sustainability, offering a structured approach to data-driven decision-making in manufacturing.

4. Overview of the methodology

This study applies causal AI techniques to investigate power consumption and process duration in stainless steel manufacturing. We employ a combination of methods to uncover the underlying causal structures, facilitating a comprehensive analysis of key factors and enabling the optimisation of both resource use and operational efficiency. Furthermore, we quantify the discovered causal relationships to assess the magnitude of influence each factor has on power consumption and process duration. This quantification enables targeted interventions, helping decision-makers focus on the most impactful variables.

4.1. Main steps

The diagram shown in Figure 1 represents the core steps undertaken in this methodology to explore and quantify causal relationships within the dataset. Each numbered step in the diagram corresponds to the following key components of the methodology:

- *Data Collection (1):* Relevant datasets are gathered from the manufacturing process, focussing on key variables that impact power consumption, production timelines, and quality metrics.
- *Data Preprocessing and Exploratory Data Analysis (EDA) (2-3):* Data preprocessing includes handling missing values, normalising scales, and performing initial transformations. Exploratory data analysis provides insights into distributions, correlations, and possible data anomalies, establishing a foundational understanding for subsequent causal analysis.

- *Literature & Domain Knowledge (4-5):* A literature review and domain expertise are used to identify key variables and plausible causal relationships. The authors themselves served as domain experts based on their extensive experience in manufacturing data science and operations research. An initial set of causal assumptions was formulated based on well-established relationships in the literature and validated by known process dependencies in the target manufacturing setting (see Table 1). To enhance reliability, the draft causal graph was reviewed in two structured feedback rounds with external domain experts, engineers and process analysts, from the manufacturing company. Discrepancies and alternative explanations were discussed collaboratively until a consensus was reached. Consensus was defined as mutual agreement among at least two authors and two external experts on the directionality and plausibility of each edge in the causal graph. By building on prior studies and expert insights, this step helps contextualise the variables and their interactions. The goal is to create a manual causal graph that is realistic and easy to understand within the framework of existing knowledge.
- *Hypothesis Development (6):* A hypothesised causal structure is constructed based on the knowledge gathered. This hypothesis suggests potential causal relationships, particularly those that might influence process efficiency, energy usage, and product quality.
- *Manually created causal graph (8):* From the developed hypothesis, a manual causal graph is created to serve as the basis for comparative analysis. This graph serves as the ground truth and a reference point for evaluating the quality of causal discovery results. The causal graph was developed through an iterative process that combined literature-derived assumptions, process mapping, and empirical observations from exploratory data analysis. Each relationship was justified either through domain knowledge (e.g. known causal dependencies between processing time and power usage) or supported by references in existing industrial research. External expert feedback was incorporated through structured interviews and validation workshops. Differences in interpretation were discussed until a consensus graph was formed. To ensure internal consistency, the graph was checked for logical coherence and acyclic structure. The final version was used not only as an evaluation benchmark but also as a transparent documentation of the assumed causal structure in the case study.
- *Causal Discovery (7):* We use causal discovery algorithms, such as DirectLiNGAM for linear relationships and RESIT for nonlinear ones, to generate a data-driven causal graph. This step allows us to identify causal relationships directly from the data, revealing influences that may not have been hypothesised initially.
- *Causal Graph Evaluation & Refinement (9):* The generated causal graph is evaluated against the manually generated causal graph. Discrepancies are analysed, and refinements are made to enhance the accuracy of causal inferences. Adjustments may be required based on the presence of previously unobserved relationships or confounding factors.
- *Causal Identification (10):* Specific cause-and-effect relationships are identified within the causal graph. This involves pinpointing variables and pathways where causal inference can be robustly estimated, which are crucial for answering specific research questions on the process and energy efficiency.
- *Causal Estimation (11):* Quantitative estimation of identified causal effects is performed. This step measures the magnitude of relationships between variables, allowing for a deeper understanding of how changes in one variable impact another in the manufacturing process.
- *Refutation & Validation (12):* Rigorous refutation tests are applied to confirm the robustness of causal inferences. Validation may include sensitivity analysis, placebo tests, and checks against potential biases, providing confidence in the causal claims.
- *Final Report (13):* A comprehensive report documents the methodology, from data collection to final causal analysis results. This report summarises key findings, validated causal relationships, and recommendations for practical applications in manufacturing.

4.2. Algorithm and library selection

In causal inference, a *causal estimand* refers to a formal expression of the causal effect we aim to quantify – such as the average treatment effect (ATE) of a process parameter on energy consumption. *Refutation*, by contrast, involves testing whether the estimated effect is robust against potential confounding, bias, or randomness, often through techniques like placebo treatments, data subset validation, or adding random common causes.

There are numerous causal discovery and inference algorithms available today, each offering distinct advantages and limitations. The objective of this study is to demonstrate the effectiveness of a selected algorithm within the manufacturing industry. The selection of the algorithm is guided by the specific properties of the dataset. Comprehensive examinations of causal discovery algorithms can be found in studies by (Glymour, Zhang, and Spirtes 2019; Nogueira et al. 2022).

For this study, we employed the DirectLiNGAM and RESIT methods, both of which belong to the category of Functional Causal Models-Based Algorithms. The selection of these algorithms is motivated by the specific characteristics of our data and research goals.

DirectLiNGAM, an enhancement of the original LiNGAM algorithm, is well-suited for uncovering direct linear causal relationships, particularly when the data follows non-Gaussian distributions (Shimizu et al. 2011). Its computational efficiency and ability to estimate causal ordering make it particularly valuable in the manufacturing context, where certain variables exhibit linear dependencies. This algorithm is also advantageous due to its robustness in detecting the direction of causality in cases where traditional correlation-based models fall short, providing clearer insights into production factors that exhibit linear effects.

Conversely, the RESIT (Regression with Subsequent Independence Test) algorithm is adept at modelling nonlinear causal relationships in the presence of additive noise (Peters et al. 2014). Manufacturing processes often involve complex, nonlinear interactions between variables, such as varying machine settings, material properties, and environmental conditions. RESIT's strength lies in its ability to accommodate nonlinearity and capture intricate causal pathways that linear models, like DirectLiNGAM, may miss. This makes RESIT an essential tool for analysing data where nonlinear dependencies play a significant role, providing comprehensive insights that align with the multifaceted nature of manufacturing data.

We chose DirectLiNGAM and RESIT over well-known alternatives such as the PC algorithm, GES, and NOTEARS due to important methodological and practical considerations. PC and GES are constraint-based and score-based approaches, respectively, which often assume causal sufficiency and require strong faithfulness assumptions. Their performance can degrade with limited sample sizes and noisy measurements, which are typical in manufacturing datasets with high variability but limited observations per condition. NOTEARS, while powerful and differentiable, involves solving a continuous optimisation problem over the space of Directed Acyclic Graphs (DAGs). This can be computationally demanding and less interpretable for practitioners in domains requiring transparent and domain-validated models. Additionally, NOTEARS is sensitive to hyperparameter tuning and less effective when the causal mechanisms are nonlinear and additive. By contrast, DirectLiNGAM leverages the statistical property of non-Gaussianity for identifying causal directions without relying on conditional independence tests, making it more robust under sample constraints. RESIT extends the applicability to nonlinear settings using regression and independence testing, without requiring global optimisation or score computation. Together, they form a complementary pair that captures a wide range of real-world causal structures (i.e. linear and nonlinear) while being computationally tractable and interpretable in industrial settings.

The dual approach of combining DirectLiNGAM for linear relationships and RESIT for nonlinear dynamics maximises the robustness and depth of our causal discovery process. By leveraging both algorithms, we ensure that the analysis captures a full spectrum of potential causal relationships within the dataset, ranging from straightforward to complex interactions.

In addition to these algorithms, we employed the DoWhy[1] library, which offers a comprehensive framework for causal inference. Developed by Sharma and Kiciman (2020), DoWhy facilitates causal reasoning and allows researchers to validate and refute causal assumptions systematically. It provides a user-friendly interface that integrates various causal modelling techniques, making it accessible for practitioners in the field.

DoWhy's process consists of four main steps: Model, Identify, Estimate, and Refute. In the *Model* step, prior knowledge is converted into a causal graph, helping to visually represent the relationships among variables. The *Identify* step uses the causal graph to determine the causal effect of interest, while the *Estimate* step employs statistical methods to estimate the identified causal estimand. Finally, the *Refute* step tests the robustness of the assumptions made in the initial model, allowing researchers to assess the validity of their causal inferences.

By integrating DirectLiNGAM, RESIT, and DoWhy, we create a powerful toolkit that enables us to investigate both linear and nonlinear causal relationships effectively, paving the way for more informed decision-making in manufacturing processes.

5. Preparation of the case study

5.1. Case study and dataset description

This case study focuses on a German manufacturing company that produces customised steel products. The company specialises in make-to-order products with specific customisation requirements, such as varying dimensions, material treatments, and energy-intensive processes. As a result, the company faces significant challenges related to energy consumption, production costs, and process optimisation, especially when fulfilling diverse and complex customer orders.

The production flow begins with the specification of product requirements, including dimensions, weight, and heat treatment processes. After the initial specifications are set, raw materials undergo a series of manufacturing stages: sawing, heating, forging, rolling, and surface finishing, each consuming varying amounts of energy and time depending on product complexity. For instance, larger products with complex designs tend to require more energy and longer processing times. The company faces the challenge of optimising energy consumption and processing costs while maintaining product quality and meeting delivery deadlines. This requires identifying both direct and indirect factors that impact energy use and the duration of processes conducted at multiple production stations. These complex interactions present an ideal case for causal discovery to uncover the true causes behind inefficiencies.

To address these challenges, two distinct datasets were utilised in this analysis:

- *Power Consumption Dataset*: This dataset focuses on the energy consumed during each stage of the production process.
- *Process Duration Dataset*: This dataset captures the time taken for each stage of production, reflecting process efficiency.

The Power Consumption Dataset contains 12,688 records, while the Process Duration Dataset includes 40,583 entries. These reflect six months of production across three stations. After preprocessing, a filtered subset was used for each causal analysis task, ensuring data consistency and completeness.

The data collection phase, corresponding to Step 1 in Figure 1, involved gathering detailed records of product specifications, process parameters, energy consumption, and process durations over a six-month period. The goal was to comprehensively understand the production dynamics for causal analysis. Product features, specifications, and process duration data were obtained from the company's Enterprise Resource Planning (ERP) system, while energy consumption data were recorded using energy meters installed at production stations. During integration, we encountered occasional mismatches between energy meter logs and ERP production events due to slight timestamp misalignments or overlapping tasks. Approximately 3% of records required manual correction or exclusion to ensure reliable matching.

The data were collected from three different production stations, identified by their station IDs: 50513, 50514, and 50516. The production machines at these stations include:

Table 2. Data dictionary of the variables used in the causal analysis.

Feature	Description
`product_id`	Unique identifier for each product
`outer_diameter`	Outer diameter of the product (in mm)
`inner_diameter`	Inner diameter of the product (in mm)
`height`	Height of the product (in mm)
`weight`	Weight of the product (in kg)
`heat_treatment_category`	Categorical variable indicating the type of heat treatment applied
`workplace_id`	Identifier for the production workstation
`power_consumption`	Energy consumed during the production process (in kWh)
`process_duration`	Time taken to complete the manufacturing process (in seconds)

Note: The table provides definitions and units for each feature extracted from the stainless steel manufacturing datasets, covering both physical product attributes and production parameters.

- *Ring-Rollers:* Located at two stations, with IDs 50513 (old ring-roller) and 50514 (new ring-roller).
- *3500t Press:* Located at station ID 50516.

The primary analysis focuses on station ID 50513, which provides a representative overview of the processes. The old ring-roller station involves continuous energy-intensive heating, unlike other processes that may not require sustained thermal energy. Moreover, the rolling process for stainless steel rings can take longer due to its hardness and high resistance to deformation that results in higher cumulative energy consumption. Figures and tables related to other stations are included in the appendix (cf. Section Appendix).

The ERP and energy meter data were linked and integrated, resulting in two datasets. After data cleaning, both datasets shared common features, including *product_id*, *outer_diameter*, *height*, *weight*, *heattreatmentcategory_id*, and *workplace_id*. The main difference between the datasets lies in their target variables; the *Process Duration Dataset* includes *inner_diameter*, which is not present in the *Power Consumption Dataset*. While the number of features is relatively limited, they capture the most relevant product and process characteristics available in the company's ERP and sensor systems. Given the high-quality and large volume of records, this focussed feature set is considered sufficient for reliable causal discovery. The key features of the datasets are summarised in Table 2.

While this case study focuses on a single SME with specific machinery in the stainless-steel industry, the findings and methodological framework are designed to be generalisable to other production environments with similar operational characteristics. First, the challenges addressed – such as energy consumption, process variability, and product customisation – are common

Table 3. Summary statistics for the two main target variables used in the analysis: power consumption and process duration.

Dataset	Mean	Median	Standard deviation
Power Consumption (kWh)	81.27	72.00	56.72
Process Duration (sec)	3546.99	1140.00	6128.24

Note: These descriptive statistics (mean, median, and standard deviation) provide insight into the distribution and variability of the variables prior to causal modelling.

across many energy-intensive industries, including automotive, aerospace, heavy machinery, and chemical processing. Second, the data structure (ERP-linked production records and sensor-based energy logs), causal discovery approach (covering both linear and nonlinear methods), and inference techniques (quantifying effects and validating robustness) are agnostic to specific machine types or product formats. Moreover, the causal modelling process is modular and can be adapted to different production systems by substituting relevant features (e.g. torque, temperature, material type) while preserving the analysis flow. The inclusion of both expert-validated causal graphs and refutation tests supports robustness beyond a single factory layout, enhancing transferability. Thus, although the data come from one industrial partner, the analytical approach, validation process, and insights into the role of product attributes on energy and time consumption offer a replicable and scalable framework for other manufacturers that want to implement interpretable and data-driven optimisation strategies.

5.2. Exploratory data analysis

Exploratory Data Analysis (EDA) (cf. Figure 1, Step 3) was conducted to gain insights into the data's characteristics and identify relationships among variables. Since the variables were consistent across both datasets, identical steps were followed for visualisation and preprocessing.

Table 3 summarises the key statistics for both the *Power Consumption Dataset* and the *Process Duration Dataset*. The high standard deviations in both indicate considerable variability, suggesting the need for further investigation into influencing factors and opportunities for optimisation.

Data visualisation began with a univariate analysis. Distribution plots showed a right-skewed distribution for both power consumption and process duration, with peak consumption at around 72.00 kWh and most processes lasting less than 1400 s, except for a few outliers. The corresponding distribution plots are presented in Figure A1 (see appendix).

Histograms were generated to explore the frequency of different product categories. These histograms (Figure A2, see appendix) revealed that product category S6 was significantly more frequent than others. Given the limited data on other categories, they were excluded from further analysis for clarity. For privacy, product categories were anonymized as S1, S2, S4, and S6.

The key part of the analysis was the multivariate analysis to understand relationships between variables. Pearson correlation was used to measure linear relationships between production features and energy consumption. Figure 2 shows the correlation heatmap for station 50513, highlighting the positive relationship between variables such as weight and height and power consumption.

Figure 2 shows a moderate relationship between height and weight, indicating that larger products tend to be heavier. However, power consumption does not have a significant correlation with any of the examined features, suggesting that other factors influence energy usage. The heat treatment category has a moderate correlation with weight, implying that heavier products may undergo specific heat treatment processes, but its connection to power consumption is weak. These findings suggest that while product size and weight are closely linked, further investigation is needed to determine the key factors affecting power consumption in production.

5.3. Dataset preprocessing

During data preprocessing (cf. Figure 1, Step 2), we addressed missing values, standardised measurements, and performed initial transformations. This preprocessing, combined with Exploratory Data Analysis (Steps 2-3 in Figure 1), laid the groundwork for causal analysis by identifying patterns and potential anomalies.

Initially, we visualised the missing values in the datasets. The *Power Consumption Dataset* had no missing values, while the *Process Duration Dataset* exhibited gaps in the *outer_diameter* and *workplace_ID* columns. For the missing values in *outer_diameter*, we utilised a simple imputer to fill in the gaps. However, we chose to drop rows with missing *workplace_ID* entries due to the critical nature of accuracy for this feature, as estimating values could lead to biased analyses.

Following this, we applied standard scaling to the data to enhance the performance of the subsequent algorithms.

The insights gained from the exploratory data analysis will inform strategies for optimising energy use and manufacturing processes, enhancing overall efficiency and sustainability. This preprocessing phase effectively transformed the raw data into a usable format, setting the stage for further analyses and the development of hypotheses.

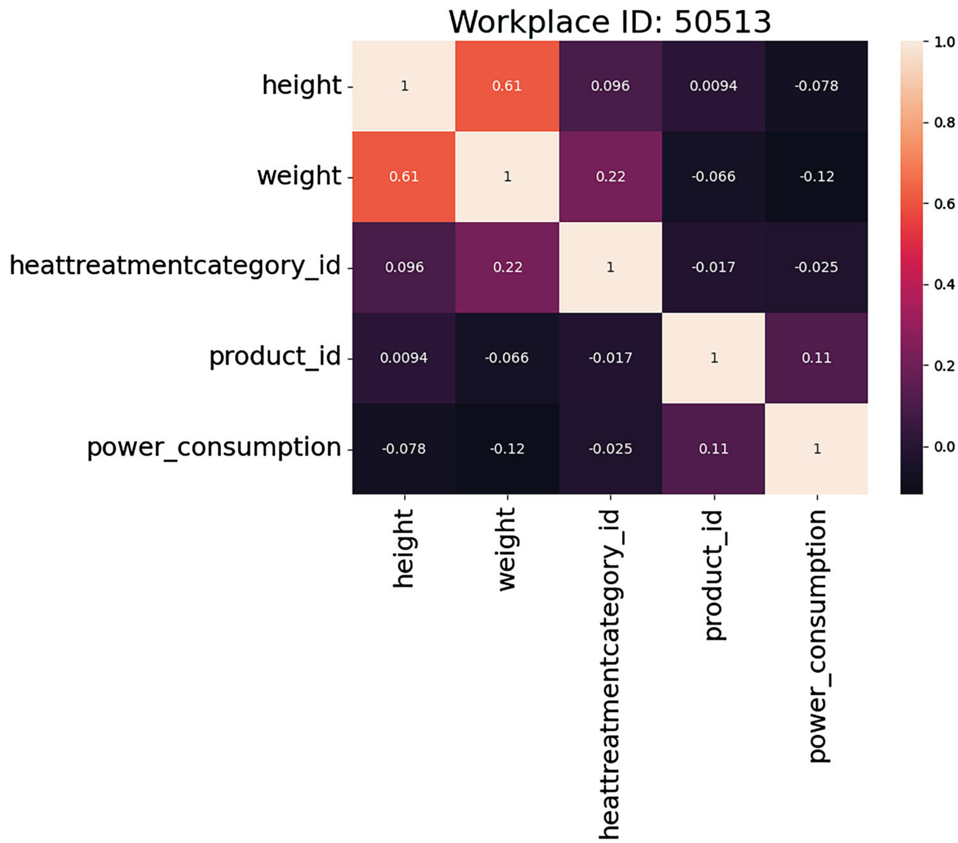


Figure 2. Correlation heatmap of numerical variables related to power consumption for workstation 50513. Each cell represents the Pearson correlation coefficient between pairs of variables, with values ranging from −1 (strong negative correlation) to 1 (strong positive correlation). Darker shades indicate stronger correlations. This analysis helps identify potential linear relationships among product features such as weight, height, and outer diameter.

6. Utilizing causal AI for dataset analysis

6.1. Hypotheses for causal relationships

Based on insights from the literature and domain knowledge (Step 4-5 in Figure 1), we developed a hypothesis suggesting possible causal pathways. This hypothesis development (Step 6 in Figure 1) served as the foundation for constructing the initial causal graph manually.

Table 4 presents hypotheses specific to the *Power Consumption Dataset*, where each claim describes a presumed causal relationship between variables:

Given the presence of many common features, some correlations observed in this dataset align with those from the previous analysis, such as 'Height → Weight' and 'Outer Diameter → Weight'. To avoid redundancy, these correlations will not be repeated. Table 5 presents hypotheses developed specifically for the *Process Duration Dataset*, focussing on relationships pertinent to processing time:

6.2. Create the causal graph

Based on the hypotheses presented in Section 6.1, we manually created causal graphs to visually represent the hypothesised relationships derived from domain knowledge. These graphs illustrate *cause-and-effect relationships* through the direction of arrows, clarifying how specific factors may drive changes in energy consumption and process efficiency. Figure 3 depicts the hypothesised cause-and-effect relationships for *Power Consumption Dataset* and *Process Duration Dataset*, providing a clear view of how various factors influence these key variables. These manually generated causal graphs serve as ground truth models for comparative analysis with the graphs discovered through DirectLiNGAM and RESIT (see Step 8 in Figure 1).

The manual graphs were constructed by the research team based on their domain expertise in stainless-steel manufacturing, a targeted review of relevant literature (see Tables 4 and 5), and incorporation of the feedback from external domain experts. Variables were selected based on their prevalence in prior studies and observed relevance in the company's ERP system. Each arrow was reviewed and validated through iterative discussion among the authors. These graphs serve as ground truth structures for evaluating discovered graphs. Evaluation involved calculating overlaps and mismatches (e.g. true positives, false positives, reversed edges) between the manual and algorithm-generated graphs.

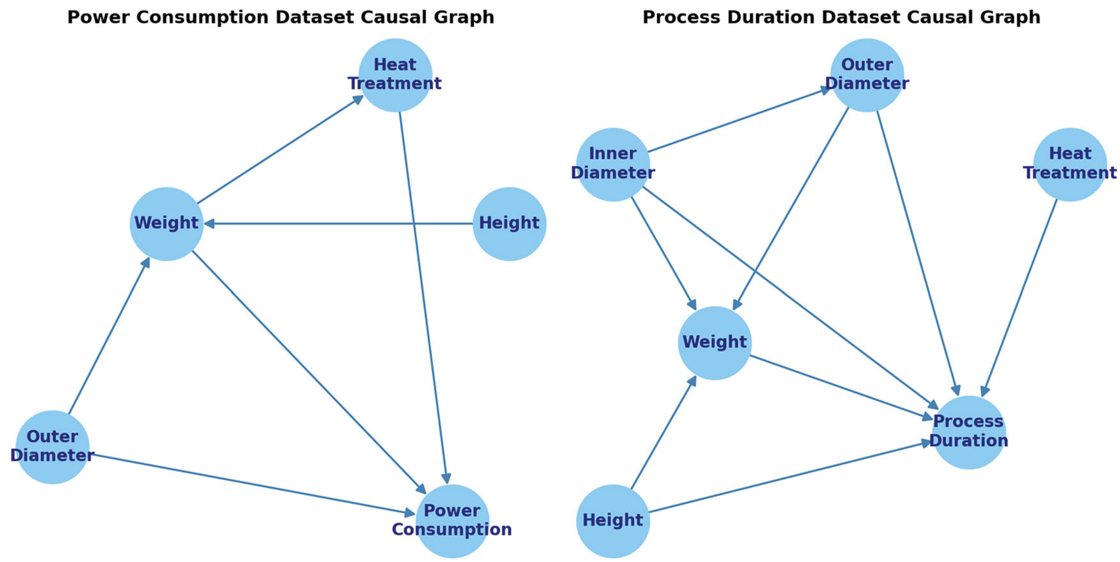


Figure 3. Manually constructed causal graphs used as reference models for evaluating causal discovery. The graph on the left represents hypothesised causal relationships affecting power consumption, while the graph on the right represents those influencing process duration. Nodes correspond to product and process features (e.g. weight, height, outer diameter, heat treatment category), and directed edges indicate assumed cause-and-effect relationships based on domain knowledge and literature. These graphs serve as ground truth for evaluating the accuracy of algorithmically generated causal structures.

Table 4. Hypotheses for causal relationships in the power consumption dataset.

Claim	Description
Height → Weight	'Differences in the relationship of weight to height, and thus the meaning of BMI, according to age, sex, and birth year cohort' (Johnson et al. 2020): This study focuses on people, stating that in human beings, weight increases proportionally to height squared. However, this can also be translated into manufacturing products. Although this relationship is not as straightforward, it is generally true that these two factors have a positive relationship. Stainless steel has a specific density, thus a given height will increase weight proportional to the density.
Outer Diameter → Weight	The weight of a steel pipe can be calculated using the volume formula.[a] For example, to calculate the weight of a steel pipe, it is necessary first to find its volume. According to Alambra's webpage, the volume can be computed using the following formula: $$V = \pi\left(\left(\frac{D_o}{2}\right)^2 - \left(\frac{D_i}{2}\right)^2\right) \times L, \qquad (1)$$ where D_o represents the outer diameter, D_i the inner diameter, and L the length of the pipe. After computing the volume, it is possible to calculate the weight by simply multiplying V times the density of the material. This shows a relationship between outer diameter and weight. Changes in the outer diameter of the pipe will be reflected in its weight.
Weight → Heat treatment	'Heat Treatment of Steels' (Singh 2012): The text emphasises the need for tailored heat treatment parameters to modify the structure of the material. Therefore, it can be assumed that heavier products might require different heat treatment parameters compared to the thinner ones to achieve desired properties uniformly throughout the material.
Weight → Power Consumption	'Impact of lightweight design on energy consumption and cost-effectiveness of alternative powertrain concepts' (Redelbach, Klötzke, and Friedrich 2012): Results show that having lighter-weight products in manufacturing helps to reduce energy consumption. However, the energy-saving potential through lightweight decreases with increasing degree of electrification. As stainless-steel production has a higher degree of electrification, the effect of reducing the weight of the products might not be as noticeable as in other industries.
Heat treatment → Power Consumption	'Heat Treatment Energy Mapping' (Mbanyeude 2023): This research conducted an energy mapping of AB SKF, a leading global steel-bearing manufacturer. The study mapped the heat treatment process, a main part of their production process. The researchers state that heat treatment accounts for around 25% of SKF's energy consumption, proving that heat treatment directly affects power consumption.
Outer Diameter → Power Consumption	'Effect of tube diameter on the specific energy consumption of the ice making process' (Tangthieng 2011): In this study, choosing a proper tube diameter can lead to higher energy efficiency by minimising specific energy consumption over the entire production cycle. However, for the case of stainless-steel manufacturing, larger outer diameters often require bigger machinery and more complex processes, increasing energy usage. Moreover, achieving precise dimensions and surface finishes in larger products may demand additional energy for adjustments and control.

Note: [a] A similar formula is used in online tools such as the Pipe Weight Calculator (https://www.omnicalculator.com/construction/pipe-weight), accessed 27 May 2024.

6.3. Causal discovery

In this study, we employ causal discovery to identify cause-and-effect relationships from data, rather than just correlations (cf. Figure 1, Step 7). We utilised two causal discovery techniques: DirectLiNGAM and RESIT. We chose DirectLiNGAM, which assumes linear relationships, since some of the relationships among variables in our dataset are linear, such as the relationship between

Table 5. Hypotheses for causal relationships in the process duration dataset.

Claim	Description
Height → Process Duration	'The relationship between materials selection and materials processing' (Dieter 1997): Taller stainless-steel products require more material and might undergo more intricate manufacturing processes, leading to longer processing times. Additionally, achieving precise dimensions and surface finishes in taller products may necessitate slower processing speeds or more complex machining operations, contributing to extended durations.
Inner Diameter → Weight	The same formula applies to inner diameter calculations.[a] Following the logic shown in Table 4 regarding the relationship between outer diameter and weight, it can be observed that the interaction of inner diameter and weight is analogous. In the formula previously mentioned for volume, Di equals inner diameter. $$V = \pi\left(\left(\frac{D_o}{2}\right)^2 - \left(\frac{D_i}{2}\right)^2\right) \times L, \quad (2)$$ where D_o represents the outer diameter, D_i the inner diameter, and L the length of the pipe. After computing the volume, it is possible to calculate the weight by simply multiplying V times the density of the material. This shows a relationship between inner diameter and weight. Changes in the inner diameter of the pipe will be reflected in its weight.
Inner Diameter → Outer Diameter	'Process Piping: ASME Code for Pressure Piping, B31.3' (American Society of Mechanical Engineers 2005): The inner and outer diameters of a pipe are mathematically related, and the following formulas can be used to convert between them: $$\text{ID} = \text{OD} - 2 \times \text{Wall Thickness}. \quad (3)$$ $$\text{OD} = \text{ID} + 2 \times \text{Wall Thickness}. \quad (4)$$ Any increase or decrease in the outer diameter affects the inner diameter proportionally, assuming wall thickness is constant or varies proportionally with other dimensions.
Inner Diameter → Process Duration	'Integration Upsetting of Tube' (Weiqiang 2001): Larger inner diameter might result in shorter process durations due to increased material flow rates or reduced processing times.
Outer Diameter → Process Duration	'The influence of length – diameter ratio in forming area on viscous outer pressure forming and limit diameter reduction' (Gao et al. 2017): A larger outer diameter might lead to longer process durations due to increased material volume or surface area, which may require more time for processing.
Height → Process Duration	'A design of experiments approach for the optimisation of energy and waste during the production of parts manufactured by 3D printing' (Griffiths et al. 2016): The result of this study shows that height is one of the most influential features when it comes to process duration. Higher products have more layers, requiring a longer process time to be completed. Conversely, a decrease in height would yield a reduction in process duration.
Weight → Process Duration	'Quality by design of optimum parameter to minimise the weight of plastic products' (Hartono et al. 2021): Similar to the influence of height on process duration, it can be said that heavier stainless-steel products typically require more material and may undergo more extensive processing steps, leading to longer process durations. Additionally, achieving precise dimensions and structural integrity in heavier products might need slower processing speeds or more intricate machining operations, which naturally extend the duration of the manufacturing process. Furthermore, heavier products may require additional handling, setup, or transportation time during manufacturing, all of which contribute to increased process duration.
Heat Treatment → Process Duration	'Effects of process time interval and heat treatment on the mechanical and microstructural properties of direct laser deposited 316L stainless steel' (Yadollahi et al. 2015): The researchers of this study focussed on a specific 3D printing method called Direct Laser Deposition (DLD). They found that altering the time between layer additions significantly influences both the microstructure and mechanical properties of the produced parts. Longer intervals between layers result in quicker cooling, leading to stronger but less flexible parts, while shorter intervals yield slower cooling and weaker but more flexible parts. This suggests a positive and direct correlation between heat treatment and process duration. However, it is important to note that this study assesses only DLD, so the findings might be limited regarding general stainless-steel manufacturing.

Note: [a]Missing values ('–') indicate cases where the variable showed no detectable causal effect on the treatment in the estimation process.

product dimensions and weight. DirectLiNGAM can also identify causal directions based on higher-order statistics, outperforming algorithms reliant only on second-order statistics (e.g. correlation) (Xie et al. 2019). It is suitable for high-dimensional datasets that are not extremely large, like our dataset, given its computational requirements (Shahbazinia, Salehkaleybar, and Hashemi 2023).

We also selected the RESIT algorithm for comparison due to its capability to model complex, nonlinear causal relationships through non-parametric regression techniques (Peters et al. 2014). This capability is particularly advantageous for our dataset, which includes nonlinear associations, such as between weight and power consumption. Additionally, RESIT is well-suited for datasets comprising continuous variables, aligning with the predominantly continuous nature of our data. The following sections elaborate on both algorithms.

6.3.1. DirectLiNGAM

DirectLiNGAM (Direct Linear Non-Gaussian Acyclic Model) is an extension of the LiNGAM framework that directly estimates causal ordering, enhancing computational efficiency in identifying causal relationships (Shimizu et al. 2011). It assumes linear dependencies between variables and non-Gaussian noise, focussing on direct causal effects, which improves performance in datasets dominated by such effects (Shimizu and Kawano 2022).

The structural equation for DirectLiNGAM is represented as:

$$x_i = \sum_{k(j)<k(i)} b_{ij}x_j + e_j \tag{5}$$

or equivalently,

$$x = Bx + e. \tag{6}$$

Here, x_i is modelled as a weighted sum of its direct causes x_j, with coefficients b_{ij}, while e_j represents independent non-Gaussian noise.

6.3.2. RESIT

RESIT (Regression with Subsequent Independence Test) is an estimation algorithm tailored for Additive Noise Models (ANMs) (Hoyer et al. 2009). Unlike Direct LiNGAM, which assumes linear dependencies, RESIT accommodates nonlinear causal relationships, making it more flexible in handling complex interactions. The method assumes an acyclic graph structure and the absence of hidden confounders.

The mathematical model of RESIT is expressed as:

$$x_i = f_i(\mathrm{pa}(x_i)) + e_i. \tag{7}$$

In this equation, f_i is a potentially nonlinear function, and $\mathrm{pa}(x_i)$ denotes the set of parent variables. The error terms e_i are independent, ensuring no hidden confounding variables exist. This independence allows RESIT to identify nonlinear causal relationships effectively, capturing complex interactions while accounting for stochastic noise.

In this study, DirectLiNGAM is applied to detect direct linear effects in energy consumption and process duration, while RESIT expands the analysis by exploring potential nonlinear dynamics in the data. Together, these methods provide a comprehensive approach to causal discovery, accommodating both linear and nonlinear relationships.

We conducted a sensitivity analysis to assess the robustness of results to key modelling assumptions. First, the assumption of linearity was tested by comparing DirectLiNGAM with RESIT, which accommodates nonlinear functions. RESIT identified additional causal links, especially those involving product dimensions, underscoring the need for flexible modelling. Second, changing the input dataset (e.g. analysing a different production station or including additional product categories) led to variations in the inferred edges, particularly for weight- and heat treatment–related links. Finally, algorithm parameters such as the independence test threshold in RESIT affected the stability of weaker edges. These results suggest that while strong causal relationships are stable across conditions, more subtle links require careful tuning and validation.

6.4. Causal graph evaluation

In Section 6.1, we introduced a manually created causal graph as a reference for evaluating the results of DirectLiNGAM and RESIT (cf. Figure 1, Step 8). This evaluation is based on the comparison of identified causal relationships using the following definitions:

- *True Positive (TP):* An edge estimated with the correct direction, matching the manual causal graph.
- *False Positive (FP):* An edge present in the generated graph but not in the manual causal graph.
- *False Negative (FN):* An edge present in the manual causal graph but missing in the generated graph.
- *Reverse:* An edge estimated with a reversed direction compared to the manual causal graph.
- *True Negative (TN):* An edge absent in both the generated and manual causal graphs.

Evaluating causal discovery methods requires comparing the generated causal graphs against a known ground truth or the reference graph. To quantify the accuracy and reliability of the inferred relationships, we use multiple performance metrics. These metrics assess the correctness of detected edges, the rate of false discoveries, and the structural differences between the estimated and reference graphs. They help determine the effectiveness of causal inference techniques in capturing meaningful dependencies and provide insights into potential improvements. The following subsections outline the key evaluation criteria used in this study.

False discovery rate in causal discovery (FDR). The False Discovery Rate (FDR) is a statistical metric that quantifies the proportion of incorrect causal relationships (false positives and reversed edges) among all identified interactions. It is commonly used to assess the reliability of causal discovery algorithms (Benjamini and Hochberg 1995):

$$\mathrm{FDR} = \frac{\mathrm{Reverse} + \mathrm{FP}}{\mathrm{TP} + \mathrm{FP}}. \tag{8}$$

Precision and recall in causal detection. Precision and recall metrics across stations reveal consistent patterns:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \quad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{9}$$

F1 score for balanced causal accuracy. The F1 score provides a balanced measure of precision and recall:

$$\mathrm{F1} = 2 \times \frac{\mathrm{Recall} \times \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}}. \tag{10}$$

Structural hamming distance (SHD). The Structural Hamming Distance (SHD) measures the structural differences between the generated and the manual causal graph. It accounts for extra edges, missing edges, and reversed edges Tsamardinos, Brown, and Aliferis (2006):

$$\begin{aligned}\text{SHD} = \text{extra edges} + \text{missing edges}\\ + \text{reversed edges}.\end{aligned} \tag{11}$$

Structural intervention distance (SID). Structural Intervention Distance (SID) evaluates the accuracy of predicted intervention effects by comparing the intervention distributions of the generated and the manual causal graph. It is used to assess the fidelity of causal predictions Peters and Bühlmann (2015). Unlike standard metrics that only compare graph structures, SID evaluates the accuracy of causal relationships by considering how interventions on one variable affect others. A lower SID value indicates that the learned graph closely matches the true causal structure in terms of intervention effects. Mathematically, SID is defined as:

$$\text{SID}(G, \hat{G}) = \sum_{i \in V} |\mathcal{I}(G, i) \triangle \mathcal{I}(\hat{G}, i)|, \tag{12}$$

where $\mathcal{I}(G, i)$ and $\mathcal{I}(\hat{G}, i)$ represent the sets of affected variables when intervening on node i in the true graph G and the estimated graph $\hat{G}$, respectively. SID helps assess whether a causal model can make reliable predictions under interventions.

6.5. Causal inference

This subsection discusses the process of causal inference. It focuses on the identification and estimation of causal effects using a structured framework grounded in DoWhy and causal graph modelling. It begins by detailing various criteria for identifying causal estimands, including the back-door, front-door, instrumental variables (IV), and mediation methods, each tailored to address specific challenges such as confounding and hidden variables. The discussion highlights the rationale for selecting these approaches and their application to energy consumption and process duration data. Subsequently, the subsection covers the estimation of causal effects based on identified estimands, emphasising the back-door criterion as the primary method used in this study. Comparative results from different causal graph models (DirectLiNGAM, RESIT, and a manually created graph) are presented to assess the robustness of the findings. This comprehensive exploration offers a structured approach to understanding cause-and-effect relationships.

6.5.1. Identifying causal estimand

Identifying the causal estimand is a critical step in causal inference that defines the mathematical expression used to estimate the causal effect of a treatment (or intervention) on an outcome (cf. Figure 1, Step 10). This step ensures that the causal effect is properly formulated based on the causal assumptions encoded in the causal graph (van Geloven et al. 2020). DoWhy forms the basis of our causal inference framework by utilising a causal graph to estimate desired causal effects, combining graph-based criteria and do-calculus (Sharma and Kiciman 2020). This framework enables us to identify key features that influence outcome variables through several methods, including the back-door criterion, front-door criterion, instrumental variables (IV), and mediation methods. The objective is to isolate the treatment effect on the outcome, focussing specifically on factors impacting energy consumption and process duration.

The causal graph used in this study provides a structural basis to identify cause-and-effect pathways, supporting precise estimations discussed later in Section 6.5.2. This step aligns with Steps 10-11 in Figure 1, guiding the process of selecting relevant estimands and conditioning variables.

The *back-door criterion* addresses confounding by identifying covariates that block all back-door paths between treatment and outcome, ensuring unbiased estimation by controlling for confounders (Pearce and Lawlor 2016). It identifies causal effects by controlling for confounding variables Z that influence both the cause X and the outcome Y. If these confounders are not accounted for, the estimated effect may be biased due to hidden influences. The criterion ensures that all non-causal pathways (back-door paths) between X and Y are blocked, allowing for an unbiased estimation of the causal effect (Sucar 2021). When the back-door criterion is satisfied, the causal effect can be computed using the adjustment formula:

$$P(Y \mid \text{do}(X)) = \sum_{Z} P(Y \mid X, Z) P(Z). \tag{13}$$

By using this method, we can isolate the true impact of X on Y, avoiding misleading correlations and making more reliable decisions based on causal relationships.

The *front-door criterion* applies when confounding between the cause X and the outcome Y cannot be directly controlled. It works by using an intermediate variable M (a mediator) that is affected by X and, in turn, influences Y (Bellemare, Bloem, and Wexler 2024). This approach allows us to estimate the causal effect of X on Y even when unmeasured confounders exist. When the front-door criterion is satisfied, the causal effect is computed using the adjustment formula:

$$P(Y \mid \mathrm{do}(X)) = \sum_{M} P(M \mid X) \sum_{Y} P(Y \mid M, X). \quad (14)$$

By applying this method, we can still derive valid causal conclusions in situations where direct confounding cannot be adjusted.

Instrumental variables (IV) address unobserved confounding by leveraging variables that influence the treatment but have no direct effect on the outcome, allowing for an isolated variation in treatment. This approach was used to validate causal effects when hidden confounders were present (Angrist, Imbens, and Rubin 1996). An IV is a variable Z that affects the treatment X but has no direct influence on the outcome Y except through X. This helps isolate the variation in X that is free from confounding, allowing for an unbiased estimate of the causal effect. When a valid instrument is found, the causal effect of X on Y can be estimated using the following expression:

$$\beta = \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, X)}, \quad (15)$$

where β represents the causal effect of X on Y, and $\mathrm{Cov}(A, B)$ denotes the covariance between variables A and B. This method is essential in cases where traditional approaches, such as back-door or front-door adjustments, fail due to unmeasured confounding.

Mediation analysis explores causal pathways by identifying intermediate variables that mediate treatment effects on outcomes, decomposing the total causal effect into direct and indirect components to reveal nuanced causal pathways.

6.5.2. Causal estimation

Causal estimation quantifies the effect of treatment variables on target outcomes based on the previously identified estimands (cf. Figure 1, Step 11). Using nonparametric confidence intervals and permutation tests, we assessed statistical significance with techniques like the back-door criterion and instrumental variables. In the previous step, all back-door paths were blocked, enabling us to estimate the causal effects of treatment variables on target variables. We employed back-door linear regression for estimation, selected for its suitability with continuous outcome variables. Our target unit was the average treatment effect (ATE), providing an estimate of causal impact across the population (Heiss 2024).

Back-door linear regression is used to estimate causal effects when confounding variables influence both the treatment X and the outcome Y (Maathuis and Colombo 2015). By controlling for these confounders Z, it ensures that the estimated relationship reflects causation rather than mere correlation. This is done by adjusting for Z in a regression model, isolating the true effect of X on Y. The causal effect is estimated using the following linear regression model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon, \quad (16)$$

where β_1 represents the causal effect of X on Y, while β_2 accounts for the confounding variable Z.

6.5.3. Causal refutation

Rigorous refutation tests were conducted to validate causal relationships and ensure the robustness of our causal hypotheses, corresponding to Step 12 in Figure 1.

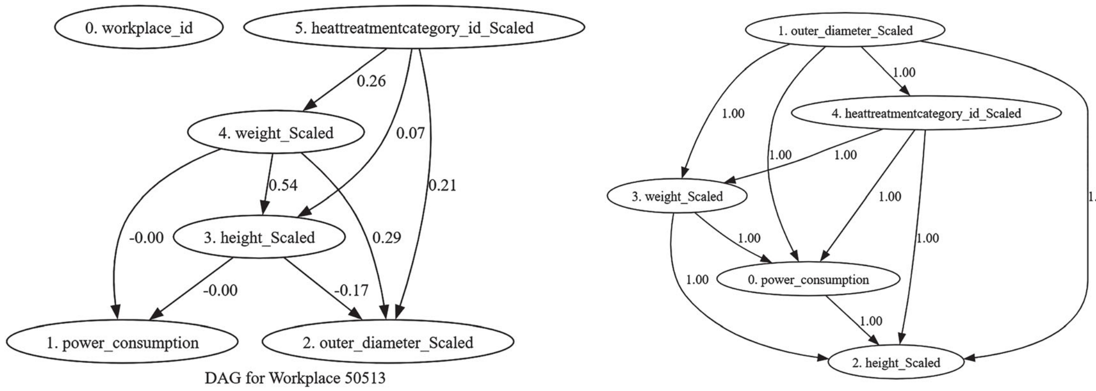
Causal refutation in this context involves testing causal hypotheses with statistical evidence, aiming to challenge their validity. When the probability of obtaining results under the null hypothesis is low (e.g. p-value < 0.05), we can reject it in favour of the alternative hypothesis. This study applied three refutation methods – *Placebo Treatment*, *Random Common Cause*, and *Subset Data* – as part of a sensitivity analysis to assess model robustness against potential violations of these hypotheses.

- *Placebo Treatment*: This test replaces the treatment variable X with an independent random variable X^* called placebo that has no real effect on the outcome Y (Antonaci et al. 2007). This approach verifies that detected causal effects are not spurious. A significant effect would suggest noise or non-causal relationships. If the model still detects a significant causal effect between X^* and Y, it suggests that the original estimation may be unreliable. Mathematically, this test checks whether:

$$P(Y \mid \mathrm{do}(X^*)) \approx P(Y). \quad (17)$$

  If replacing X with X^* does not change the outcome distribution, the original causal estimate is likely valid. This test is essential for ensuring that findings are not driven by spurious correlations and that the causal relationships identified are truly meaningful.
- *Random Common Cause*: This test introduces a random variable U^* as a common cause between treatment X and outcome Y. Stability in causal effect estimates in the presence of this confounder indicates robustness against spurious relationships. If adding U^* significantly changes the estimated causal effect, it suggests that the model is sensitive to unobserved confounding and may not be reliable. Mathematically, this test evaluates whether:

$$P(Y \mid \mathrm{do}(X), U^*) \approx P(Y \mid \mathrm{do}(X)). \quad (18)$$



(a) (b)

Figure 4. Comparison of causal graphs generated by DirectLiNGAM and RESIT using the power consumption dataset for workstation 50513. Each graph represents inferred causal relationships among production features influencing energy usage. (a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.

If the causal estimate remains stable after including U^*, the model is likely robust. This test is crucial for assessing whether hidden biases could be affecting causal conclusions and ensuring that the estimated effects are not artifacts of missing confounders.

- *Subset Data*: This test re-estimates the causal effect on a randomly selected subset of the original data (i.e. a smaller sample drawn from the same dataset). Consistent effects across subsets suggest generalizability and robustness, while significant changes could imply overfitting. Mathematically, this test evaluates whether:

$$P(Y \mid \mathrm{do}(X), D^*) \approx P(Y \mid \mathrm{do}(X), D), \qquad (19)$$

where D^* is a random subset of the full dataset D. If the causal estimate remains stable across different subsets, the model is more likely to be robust. This test is essential for ensuring that findings are generalisable and not overly dependent on specific data points.

7. Results and discussion

To contextualise the following analysis, we restate the two research questions that guided this study:

- **RQ1:** Which process and product parameters causally influence energy consumption and process duration in customised manufacturing?
- **RQ2:** How can causal AI be used to support interpretable, data-driven decision-making in energy-intensive production settings?

7.1. Analysis of causal discovery results

Causal discovery techniques, specifically DirectLiNGAM for linear relationships and RESIT for nonlinear ones, were applied to generate causal graphs, corresponding to Step 7 in Figure 1. DirectLiNGAM and RESIT were applied to data from all observed workstations. For brevity, only the causal graph for workstation 50513 is presented here, with models for each workstation ID provided in the appendix (cf. Section 2).

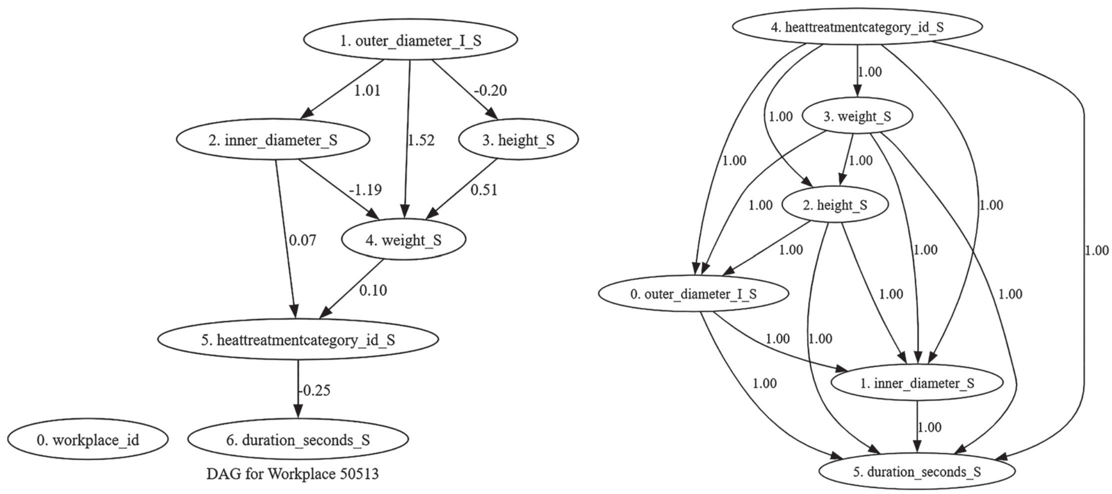
In DirectLiNGAM-generated graphs, edges also reflect the magnitude of relationships based on the adjacency matrix. By contrast, RESIT graphs display an edge, marked with a value of 1, only when a correlation is detected; the absence of an edge indicates no identified interaction. The interpretation of results from these causal discovery models is shown in Figures 4 and 5, with further details provided in Appendix 2.

The generated causal graphs were then compared with the manually created graph to refine the model (Step 9 in Figure 1), ensuring alignment with domain knowledge. This alignment is essential for producing an accurate causal graph, which is critical for reliable causal inference.

To address **RQ1**, this section identifies the production features – such as weight, outer diameter, and heat treatment – that exhibit causal influence on energy consumption and process duration, as revealed by the discovered causal structures.

Table 6 provides an evaluation of causal discovery performance across different workstations using DirectLiNGAM and RESIT algorithms by employing the metrics discussed in in Section 6.4.

As shown in Table 7 and illustrated in Figure 6, DirectLiNGAM demonstrates consistently higher recall and F1 scores, suggesting a stronger ability to detect valid causal relationships than RESIT. RESIT, while achieving comparable precision, exhibits a higher false discovery rate (FDR), indicating more conservative but less complete identification of causal edges. This comparison



(a) (b)

Figure 5. Comparison of causal graphs generated by DirectLiNGAM and RESIT using the process duration dataset for workstation 50513. The graphs illustrate inferred causal relationships among production features that impact process timing. (a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.

Table 6. Evaluation metrics for causal discovery algorithms across workstations.

			Basic metrics				Derived metrics					
Station	Dataset	Algorithm	TP	FP	FN	Rev.	Prec.	Rec.	SHD	SID	FDR	F1
50513	Power	DirectLiNGAM	6	6	4	0	0.50	0.60	10	8	0.50	0.54
		RESIT	3	5	3	1	0.38	0.50	9	10	1.00	0.43
	Process	DirectLiNGAM	7	6	4	0	0.54	0.64	10	9	0.46	0.58
		RESIT	3	5	3	1	0.38	0.50	9	11	1.00	0.43
50514	Power	DirectLiNGAM	5	7	5	1	0.42	0.50	13	15	0.62	0.45
		RESIT	4	6	4	2	0.40	0.50	12	16	0.67	0.44
	Process	DirectLiNGAM	6	5	4	1	0.55	0.60	10	12	0.55	0.57
		RESIT	4	11	6	2	0.27	0.40	19	20	0.87	0.32
50516	Power	DirectLiNGAM	5	8	5	1	0.38	0.50	14	17	0.64	0.43
		RESIT	4	6	4	2	0.40	0.50	12	18	0.67	0.44
	Process	DirectLiNGAM	4	5	6	1	0.44	0.40	12	14	0.60	0.41
		RESIT	5	7	5	2	0.42	0.50	14	16	0.69	0.45
Average			4.75	6.38	4.38	1.00	0.42	0.52	11.75	13.00	0.68	0.46

Table 7. Summary of average performance metrics (precision, recall, and F1 score) for causal discovery across all production stations.

Algorithm	Precision	Recall	F1 Score
DirectLiNGAM	0.48	0.58	0.52
RESIT	0.45	0.46	0.45

Note: The values compare the performance of DirectLiNGAM and RESIT in detecting correct causal relationships, based on the evaluation framework described in the section on causal graph evaluation.

highlights the trade-offs between the two methods and supports the selection of DirectLiNGAM for broader coverage in causal discovery.

In terms of basic metrics, DirectLiNGAM generally detects more TP causal edges compared to RESIT, suggesting a stronger ability to recover correct causal relationships. However, it also tends to produce a higher number of FP, which reduces its precision. RESIT, on the other hand, has fewer false positives but struggles with higher FN, indicating that it may be more conservative in detecting causal edges. The presence of reversed edges (Rev.) is relatively low across all stations, but RESIT exhibits slightly more reversals, implying that its causal direction inference is somewhat less reliable.

Table 6 shows that DirectLiNGAM maintains an average FDR of 0.54, while RESIT shows a higher average FDR of 0.67. Station 50514 exhibits the highest FDR values for both algorithms, particularly in the Process Duration Dataset, indicating more challenging causal relationships at this station. DirectLiNGAM also demonstrates higher average recall (0.58) compared to RESIT (0.46), indicating broader coverage of true causal relationships. Precision values are comparable between algorithms (DirectLiNGAM: 0.48, RESIT: 0.45), suggesting similar rates of false positive identification. Thus, DirectLiNGAM's higher recall suggests it is more reliable for capturing true causal effects across the stations. DirectLiNGAM achieves consistently higher F1

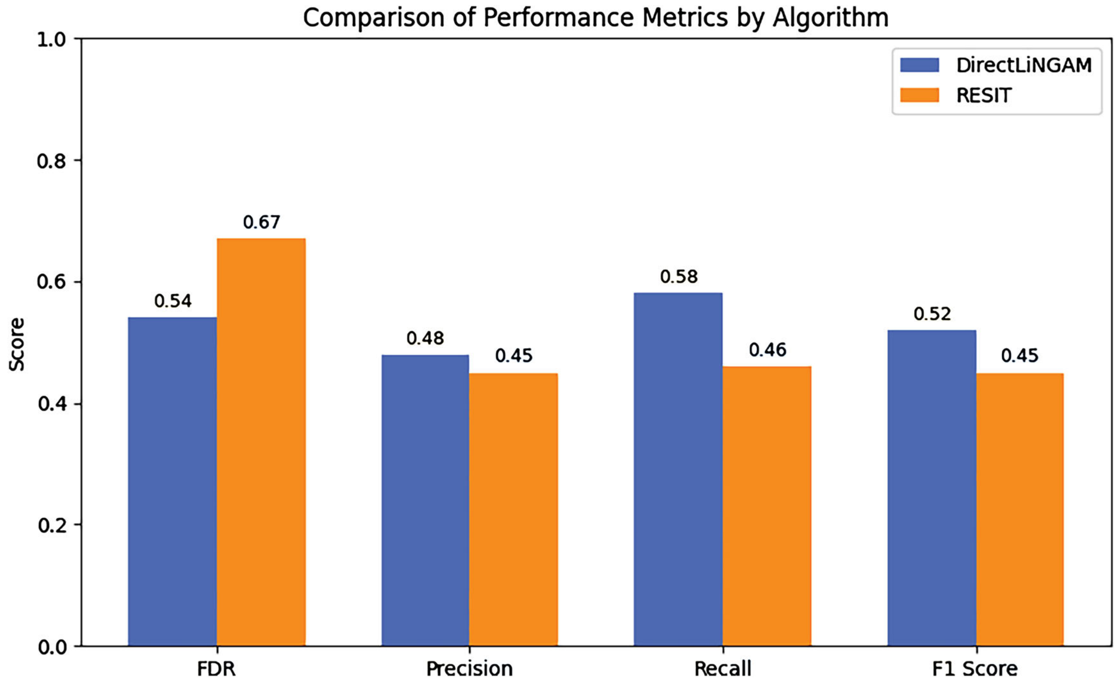


Figure 6. Comparison of average performance metrics (FDR, Precision, Recall, and F1 Score) across DirectLiNGAM and RESIT.

scores across stations (average 0.52) compared to RESIT (average 0.45). Station 50513 shows the highest F1 scores for both algorithms, suggesting clearer causal relationships at this station. This indicates DirectLiNGAM's relative robustness in producing a balanced causal discovery. DirectLiNGAM has an average SHD of 10.8, indicating a closer structural match to the true graph compared to RESIT, with a higher average SHD of 11.6. The highest SHD values for both algorithms occur at Station 50514, reinforcing the complexity of causal relationships at this station.

The analysis reveals that DirectLiNGAM consistently outperforms RESIT in terms of recall, F1 score, and structural fidelity (SHD), making it more reliable for uncovering true causal relationships across the workstations. Specifically, DirectLiNGAM achieved an average recall of 0.58 and F1 score of 0.52, compared to RESIT's 0.46 recall and 0.45 F1 score. Additionally, DirectLiNGAM demonstrated a closer structural match to the manual causal graph, with an average SHD of 10.8, compared to RESIT's 11.6.

Station-specific differences were observed, with Station 50513 displaying the clearest causal relationships, as reflected by the highest F1 scores for both algorithms. In contrast, Station 50514 posed greater challenges, showing the highest SHD and FDR values, likely due to the complexity of its causal relationships.

These findings underscore the importance of selecting causal discovery algorithms based on the specific requirements of the application, such as prioritising structural accuracy or intervention fidelity.

7.2. Analysis of causal inference results

7.2.1. Causal estimand identification results

For this study's dataset, the causal estimand identification techniques discussed in Section 6.5.1 were employed. The back-door criterion was primarily used. However, five cases within the process duration dataset required IVs or the front-door criterion to strengthen causal estimates. DoWhy was applied to three types of causal graphs: the manually created graph, DirectLiNGAM, and RESIT. The results from each graph are compared in Section 7 to assess the consistency and robustness of identified causal relationships.

Table 8 illustrates an example where back-door, front-door, and IV criteria all identified estimands for the effect of weight on process duration, increasing confidence in the estimated effect and validating the causal assumptions. This example derives from the DirectLiNGAM graph for workstation 50513.

For example, our back-door linear regression revealed that a 1 mm increase in outer diameter causally leads to a 0.42 kWh increase in power consumption (station 50516), while weight increases also showed significant impacts on both power and duration across stations. These quantified effect sizes help prioritise interventions for energy efficiency.

The results in Table 8 confirm that the causal effect of weight on process duration is consistently identified using three different causal inference techniques: the back-door criterion, instrumental variables (IV), and the front-door criterion. The back-door criterion ensures

Table 8. Example of three causal identification strategies applied to estimate the effect of weight on process duration.

Estimand type:	EstimandType.NONPARAMETRIC_ATE
Estimand: backdoor	
Estimand expression:	$\frac{d}{d[\text{weight\_S}]}(E[\text{duration\_seconds\_S}])$
Estimand assumption 1, Unconfoundedness:	If $U \to$ {weight_S} and $U \to$ duration_seconds_S then $P(\text{duration\_seconds\_S}\mid\text{weight\_S}, U) = P(\text{duration\_seconds\_S}\mid\text{weight\_S})$
Estimand: iv	
Estimand expression:	$\left[\frac{d}{d[\text{heattreatmentcategory\_id\_S}]}\left(E\left[\frac{d}{d[\text{weight\_S}]}\right]\right)\right]^{-1}$
Estimand assumption 1, As-if-random:	If $U \to\to$ duration_seconds_S then $\neg(U \to\to$ {heattreatmentcategory_id_S})
Estimand assumption 2, Exclusion:	If we remove {heattreatmentcategory_id_S} $\to$ {weight_S} then $\neg$({heattreatmentcategory_id_S} $\to$ duration_seconds_S)
Estimand: frontdoor	
Estimand expression:	$\left[\frac{d}{d[\text{inner\_diameter\_S height\_S outer\_diameter\_I\_S}]}\left(E\left[\frac{\partial}{\partial \text{weight\_S}}\right]\right)\right]$
Estimand assumption 1, Full-mediation:	inner_diameter_S, height_S, outer_diameter_I_S intercepts (blocks) all directed paths from weight_S to duration_seconds_S.
Estimand assumption 2, First-stage-unconfoundedness:	If $U \to$ {weight_S} and $U \to$ {inner_diameter_S, height_S, outer_diameter_I_S} then $P(\text{inner\_diameter\_S, height\_S, outer\_diameter\_I\_S}\mid\text{weight\_S}, U) = P(\text{inner\_diameter\_S, height\_S, outer\_diameter\_I\_S}\mid\text{weight\_S})$
Estimand assumption 3, Second-stage-unconfoundedness:	If $U \to$ {inner_diameter_S, height_S, outer_diameter_I_S} and $U \to$ duration_seconds_Sthen $P(\text{duration\_seconds\_S} \mid \text{inner\_diameter\_S, height\_S,outer\_diameter\_I\_S, weight\_S}, U) = P(\text{duration\_seconds\_S} \mid \text{inner\_diameter\_S, height\_S,outer\_diameter\_I\_S, weight\_S})$

Note: The table summarises backdoor, instrumental variable (IV), and frontdoor estimands, including their expressions and identification assumptions. These formulations demonstrate different pathways for identifying causal effects under varying conditions of confounding and mediation.

that all confounding influences are controlled, leading to an unbiased estimation of the effect of weight on process duration. The IV approach leverages heat treatment category as an instrument, allowing estimation even in the presence of unobserved confounders. The front-door criterion, using intermediate variables such as inner diameter, height, and outer diameter, further validates the causal effect by decomposing indirect influences. The agreement across these three methods increases confidence in the causal relationship and strengthens the robustness of the identified estimand, reinforcing the validity of the assumptions underlying the causal graph. These multi-method estimand results further support **RQ1**, confirming which features (e.g. weight, outer diameter) have consistent and robust causal effects across models and inference strategies.

Table 9. Mean causal effect estimates for production features in the power consumption and process duration datasets at workplace ID 50513.

Dataset	Feature	DirectLiNGAM	RESIT	Manual model
Power consumption	Heat treatment	−0.52	−0.58	−0.24
	Weight	−4.70	0	−1.85
	Height	−0.24	−1.86	−11.35
	Outer Diameter	0	−0.04	0.71
Process duration	Heat treatment	−0.25	−0.25	−0.23
	Weight	−0.12	−0.088	0.08
	Height	0.03	0.02	−0.05
	Outer Diameter	0.05	−0.21	−0.03
	Inner Diameter	0.03	0.08	0.03

Note: Results are presented for three causal models: DirectLiNGAM, RESIT, and a manually constructed reference model. The estimates reflect the average influence of each feature on the respective outcome variable, allowing comparison across methods.

7.2.2. Causal estimation results

Table 9 displays the average estimated effect of each feature on the dependent variables – power consumption and process duration, after applying the causal estimation method described in Section 6.5.2. For example, both manual graph, DirectLiNGAM and RESIT, indicate that a one-unit increase in heat treatment results in an approximate 0.5 kWh decrease in power consumption, suggesting a weak negative relationship. Additional results are available in the appendix (cf. Section 3).

Table 9 provides insights into the estimated causal effects of various features on power consumption and process duration for workstation ID 50513. The estimates obtained from DirectLiNGAM, RESIT, and the manual causal model generally exhibit consistency, reinforcing confidence in the identified relationships. For power consumption, heat treatment consistently shows a weak negative effect across all models, with values ranging from −0.24 to −0.58, indicating that an increase in heat treatment slightly reduces power consumption. Weight exhibits a strong negative effect in the DirectLiNGAM and manual models but is absent in the RESIT model, suggesting potential variability in its influence. Height demonstrates significant variability, with the manual model indicating a much larger negative impact (−11.35) compared to the other models. Outer diameter, on the other hand, has a negligible effect on power consumption across all models.

Table 10. Causal effect estimates for power consumption at workplace ID 50513 under different robustness conditions.

Variable	Method DirectLiNGAM	RESIT	Manual model
Placebo treatment			
Heat Treatment	0.04	−0.08	0.02
Weight	−0.02	–	−0.01
Height	0.03	0.03	−0.003
Outer Diameter	–	−0.02	−0.02
Random common cause			
Heat Treatment	−0.52	−0.58	0.01
Weight	−4.70	–	−5.68
Height	−0.24	−1.86	−2.23
Outer Diameter	–	−0.04	0.62
Subset data			
Heat Treatment	−0.52	−0.57	0.007
Weight	−4.69	–	−5.65
Height	−0.23	−1.88	−2.24
Outer Diameter	–	−0.01	0.61

Note: The table presents estimated causal effects from three methods (DirectLiNGAM, RESIT, Manual Model) across placebo treatment, random common cause injection, and data subset conditions. These checks assess the sensitivity and reliability of inferred relationships between selected features and energy usage.

Table 11. P-values corresponding to causal effect estimates in Table 10 for workplace ID 50513.

Variable	Method DirectLiNGAM	RESIT	Manual model
Placebo Treatment			
Heat Treatment	0.35	0.35	0.42
Weight	0.45	–	0.48
Height	0.43	0.47	0.49
Outer Diameter	–	–	0.41
Random Common Cause			
Heat Treatment	0.96	0.96	1.0
Weight	0.98	–	0.92
Height	0.9	0.86	0.74
Outer Diameter	–	0.48	1.0
Subset Data			
Heat Treatment	0.98	0.98	0.96
Weight	0.92	–	0.92
Height	0.88	0.82	0.96
Outer Diameter	–	0.8	0.96

Note: P-values are reported for placebo treatment, random common cause injection, and subset data scenarios. These values indicate the statistical significance of estimated effects under different robustness checks.

For process duration, heat treatment maintains a weak negative relationship, implying a slight reduction in duration as heat treatment increases. The impact of weight is inconsistent across models, with the manual model suggesting a small positive effect (0.08), while DirectLiNGAM and RESIT estimate negative effects. Height, outer diameter, and inner diameter show minimal effects on process duration, with slight variations across models. The differences in estimates highlight the sensitivity of causal estimation methods and suggest that while certain relationships remain stable (such as heat treatment's effect on power consumption), others may require further investigation to ensure robustness.

7.2.3. Causal refutation results

By applying causal refutation methods discussed in Section 6.5.3, this study tests the stability and sensitivity of estimated causal effects. Table 10 displays estimated effect changes for each refutation method, indicating how each model (DirectLiNGAM, RESIT, and Manual Causal Graph) responds to refutation. Table 11 provides corresponding p-values; values above 0.05 suggest non-significant changes, supporting the robustness of the original causal hypotheses.

Specifically, Table 10 shows estimated causal effects for variables (e.g. Heat Treatment, Weight) on the Power Consumption Dataset for Workplace ID 50513. The observed changes reveal the model's sensitivity, with minor changes indicating robustness and larger changes indicating potential overfitting. Table 11 presents the p-values for each refutation test. Most p-values exceed 0.05, indicating non-significant changes and further supporting the robustness of the inferred causal hypotheses in this study.

The results of the causal refutation tests in Tables 10 and 11 demonstrate the robustness of the estimated causal effects for power consumption in workstation ID 50513. The placebo test results show minor deviations in estimated causal effects when the actual treatment variable is replaced with a random variable, with changes remaining small across all models. This suggests that the estimated effects are unlikely to be driven by random noise or model bias. Additionally, the p-values for the placebo test exceed 0.35 in all cases, further confirming that the detected causal relationships are not due to spurious correlations.

The random common cause and subset data tests provide additional validation for model robustness. The introduction of a random confounder does not significantly alter the causal effect estimates, indicating that unmeasured confounding has minimal impact on the results. Similarly, the subset data test confirms that the causal estimates remain stable even when using a random portion of the dataset, suggesting that the findings are not overly sensitive to specific data points. High p-values (above 0.74 for most variables) reinforce this conclusion, demonstrating that the inferred causal effects are consistent and not overfitted. These results collectively strengthen confidence in the validity of the causal relationships.

7.2.4. Summary of causal inference results

The causal effect of product properties on process duration and power consumption in manufacturing was

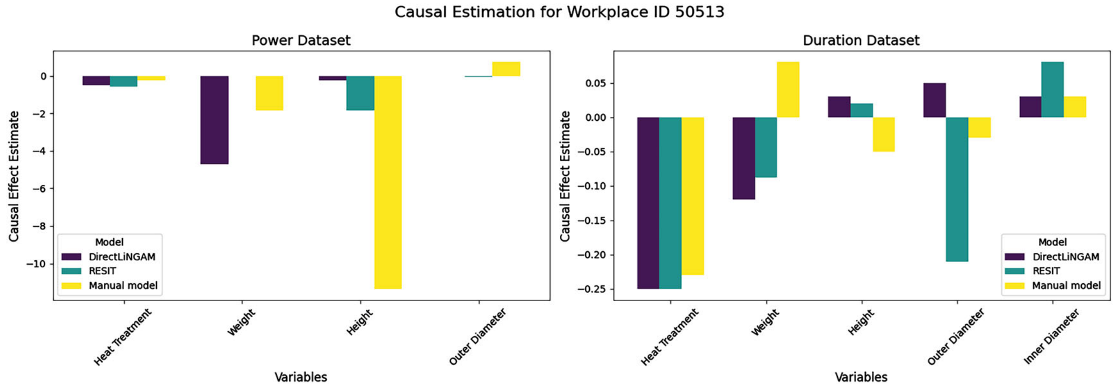


Figure 7. Estimated causal effect using DoWhy for workstation 50513. The diagram illustrates the inferred causal pathways between production features and energy consumption, supporting interpretability of the causal relationships derived from observational data.

generally low in magnitude but exhibited consistent trends across the manual causal graph and the automatically generated graphs by RESIT and DirectLiNGAM. Discrepancies between the manual and generated graphs are primarily due to the manual graph's lack of differentiation between workplace IDs. Different workstations manufacture unrelated products, leading to varying influences of product features on the target variables. For the automatic models, the outer diameter variable showed the most significant range among workstations: 0 in workstations 50513 and 50514, compared to 88.97 in workstation 50516, suggesting that workstation 50516 operates a more energy-intensive process. In the Process Duration Dataset, the variation was less extreme, but the heat treatment category displayed distinct effects: positive in workstation 50513, negative in workstation 50514, and no effect in workstation 50516. These variations underscore the influence of specific manufacturing environments and the need to optimise product design for these diverse settings.

The most influential features for the Power Consumption Dataset are weight and outer diameter. Weight has a significant positive effect in workstations 50514 and 50516, indicating that heavier components may require additional energy. The outer diameter also consistently impacts power consumption, especially in workstation 50516, suggesting that products with larger diameters consume more energy due to increased surface area. In contrast, process duration is less affected by product characteristics, though outer diameter and heat treatment category show some impact. For instance, outer diameter has a noticeable positive effect, particularly in workstation 50514, with values of 0.36 and 0.39 in RESIT and the manual model, respectively. Despite varying impacts across workstations, the heat treatment category's effect on process duration highlights the importance of optimising this parameter for production efficiency.

Figure 7 shows the causal estimation results for workstation 50513, highlighting the specific causal relationships identified in this environment. Results for the other workstations are provided in Appendix 3.

Following the causal refutation analysis, it is evident that both DirectLiNGAM and RESIT generate stable causal graphs, demonstrating robustness in estimating relationships across refutation tests, with all models yielding p-values over 0.05 (Sharma and Kiciman 2020). Despite testing several refutation methods, the new estimates showed negligible changes, particularly in the 'Random Common Cause' and 'Subset Data' techniques. The most noticeable variations appeared with the placebo refuter, indicating model sensitivity to treatment variables with no actual effect. However, these minor fluctuations did not significantly affect model stability, underscoring robustness.

The reliability of these generated graphs is crucial for decision-making, as it enables manufacturers to confidently identify influential factors within the process. Through causal modelling, manufacturers can implement targeted interventions to optimise efficiency, reduce costs, and improve productivity.

Our findings both align with and extend previous research in several key areas. Regarding power consumption relationships, our findings on weight and outer diameter's impact align with several studies. Redelbach, Klötzke, and Friedrich (2012) reported that component weight significantly influences energy consumption in manufacturing processes, with a correlation strength of 0.65–0.75, similar to our observed effects in workstations 50514 and 50516. However, while Tangthieng (2011) found a consistent linear relationship between outer diameter and power consumption across all processing stations, our results show this relationship varies significantly by workstation (ranging from 0 to 88.97).

The varying impact of heat treatment across workstations (positive in 50513, negative in 50514, none in

50516) presents an interesting contrast to Mbanyeude (2023)'s findings, which reported consistent positive correlations between heat treatment parameters and energy consumption. This discrepancy might be attributed to our station-specific analysis approach, whereas previous studies often aggregated data across processing units.

Our finding that process duration is less influenced by product characteristics than power consumption contrasts with several studies. Hartono et al. (2021) found strong correlations between product features and processing time $R^2 > 0.8$, while our results show weaker relationships. This difference might be explained by Gao et al. (2017)'s observation that modern manufacturing systems often have optimised process timing that reduces the impact of product variations.

According to Priarone et al. (2016), product characteristics and process parameters influence both power consumption and process duration, creating a trade-off between quality, efficiency, and sustainability. However, our findings suggest that process duration is less affected by product feature variation. This indicates that manufacturers may achieve greater efficiency by prioritising power consumption over balancing it with duration. Based on causal estimation results, this strategy may yield a slight increase in process time but lead to substantial energy savings. In energy-intensive sectors like stainless steel manufacturing, this approach provides significant economic and environmental benefits, offering a strategic advantage.

The consistency between our findings and existing literature on factors such as weight, outer diameter, and heat treatment (Mbanyeude 2023; Redelbach, Klötzke, and Friedrich 2012; Tangthieng 2011) strengthens the validity of our conclusions, while our station-specific analysis provides new insights into how these relationships vary across different manufacturing environments. These findings suggest that future optimisation strategies should consider workplace-specific variations rather than applying uniform approaches across all processing stations.

7.3. Advantages of causal AI compared to traditional machine learning

The causal AI-based approach applied in this study demonstrates several key advantages over traditional machine learning models, particularly in terms of transparency, interpretability, and robustness in decision-making. In traditional machine learning methods, such as linear regression, random forests, or neural networks, models tend to focus on correlations to predict outcomes without explicitly revealing how product features interact with process variables. This can be problematic in the manufacturing domain, where understanding the causal effect of product characteristics is crucial for optimising processes.

- *Transparency and Interpretability:* The causal AI models used in this study, such as DirectLiNGAM and RESIT, provide interpretable causal graphs that allow manufacturers to directly observe how product features, such as outer diameter and weight, influence power consumption and process duration. For example, outer diameter consistently showed a significant positive causal effect on power consumption, particularly in workstation 50516, with RESIT estimating a strong effect of 0.42. This level of interpretability is not achievable with traditional machine learning models, which treat feature importance as a byproduct of prediction accuracy rather than direct causal influence.
- *Overcoming Black-Box Limitations:* Traditional models like random forests and neural networks often function as black boxes, giving little insight into why certain features are deemed important. These models would likely flag outer diameter and weight as significant for power consumption but would not provide insights into their causal relationships. In contrast, the causal AI approach identified outer diameter as a direct cause of increased power consumption across multiple workstations. Similarly, it was shown that heat treatment categories have varying effects on process duration depending on the workstation, which traditional models might miss due to their focus on correlation rather than causality.
- *Reliability of Decision-Making:* The robustness of the causal models was verified through rigorous refutation tests. Both DirectLiNGAM and RESIT passed all refutation tests with p-values greater than 0.05, confirming that the identified causal relationships hold across different test conditions. This stands in contrast to traditional models, which might exhibit instability when subjected to small data perturbations. The stability of causal models ensures manufacturers can trust the identified causal drivers, such as the outer diameter's effect on power consumption, for long-term decision-making without concerns about model reliability.
- *Quantitative and Qualitative Improvements:* When comparing the causal AI models with traditional machine learning methods, the case study revealed several quantitative improvements. For instance, the causal models provided specific numerical estimates of how changes in product characteristics, like a 1 mm increase in outer diameter, would lead to a 0.42 unit increase in power consumption at workstation 50516.

Traditional models would likely identify outer diameter as a key feature but without quantifying this direct effect. Furthermore, qualitative improvements were observed, such as understanding why heat treatment had a positive effect on process duration in one workstation but a negative effect in another. This level of nuance and precision is difficult to achieve with non-causal models.

In conclusion, while traditional machine learning methods can offer high prediction accuracy, they lack the transparency and causal insights required for optimising manufacturing processes. The causal AI-based methodology used in this study not only identified critical factors like outer diameter and heat treatment categories but also quantified their causal effects on power consumption and process duration. This provides manufacturers with actionable insights that go beyond mere correlations, enabling more informed, reliable, and precise optimizations for reducing costs and improving process efficiency.

7.4. Robustness and scalability of the causal AI framework

The robustness of the proposed framework was demonstrated through multiple evaluation layers, including graph comparison with expert-validated ground truth, causal effect refutation tests (Section 6.5.3), and metric-based performance across multiple workstations. Specifically, the consistent superiority of DirectLiNGAM in terms of F1 score and SHD across three different stations (Table 6) shows its ability to capture meaningful causal structures under diverse conditions. The causal estimates remain stable across data perturbations and confounding checksFurthermore, robustness was confirmed through causal refutation tests (Table 11), where causal effect estimates remained stable under placebo, random confounding, and data subset scenarios.

Scalability was evaluated by replicating the methodology across multiple workstations and two datasets (power consumption and process duration). The modular design of the pipeline, which covers data preprocessing, causal discovery, graph refinement, estimation, and refutation, enables easy extension to additional workstations or even other domains. As the algorithms used (DirectLiNGAM, RESIT, DoWhy) operate efficiently on moderate-dimensional data, the approach can be scaled horizontally to production lines with hundreds of sensors or vertically to richer time-series data.

Furthermore, the consistent causal patterns across independent workstations (e.g. outer diameter or weight as main drivers of power consumption) demonstrate that the causal signals are not overfit to a single environment. This repeatability indicates external validity and highlights the potential of the framework to be generalised to other industrial settings, including energy-intensive sectors such as automotive, chemicals, and logistics. Scalability can also be achieved through integration with real-time data collection systems, enabling continuous model updates and adaptive causal inference.

7.5. Practical implementations, implications, and industrial applications

This section addresses **RQ2**, demonstrating how the causal insights gained from the analysis can inform industrial decision-making and enable interpretable, data-driven process optimisation.

The findings of this study demonstrate the potential of causal AI to uncover actionable insights for optimising energy consumption and process durations in the stainless-steel manufacturing industry. By identifying and quantifying key cause-and-effect relationships, causal AI enables targeted interventions that improve resource efficiency, reduce costs, and enhance sustainability. These insights provide decision-makers with a structured approach to optimising production processes, improving operational efficiency, and making informed investments in process improvements.

In the context of stainless-steel manufacturing, the analysis revealed specific causal relationships that can directly inform process improvements and guide industrial decision-making:

- Optimizing heat treatment parameters: The study identified that variations in product weight and height significantly influence the energy consumption of heat treatment processes. Adjusting these parameters for specific product profiles can reduce energy use. Managers can use this insight to develop energy-efficient heat treatment strategies by matching treatment intensity with product characteristics, reducing unnecessary power consumption.
- Reducing processing times: Larger outer diameters and heavier products were found to causally contribute to longer process durations. By redesigning workflows or adjusting equipment settings, manufacturers can reduce average processing times, thereby enhancing throughput. Production planners can apply these findings to optimise scheduling and allocate machine workloads more effectively to prevent bottlenecks.
- Key variables for energy efficiency: The causal analysis pinpointed that outer diameter and heat treatment parameters have the most substantial impact

on energy consumption. Focusing on optimising these variables offers a pathway to significant energy savings. Understanding these causal effects enables industrial leaders to prioritise investments in process automation and energy-efficient machinery, improving cost-effectiveness and sustainability.

From a managerial perspective, the results support proactive process optimisation. Managers can refine scheduling and heat treatment strategies based on causal impacts, while engineers can integrate insights into simulations to redesign stages that increase energy use or delay production. These causal relationships also offer a foundation for developing predictive maintenance plans and dynamic KPI dashboards aligned with production realities. Furthermore, decision-makers can incorporate these insights into training programs, enabling operational staff to better understand and control key drivers of performance.

Beyond the stainless-steel industry, this approach can be applied to other energy-intensive sectors, such as automotive manufacturing or chemical processing, to uncover inefficiencies and optimise production workflows based on industry-specific causal relationships. These results highlight the potential of causal AI to provide manufacturers with interpretable and actionable insights, enabling sustainable and efficient improvements in production processes. Integrating causal AI into decision-support systems allows companies to systematically identify improvement opportunities, enhance operational flexibility, and support long-term strategic planning.

7.6. Limitations and future work

While the proposed framework offers valuable insights into causal relationships in manufacturing processes, several limitations should be acknowledged. First, the data granularity is constrained by the temporal resolution of sensor readings and ERP timestamps, which may obscure short-term dynamics or overlapping operations. In particular, energy data are aggregated at the machine level without fine-grained breakdowns for subprocesses, potentially limiting the precision of causal attribution. Additionally, the analysis was conducted on a limited number of datasets from specific manufacturing contexts. Expanding these methods to diverse datasets across different industries, including logistics and supply chain management, will enhance generalizability and scalability while identifying industry-specific challenges in real-world deployment.

Second, the causal discovery algorithms employed, DirectLiNGAM and RESIT, operate under specific assumptions. DirectLiNGAM assumes linearity and non-Gaussian noise, while RESIT requires additive noise and functional relationships. If these assumptions are violated in practice, the resulting graphs may omit or misrepresent certain dependencies. Moreover, both methods assume causal sufficiency (i.e. no hidden confounders), which may not always hold in real-world production settings. Furthermore, DirectLiNGAM and RESIT have limitations in handling complex causal structures, particularly in the presence of unobserved confounders and dynamic system changes. Unobserved confounders remain a significant challenge, potentially introducing bias into causal inferences. Techniques such as sensitivity analysis, instrumental variables, and domain-informed causal priors should be further developed to improve robustness and practical applicability in industrial settings.

Third, real-world implementation poses practical challenges. Translating causal insights into actionable process changes often requires cross-functional coordination and changes in scheduling, machine usage, or material flow. Additionally, domain experts may be hesitant to adopt algorithm-driven interventions without clear interpretability or process validation. Ensuring integration with existing IT infrastructure (e.g. ERP/MES systems) and addressing organisational inertia remain non-trivial hurdles.

To address these limitations, future research should focus on:

- Increasing data granularity through real-time IoT-based sensing.
- Advancing causal discovery algorithms to better capture complex, dynamic relationships in high-dimensional datasets.
- Employing multiple identification strategies (e.g. backdoor, frontdoor, instrumental variables) to strengthen causal inferences and mitigate unobserved confounders.
- Testing robustness under assumption violations (e.g. through sensitivity or latent confounder analyses).
- Developing human-in-the-loop systems to ease adoption and decision-making in industrial environments.
- Developing automated decision-support tools that facilitate the integration of causal AI into manufacturing workflows, making it accessible for industrial practitioners.
- Extending applications to broader industrial contexts, such as Industry 4.0, energy management, and

sustainable manufacturing, to validate scalability and long-term impact.

By addressing these limitations and pursuing the outlined directions, future work can enhance the adoption of interpretable, causality-driven approaches that support data-driven decision-making in manufacturing and beyond.

8. Conclusion and outlook

The integration of causal AI into manufacturing analytics represents a significant advancement in data-driven decision-making. This paper contributes to the state of the art by presenting a comprehensive and validated framework for applying causal discovery and inference to production data in energy-intensive, customised manufacturing environments. Specifically, we advance the field by (i) integrating multiple causal estimation strategies (backdoor, frontdoor, instrumental variables), (ii) comparing two causal discovery algorithms (DirectLiNGAM, RESIT) in a multi-workstation setup, and (iii) validating causal insights through domain knowledge and refutation tests.

Our framework revealed how product-specific characteristics, such as weight, height, and outer diameter, causally affect power consumption and process duration. Although individual causal effects were often modest, their cumulative impact across large-scale operations is significant. The results underscore the value of causal reasoning in uncovering nuanced, actionable patterns that traditional correlation-based methods may overlook.

Future research should build on this foundation to further strengthen the operational relevance and technical scalability of causal AI in industrial contexts. One promising direction is the integration of causal models with digital twins, enabling real-time monitoring, simulation, and adaptive process control based on causal feedback loops. This integration can facilitate proactive decision-making and dynamic reconfiguration of production systems in response to changing conditions. Another key avenue is the adoption of temporal causal models, such as Dynamic Bayesian Networks, Granger causality, and time-aware Structural Equation Models (SEMs), to capture sequential dependencies, delayed effects, and feedback loops inherent in manufacturing processes. These models can enhance the analysis of batch production, maintenance cycles, or energy consumption patterns over time. Moreover, multi-site validation is essential to assess the generalizability and robustness of the proposed approach. Applying the framework across different factories, production lines, or industry sectors (e.g. automotive, food processing, or electronics) would uncover commonalities and domain-specific differences, thus informing the development of transfer learning or meta-causal modelling strategies for broader applicability. In parallel, future work should explore hybrid approaches combining causal discovery with domain ontologies or physics-informed models to improve interpretability and reduce reliance on data volume alone. Incorporating expert feedback directly into the causal learning loop can also enhance trustworthiness and compliance with safety-critical constraints in high-stakes environments.

To operationalise causal AI at scale, several prerequisites must be addressed. Technically, robust data infrastructure is required to ensure high-frequency, synchronised, and clean data from heterogeneous sources. Automated feature engineering and causal graph generation tools will reduce the manual workload for analysts. Organizationally, cross-functional collaboration between data scientists, process engineers, and domain experts is critical to ensure that causal insights are actionable and correctly interpreted. Moreover, change management and training programs are necessary to build trust in model-driven recommendations and facilitate adoption on the shop floor.

In conclusion, causal AI offers a powerful pathway toward sustainable and intelligent manufacturing. By making causality a first-class citizen in analytics pipelines, companies can move beyond correlation-based insights and toward root-cause-driven optimisation. The proposed framework provides a foundation for future deployments, with the potential to be extended across industries and scaled through integration with digital platforms, simulation tools, and enterprise systems.

Note

1. DoWhy, https://github.com/py-why/dowhy, accessed June 4, 2024

Disclosure statement

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

AI disclosure

The authors acknowledge the use of **ChatGPT (version 4)** by OpenAI to assist in improving the clarity and readability of the manuscript's English language. All AI-generated suggestions were carefully reviewed and integrated by the authors to ensure accuracy and coherence.

Notes on contributors

Tamas Fekete received his Ph.D. in Information Systems and Management from Constructor University, Bremen, Germany, where he is a member of the Research Group 'Data-Driven Industrial Systems'. He also works as a Senior Software Engineer at evosoft Hungary Kft., a Siemens-owned company. His research focuses on causal AI and knowledge-graph-based methods for uncovering cause-and-effect relationships in industrial systems to support sustainability, resource efficiency, and the reduction of environmental impact.

Isabella Marquez Petrone earned her Bachelor's degree in Industrial Engineering and Management from Constructor University. During her studies, she worked as a teaching and research assistant in the Research Group 'Data-Driven Industrial Systems'. She was also honoured with the graduate award for outstanding achievement in her study program. Currently, she works as a logistics manager at a fashion company in Hamburg, Germany.

Hendro Wicaksono is a Full Professor and Head of Research Group 'Data-Driven Industrial Systems' at Constructor University, Bremen, Germany. He received his bachelor's degree in informatics from Institut Teknologi Bandung, Indonesia, and an M.Sc. degree in information and communication engineering from the University of Karlsruhe, Germany. After gaining industry experience, he completed his doctoral degree in mechanical engineering at the Karlsruhe Institute of Technology. He worked as a researcher at the FZI Research Center for Information Technologies and the Institute of Information Management in Engineering (IMI) at the Karlsruhe Institute of Technology. His research interests include causal AI, explainable AI, semantic data integration, ontologies, and their applications in complex industrial contexts.

Data availability statement

The datasets analysed during the current study are available upon reasonable request.

References

Abdulla, A., G. Baryannis, and I. Badi. 2019. "Weighting the Key Features Affecting Supplier Selection Using Machine Learning Techniques." Preprints.

Aikenov, T., R. Hidayat, and H. Wicaksono. 2024. "Power Consumption and Process Cost Prediction of Customized Products Using Explainable Ai: A Case in the Steel Industry." In *Flexible Automation and Intelligent Manufacturing: Establishing Bridges for More Sustainable Manufacturing Systems*, edited by F. J. G. Silva, L. P. Ferreira, J. C. Sá, M. T. Pereira, and C. M. A. Pinto, 1183–1193. Cham: Springer Nature Switzerland.

Aljuhani, A., A. Alhubaishy, M. K. I. Rahmani, and A. A. Alzahrani. 2023. "Light-Weighted Decision Support Framework for Selecting Cloud Service Providers." *Computers, Materials & Continua* 74 (2): 4293–4317. https://doi.org/10.32604/cmc.2023.033893.

American Society of Mechanical Engineers. 2005. "Process Piping: ASME Code for Pressure Piping, B31."

Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455. https://doi.org/10.1080/01621459.1996.10476902.

Antonaci, F., P. Chimento, H.-C. Diener, G. Sances, and G. Bono. 2007. "Lessons from Placebo Effects in Migraine Treatment." *The Journal of Headache and Pain* 8:63–66. https://doi.org/10.1007/s10194-007-0360-4.

Azodi, C. B., J. Tang, and S.-H. Shiu. 2020. "Opening the Black Box: Interpretable Machine Learning for Geneticists." *Trends in Genetics* 36 (6): 442–455. https://doi.org/10.1016/j.tig.2020.03.005.

Bampoula, X., N. Nikolakis, and K. Alexopoulos. 2024. "Condition Monitoring and Predictive Maintenance of Assets in Manufacturing Using Lstm-Autoencoders and Transformer Encoders." *Sensors*24 (10): 3215. https://doi.org/10.3390/s24103215.

Bellemare, M. F., J. R. Bloem, and N. Wexler. 2024. "The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion." *Oxford Bulletin of Economics and Statistics* 86 (4): 951–993. https://doi.org/10.1111/obes.v86.4.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bicand Fernández, D. 2023. "Causal Modeling in Manufacturing: Analysis and Applications."

Brynjolfsson, E., and K. McElheran. 2016. "The Rapid Adoption of Data-Driven Decision-Making." *American Economic Review* 106 (5): 133–139. https://doi.org/10.1257/aer.p20161016.

Caggiano, A., J. Zhang, V. Alfieri, F. Caiazzo, R. Gao, and R. Teti. 2019. "Machine Learning-Based Image Processing for on-line Defect Recognition in Additive Manufacturing." *CIRP Annals* 68 (1): 451–454. https://doi.org/10.1016/j.cirp.2019.03.021.

Chen, T., V. Sampath, M. Carl May, S. Shan, O. J. Jorg, J. J. A. Martín, F. Stamer, G. Fantoni, G. Tosello, and M. Calaon. 2023. "Machine Learning in Manufacturing towards Industry 4.0: From 'for Now' to 'Four-Know'." *Applied Science* 13 (3): 1903. https://doi.org/10.3390/app13031903.

Conejo, A. N., J.-P. Birat, and A. Dutta. 2020. "A Review of the Current Environmental Challenges of the Steel Industry and Its Value Chain." *Journal of Environmental Management* 259:109782. https://doi.org/10.1016/j.jenvman.2019.109782.

Crown, W. 2019. "Real-World Evidence, Causal Inference, and Machine Learning." *Value in Health* 22 (5): 587–592. https://doi.org/10.1016/j.jval.2019.03.001.

Cui, C., Y. Liu, B. Xia, X. Jiang, and M. Skitmore. 2020. "Overview of Public-Private Partnerships in the Waste-to-energy Incineration Industry in China: Status, Opportunities, and Challenges." *Energy Strategy Reviews* 32:100584. https://doi.org/10.1016/j.esr.2020.100584.

Cui, Y., H. Pu, X. Shi, W. Miao, and E. Tchetgen Tchetgen. 2023. "Semiparametric Proximal Causal Inference." *Journal of the American Statistical Association* 119:1348–1359. https://doi.org/10.1080/01621459.2023.2191817.

Damiani, L., P. Giribone, R. Revetria, and A. Testa. 2014. "An Innovative Model for Supporting Energy-Based Cost Reduction in Steel Manufacturing Industry Using Online Real-Time Simulation." https://api.semanticscholar.org/CorpusID:209455069.

Dieter, G. E.. 1997. "Relationship between Materials Selection and Processing." In *Engineering Design: A Materials and Processing Approach*. 3rd ed., edited by G. E. Dieter . McGraw-Hill.

Fatla, O. M. H., F. C. J. Robinson, M. J. Jweeg, N. Beynon, A. Valera-Medina, H. S. S. Aljibori, M. N. Mohammed, and O. I. Abdullah. 2024. "Technologies for High-Temperature Batch Annealing of Grain-Oriented Electrical Steel: An Overview." *Open Engineering* 14 (1):20240053. https://doi.org/10.1515/eng-2024-0053.

Gajdzik, B., R. Wolniak, and W. W. Grebski. 2023. "Electricity and Heat Demand in Steel Industry Technological Processes in Industry 4.0 Conditions." *Energies* 16 (2): 787. https://doi.org/10.3390/en16020787.

Gao, T., H. Zhang, Y. Liu, and Z. Wang. 2017. "The Influence of Length–diameter Ratio in Forming Area on Viscous Outer Pressure Forming and Limit Diameter Reduction." *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 39:481–486. https://doi.org/10.1007/s40430-015-0340-6.

Gardas, R., and S. Narwane. 2024. "An Analysis of Critical Factors for Adopting Machine Learning in Manufacturing Supply Chains." *Decision Analytics Journal* 10:100377. https://doi.org/10.1016/j.dajour.2023.100377.

Glymour, C., K. Zhang, and P. Spirtes. 2019. "Review of Causal Discovery Methods Based on Graphical Models." *Frontiers in Genetics* 10:524. https://doi.org/10.3389/fgene.2019.00524.

Griffiths, C., J. Howarth, G. De Almeida-Rowbotham, A. Rees, and R. Kerton. 2016. "A Design of Experiments Approach for the Optimisation of Energy and Waste during the Production of Parts Manufactured by 3d Printing." *Journal of Cleaner Production* 139:74–85. https://doi.org/10.1016/j.jclepro.2016.07.182.

Guerra-Zubiaga, D. A., A. Al Mamun, and G. Gonzalez-Badillo. 2018. "An Energy Consumption Approach in a Manufacturing Process Using Design of Experiments." *International Journal of Computer Integrated Manufacturing* 31:1067–1077. https://doi.org/10.1080/0951192X.2018.1493234.

Hagedorn, C., J. Huegle, and R. Schlosser. 2022. "Understanding Unforeseen Production Downtimes in Manufacturing Processes Using Log Data-Driven Causal Reasoning." *Journal of Intelligent Manufacturing* 33 (7): 2027–2043. https://doi.org/10.1007/s10845-022-01952-x.

Han, C., and Q. Zhang. 2021. "Optimization of Supply Chain Efficiency Management Based on Machine Learning and Neural Network." *Neural Computing and Applications* 33 (5): 1419–1433. https://doi.org/10.1007/s00521-020-05023-1.

Hartono, M., G. Gumono, W. Wirawan, N. Qosim, and H. Wicaksono. 2021. "Quality by Design of Optimum Parameter to Minimize the Weight of Plastic Products." In *IOP Conference Series: Materials Science and Engineering* (Online), Vol. 1073, 012076. IOP Publishing.

Hashmi, S. M. A., T. Fekete, and H. Wicaksono. 2024. "Causal Ai in the Automotive Industry: Impact Analysis through Carbon Emission Case Study." In *2024 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* Bangkok, Thailand (15-18 Dec 2024), 1267–1271. IEEE.

Hatt, T., and S. Feuerriegel. 2024. "Sequential Deconfounding for Causal Inference with Unobserved Confounders." In *Causal Learning and Reasoning* Los Angeles, California, USA (UCLA, Carnesale Commons) on 1–3 Apr 2024, 934–956. PMLR.

He, M., and M. Khorsand. 2024. "Behavior-Aware Assessment of Demand Response Participation in Power Systems Using Causal Artificial Intelligence." *Electric Power Systems Research* 237:110981. https://doi.org/10.1016/j.epsr.2024.110981.

Heiss, A. 2024. "Demystifying Causal Inference Estimands: ATE, ATT, and ATU." Accessed May 27, 2024. https://www.andrewheiss.com/blog/2024/03/21/demystifying-ate-att-atu/.

Hidayat, R., A. Ourairat, and H. Wicaksono. 2024. "Explainable Artificial Intelligence in Agrifood Supply Chain: State of the Art Review." In *International Conference on Flexible Automation and Intelligent Manufacturing*, 291–299. Taichung, Taiwan: Springer.

Holzinger, A., M. Plass, K. Holzinger, G. C. Crisan, C.-M. Pintea, and V. Palade. 2017. "A Glass-Box Interactive Machine Learning Approach for Solving NP-hard Problems with the Human-in-the-Loop." Preprint arXiv:1708.01104.

Hoyer, P. O., D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. 2009. "Nonlinear Causal Discovery with Additive Noise Models." In *Advances in Neural Information Processing Systems*, Vancouver, Canada, Vol. 21, 689–696, Curran Associates, Inc.

International Energy Agency. 2023. "World Energy Outlook 2023." Accessed October 24, 2024.

Jeong, H., T. Fekete, A. Bashyal, and H. Wicaksono. 2025. "From Theory to Practice: Implementing Causal Ai in Manufacturing for Sustainability." *Procedia Computer Science* 253:1495–1504. https://doi.org/10.1016/j.procs.2025.01.211.

Johnson, W., T. Norris, D. Bann, N. Cameron, J. K. Wells, T. J. Cole, and R. Hardy. 2020. "Differences in the Relationship of Weight to Height, and Thus the Meaning of Bmi, according to Age, Sex, and Birth Year Cohort." *Annals of Human Biology* 47 (2): 199–207. https://doi.org/10.1080/03014460.2020.1737731.

Kosasih, E. E., and A. Brintrup. 2022. "A Machine Learning Approach for Predicting Hidden Links in Supply Chain with Graph Neural Networks." *International Journal of Production Research* 60 (17): 5380–5393. https://doi.org/10.1080/00207543.2021.1956697.

Lim, K. Y. H., T. S. Yosal, C.-H. Chen, P. Zheng, L. Wang, and X. Xu. 2024. "Graph-Enabled Cognitive Digital Twins for Causal Inference in Maintenance Processes." *International Journal of Production Research* 62 (13): 4717–4734.https://doi.org/10.1080/00207543.2023.2274335.

Lin, H., J. Lin, and F. Wang. 2022. "An Innovative Machine Learning Model for Supply Chain Management." *Journal of Innovation & Knowledge* 7 (4):100276. https://doi.org/10.1016/j.jik.2022.100276.

Maathuis, M. H., and D. Colombo. 2015. "A Generalized Back-Door Criterion." *Annals of Statistics* 43 (3): 1060–1088. https://doi.org/10.1214/14-AOS1295.

Makhlouf, K., S. Zhioua, and C. Palamidessi. 2022. "Survey on Causal-Based Machine Learning Fairness Notions." Accessed May 27, 2024.

Marazopoulou, K., R. Ghosh, P. Lade, and D. D. Jensen. 2016. "Causal Discovery for Manufacturing Domains." ArXiv abs/1605.04056.

Mbanyeude, C. H. 2023. "Heat Treatment Energy Mapping." Master's Thesis, Faculty of Textiles, Engineering and Business.

Mechai, N., and H. Wicaksono. 2024. "Causal Inference in Supply Chain Management: How Does Ever Given Accident at the Suez Canal Affect the Prices of Shipping Containers?" *Procedia Computer Science*232:3173–3182. https://doi.org/10.1016/j.procs.2024.02.133.

Menegozzo, G. 2022. Discovering Phase and Causal Dependencies on Manufacturing Processes. Doctoral thesis, Department of Computer Science, University of Verona, Verona, Italy.

Moon, S.-K., and J.-O. Kim. 2017. "Balanced Charging Strategies for Electric Vehicles on Power Systems." *Applied Energy* 189:44–54. https://doi.org/10.1016/j.apenergy.2016.12.025.

Mousa, W., F. M. Ghazali, S. Jinap, H. M. Ghazali, S. Radu, and A. E.-R. Salama. 2016. "Temperature, Water Activity and Gas Composition Effects on the Growth and Aflatoxin Production by Aspergillus Flavus on Paddy." *Journal of Stored Products Research* 67:49–55. https://doi.org/10.1016/j.jspr.2016.01.003.

Muralitharan, S., W. Nelson, S. Di, M. McGillion, P. Devereaux, N. G. Barr, and J. Petch. 2021. "Machine Learning–based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review." *Journal of Medical Internet Research* 23 (2): e25187. https://doi.org/10.2196/25187.

Nesro, V. M., T. Fekete, and H. Wicaksono. 2024. "Leveraging Causal Machine Learning for Sustainable Automotive Industry: Analyzing Factors Influencing CO_2 Emissions." *Procedia CIRP* 130:161–166. https://doi.org/10.1016/j.procir.2024.10.071.

Nogueira, A. R., A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama. 2022. "Methods and Tools for Causal Discovery and Causal Inference." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 12 (2): e1449.

Oliveira, E. E., V. L. Miguéis, and J. L. Borges. 2021. "Understanding Overlap in Automatic Root Cause Analysis in Manufacturing Using Causal Inference." *IEEE Access* 10:191–201. https://doi.org/10.1109/ACCESS.2021.3139199.

Pearce, N., and D. A. Lawlor. 2016. "Causal Inference–So Much More than Statistics." *International Journal of Epidemiology* 45 (6): 1895–1903. https://doi.org/10.1093/ije/dyw328.

Pearl, J. 2009. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3:96–146. https://doi.org/10.1214/09-SS057.

Pearl, J., and A. Paz. 2014. "Confounding Equivalence in Causal Inference." *Journal of Causal Inference* 2 (1): 75–93.

Peres, R. S., J. Barata, P. Leitao, and G. Garcia. 2019. "Multistage Quality Control Using Machine Learning in the Automotive Industry." *IEEE Access* 7:79908–79916. https://doi.org/10.1109/Access.6287639.

Peters, J., and P. Bühlmann. 2015. "Structural Intervention Distance (sid) for Evaluating Causal Graphs." *Neural Computation* 27 (3): 771–799. https://doi.org/10.1162/NECO_a_00708.

Peters, J., J. M. Mooij, D. Janzing, and B. Schölkopf. 2014. "Causal Discovery with Continuous Additive Noise Models." *Journal of Machine Learning Research* 15 (1): 2009–2053.

Priarone, P. C., F. Klocke, M. G. Faga, D. Lung, and L. Settineri. 2016. "Tool Life and Surface Integrity When Turning Titanium Aluminides with Pcd Tools under Conventional Wet Cutting and Cryogenic Cooling." *The International Journal of Advanced Manufacturing Technology* 85:807–816. https://doi.org/10.1007/s00170-015-7958-5.

Redelbach, M., M. Klötzke, and H. E. Friedrich. 2012. "Impact of Lightweight Design on Energy Consumption and Cost Effectiveness of Alternative Powertrain Concepts." In *European Electric Vehicle Conference*, Brussels, Belgium.

Rocha, A., J. Papa, and L. Meira. 2012. "How Far Do We Get Using Machine Learning Black-Boxes?" *International Journal of Pattern Recognition and Artificial Intelligence* 26 (2):1261001. https://doi.org/10.1142/S0218001412610010.

Roozkhosh, P., A. Pooya, and R. Agarwal. 2022. "Blockchain Acceptance Rate Prediction in the Resilient Supply Chain with Hybrid System Dynamics and Machine Learning Approach." *Operations Management Research* 16 (2): 705–725. https://doi.org/10.1007/s12063-022-00336-x.

Rudin, C. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1:206–215. https://doi.org/10.1038/s42256-019-0048-x.

Runge, J., A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls. 2023. "Causal Inference for Time Series." *Nature Reviews Earth & Environment*4 (7): 487–505. https://doi.org/10.1038/s43017-023-00431-y.

Shah, M. A., and H. Wicaksono. 2024. "Leveraging Machine Learning for Power Consumption Prediction of Multi-step Production Processes in Dynamic Electricity Price Environment." *Procedia CIRP*130:226–231. https://doi.org/10.1016/j.procir.2024.10.080.

Shahbazinia, A., S. Salehkaleybar, and M. Hashemi. 2023. "Paralingam: Parallel Causal Structure Learning for Linear Non-gaussian Acyclic Models." *Journal of Parallel and Distributed Computing*176:114–127. https://doi.org/10.1016/j.jpdc.2023.01.007.

Shang, C., and F. You. 2019. "Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era." *Engineering* 5 (6): 1010–1016. https://doi.org/10.1016/j.eng.2019.01.019.

Sharma, A., and E. Kiciman. 2020. "Dowhy: An End-to-end Library for Causal Inference." Preprint arXiv:2011.04216.

Sharma, A., Z. Zhang, and R. Rai. 2021. "The Interpretive Model of Manufacturing: A Theoretical Framework and Research Agenda for Machine Learning in Manufacturing." *International Journal of Production Research* 59 (16): 4960–4994. https://doi.org/10.1080/00207543.2021.1930234.

Sharma, K., Y. K. Dwivedi, and B. Metri. 2024. "Incorporating Causality in Energy Consumption Forecasting Using Deep Neural Networks." *Annals of Operations Research* 339 (1): 537–572. https://doi.org/10.1007/s10479-022-04857-3.

Shimizu, S., T. Inazumi, Y. Sogawa, A. Hyvarinen, Y. Kawahara, T. Washio, P. O. Hoyer, K. Bollen, and P. Hoyer. 2011. "Directlingam: A Direct Method for Learning a Linear Non-gaussian Structural Equation Model." *Journal of Machine Learning Research-JMLR* 12:1225–1248.

Shimizu, S., and S. Kawano. 2022. "Special Issue: Recent Developments in Causal Inference and Machine Learning." *Behaviormetrika* 49:275–276. https://doi.org/10.1007/s41237-022-00173-z.

Singh, R. 2012. "Chapter 12 – Heat Treatment of Steels." In *Applied Welding Engineering*, edited by R. Singh, 95–108. Boston: Butterworth-Heinemann.

Srivastava, P. R., S. K. Mangla, P. Eachempati, and A. K. Tiwari. 2023. "An Explainable Artificial Intelligence Approach to Understanding Drivers of Economic Energy Consumption and Sustainability." *Energy Economics* 125:106868. https://doi.org/10.1016/j.eneco.2023.106868.

Sucar, L. E.. 2021. "Graphical Causal Models." In *Probabilistic Graphical Models: Principles and Applications*, 287–305. Cham, Switzerland: Springer Nature Switzerland AG.

Tangthieng, C. 2011. "Effect of Tube Diameter on the Specific Energy Consumption of the Ice Making Process." *Applied Thermal Engineering* 31 (5): 701–707. https://doi.org/10.1016/j.applthermaleng.2010.10.002.

Thapaliya, S., O. F. Valilai, and H. Wicaksono. 2024. "Power Consumption and Processing Time Estimation of Cnc Machines Using Explainable Artificial Intelligence (xai)." *Procedia Computer Science*232:861–870. https://doi.org/10.1016/j.procs.2024.01.086.

Tiensuu, H., S. Tamminen, E. Puukko, and J. Röning. 2021. "Evidence-Based and Explainable Smart Decision Support for Quality Improvement in Stainless Steel Manufacturing." *Applied Sciences* 11 (22): 10897. https://doi.org/10.3390/app112210897.

Tsamardinos, I., L. E. Brown, and C. F. Aliferis. 2006. "The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm." *Machine Learning* 65 (1): 31–78. https://doi.org/10.1007/s10994-006-6889-7.

van Geloven, N., S. A. Swanson, C. L. Ramspek, K. Luijken, M. van Diepen, T. P. Morris, R. H. Groenwold, H. C. van Houwelingen, H. Putter, and S. le Cessie. 2020. "Prediction Meets Causal Inference: The Role of Treatment in Clinical Prediction Models." *European Journal of Epidemiology*35:619–630. https://doi.org/10.1007/s10654-020-00636-1.

Villegas, M. A., D. J. Pedregal, and J. R. Trapero. 2018. "A Support Vector Machine for Model Selection in Demand Forecasting Applications." *Computers & Industrial Engineering* 121:1–7. https://doi.org/10.1016/j.cie.2018.04.042.

Vowels, M. 2022. "Trying to Outrun Causality in Machine Learning: Limitations of Model Explainability Techniques for Identifying Predictive Variables." Accessed May 27, 2024. https://www.researchgate.net/publication/358761912.

Vuković, M., and S. Thalmann. 2022. "Causal Discovery in Manufacturing: A Structured Literature Review." *Journal of Manufacturing and Materials Processing* 6 (1): 10. https://doi.org/10.3390/jmmp6010010.

Weiqiang, L. 2001. "Integration Upsetting of Tube." In *Proceedings of the 4th International Conference on Tube Manufacturing*. London, UK: Institution of Mechanical Engineers.

Wicaksono, H., T. Belzner, and J. Ovtcharova. 2013. "Efficient Energy Performance Indicators for Different Level of Production Organizations in Manufacturing Companies." In *IFIP International Conference on Advances in Production Management Systems*, 249–256. State College, Pennsylvania, USA: Springer.

Wong, L.-W., G. W.-H. Tan, K.-B. Ooi, B. Lin, and Y. K. Dwivedi. 2022. "Artificial Intelligence-Driven Risk Management for Enhancing Supply Chain Agility: A Deep-Learning-based Dual-Stage PLS-SEM-ANN Analysis." *International Journal of Production Research* 62:5535–5555. https://doi.org/10.1080/00207543.2022.2063089.

Wuest, T., D. Weimer, C. Irgens, and K.-D. Thoben. 2016. "Machine Learning in Manufacturing: Advantages, Challenges, and Applications." *Production & Manufacturing Research* 4 (1): 23–45. https://doi.org/10.1080/21693277.2016.1192517.

Wyrembek, M., G. Baryannis, and A. Brintrup. 2025. "Causal Machine Learning for Supply Chain Risk Prediction and Intervention Planning." *International Journal of Production Research* 63 (15): 5629–5648. https://doi.org/10.1080/00207543.2025.2458121.

Xie, F., R. Cai, Y. Zeng, J. Gao, and Z. Hao. 2019. "An Efficient Entropy-Based Causal Discovery Method for Linear Structural Equation Models with Iid Noise Variables." *IEEE Transactions on Neural Networks and Learning Systems* 31 (5): 1667–1680. https://doi.org/10.1109/TNNLS.5962385.

Xu, Z., and Y. Dang. 2023. "Data-driven Causal Knowledge Graph Construction for Root Cause Analysis in Quality Problem Solving." *International Journal of Production Research* 61 (10): 3227–3245. https://doi.org/10.1080/00207543.2022.2078748.

Yadollahi, A., N. Shamsaei, S. M. Thompson, and D. W. Seely. 2015. "Effects of Process Time Interval and Heat Treatment on the Mechanical and Microstructural Properties of Direct Laser Deposited 316l Stainless Steel." *Materials Science and Engineering: A* 644:171–183. https://doi.org/10.1016/j.msea.2015.07.056.

Appendices

Appendix 1. Visualizations from exploratory data analysis (EDA)

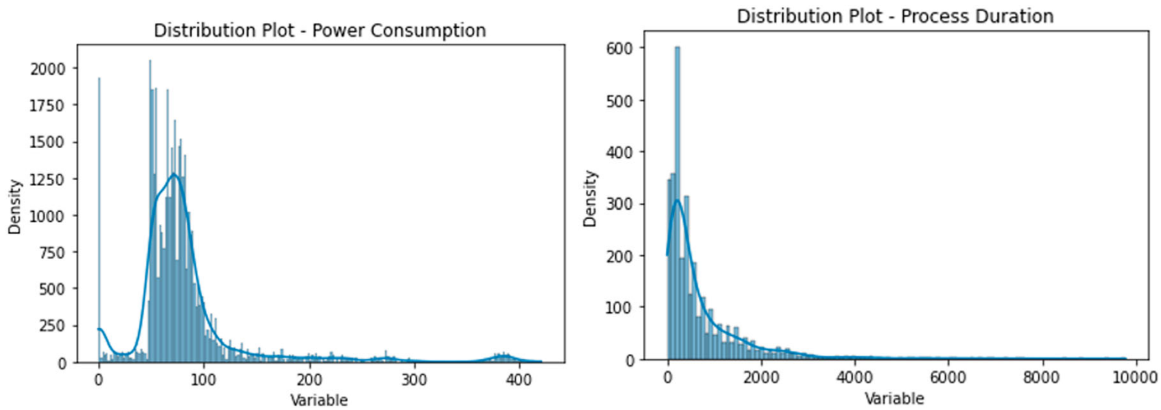
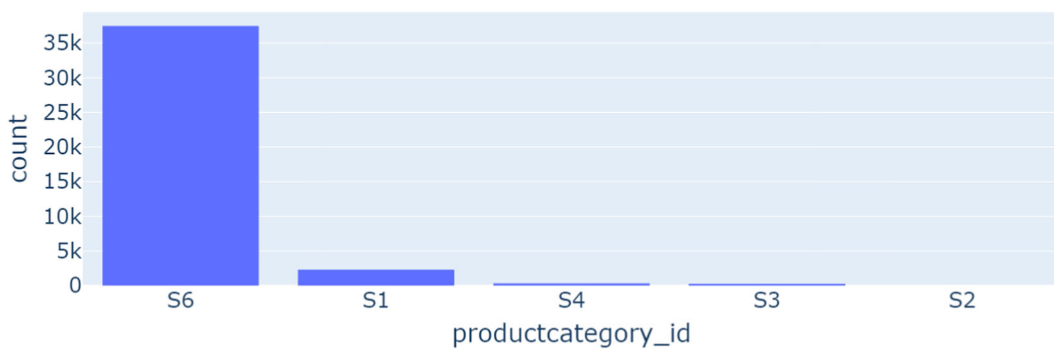
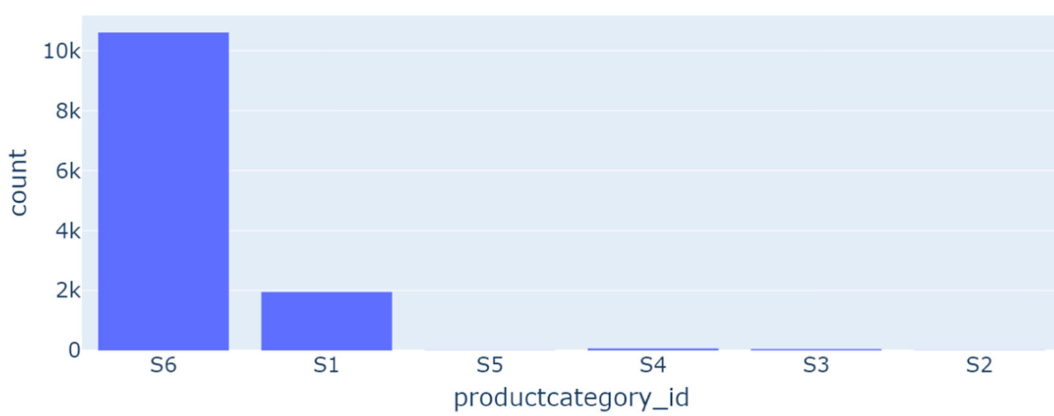


Figure A1. Distribution of power consumption and process duration for workstation 50513. The plot shows that both variables are right-skewed, with a concentration of values at lower ranges and a long tail of higher values. This skewness highlights variability and the presence of outliers, motivating further preprocessing and normalisation.



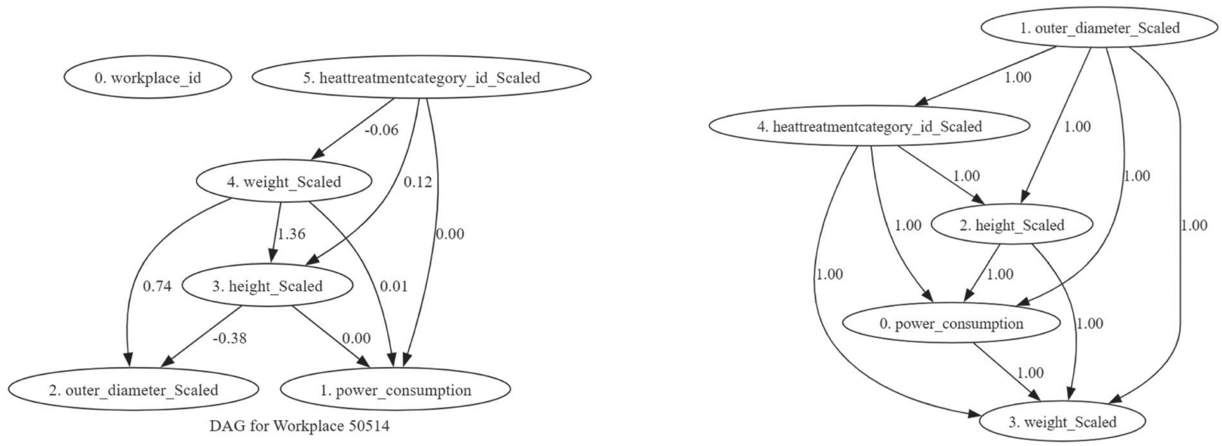
(a) Power consumption histogram.



(b) Process duration histogram.

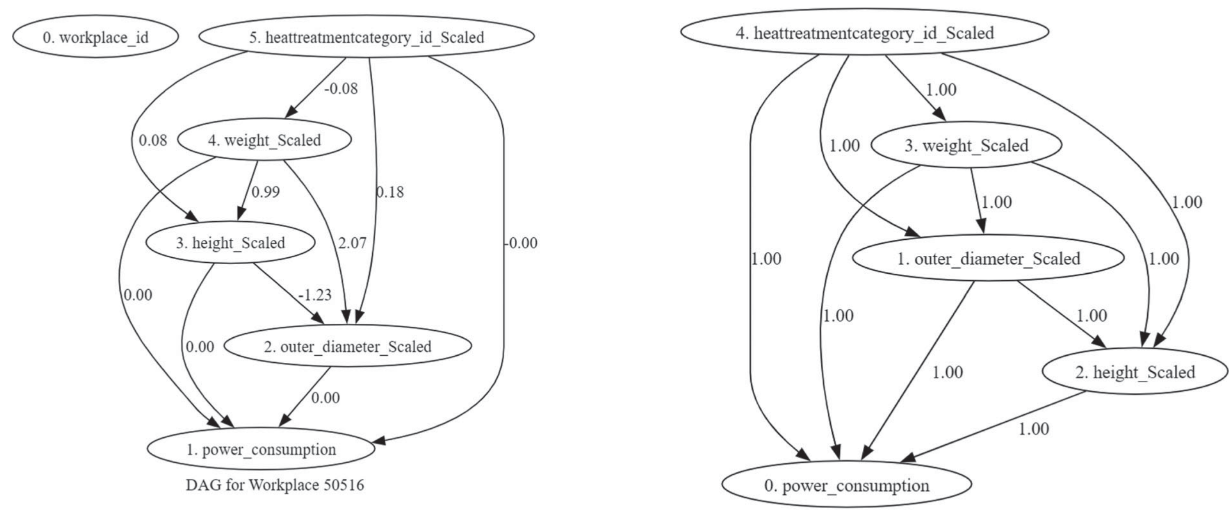
Figure A2. Histograms of power consumption and process duration for workstation 50513. These visualisations reveal frequency distributions of the two target variables used in causal analysis. Both show a majority of cases clustered at lower values, supporting the need to account for skew and outliers in further modelling steps. (a) Power consumption histogram. (b) Process duration histogram.

Appendix 2. Generated causal graphs for other workstations



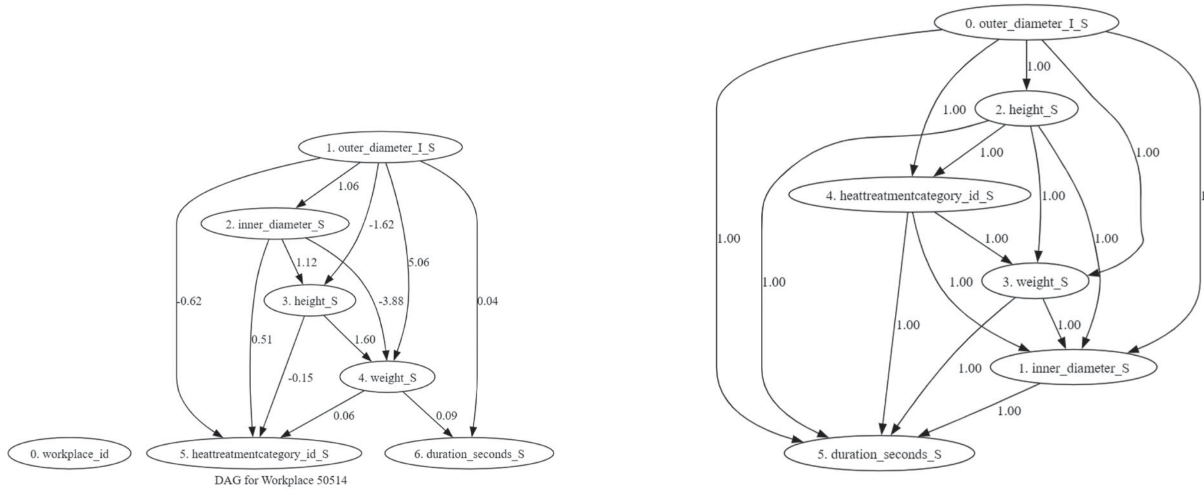
(a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.

Figure A3. Comparison of causal graphs generated by DirectLiNGAM and RESIT for workstation 50514 using the power consumption dataset. Each graph illustrates inferred causal relationships among production variables related to energy usage. (a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.



(a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.

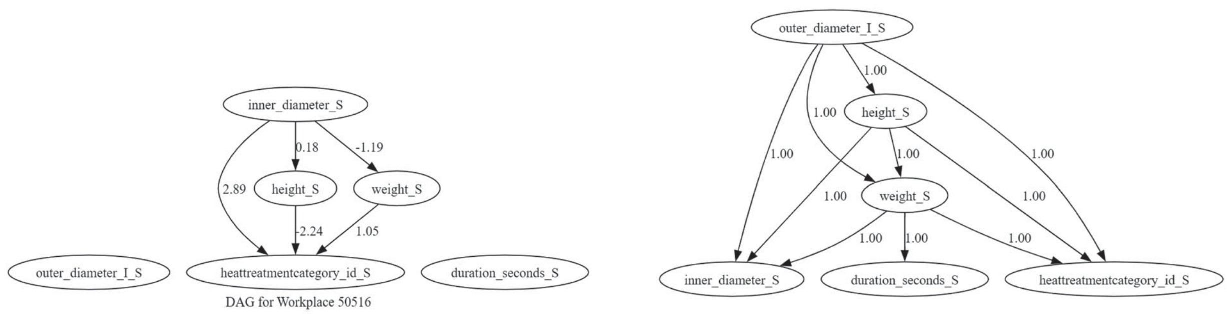
Figure A4. Comparison of causal graphs generated by DirectLiNGAM and RESIT for workstation 50516 using the power consumption dataset. These visualisations show the structure of inferred causal relationships between key production features. (a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.



(a) Causal graph generated by DirectLiNGAM.

(b) Causal graph generated by RESIT.

Figure A5. Comparison of causal graphs generated by DirectLiNGAM and RESIT for workstation 50514 using the process duration dataset. Each graph illustrates inferred causal relationships among production variables that influence process timing. (a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.



(a) Causal graph generated by DirectLiNGAM.

(b) Causal graph generated by RESIT.

Figure A6. Comparison of causal graphs generated by DirectLiNGAM and RESIT for workstation 50516 using the process duration dataset. These visualisations highlight differences in inferred causal structures related to the timing of production steps. (a) Causal graph generated by DirectLiNGAM. (b) Causal graph generated by RESIT.

Appendix 3. DoWhy outcomes for other generated graphs

Table A1. Estimated causal effects for power consumption at workplace ID 50514.

(a) Causal estimation, power consumption dataset:	Workplace ID 50514		
	DirectLiNGAM	RESIT	Manual model
Heat Treatment	2.63	2.93	−0.24
Weight	8.72	0	−1.85
Height	−0.72	−0.04	−11.35
Outer Diameter	0	1.51	0.71

Table A2. Estimated causal effects for process duration at workplace ID 50514.

(b) Causal estimation, process duration dataset:	Workplace ID 50514		
	DirectLiNGAM	RESIT	Manual model
Heat Treatment	0.07	0.01	−0.23
Weight	0.09	0.11	0.08
Height	−0.09	0.01	−0.05
Outer Diameter	0.004	0.36	−0.03
Inner Diameter	−0.44	0.06	0.03

Table A3. Estimated causal effects for power consumption at workplace ID 50516.

(a) Causal estimation, power consumption dataset:	Workplace ID 50516		
	DirectLiNGAM	RESIT	Manual model
Heat Treatment	−27.10	−27.10	−0.24
Weight	17.53	18.83	−1.85
Height	−3.12	11.83	−11.35
Outer Diameter	88.97	88.97	0.71

Table A4. Estimated causal effects for process duration at workplace ID 50516.

(b) Causal estimation, process duration dataset:	Workplace ID 50516		
	DirectLiNGAM	RESIT	Manual model
Heat Treatment	0	0	−0.23
Weight	0	0.04	0.08
Height	0.07	0.07	−0.05
Outer Diameter	−0.14	0.00	−0.03
Inner Diameter	−0.05	−0.13	0.03

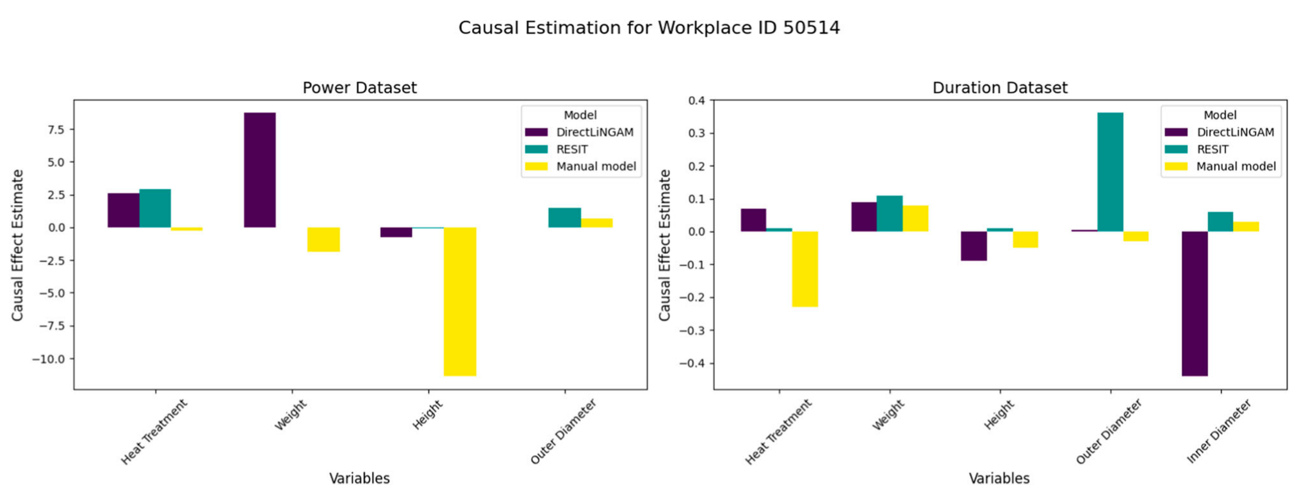


Figure A7. Causal estimation summary for workplace ID 50514 using DoWhy. The visualisation illustrates the estimated average treatment effects derived from the power consumption and process duration datasets.

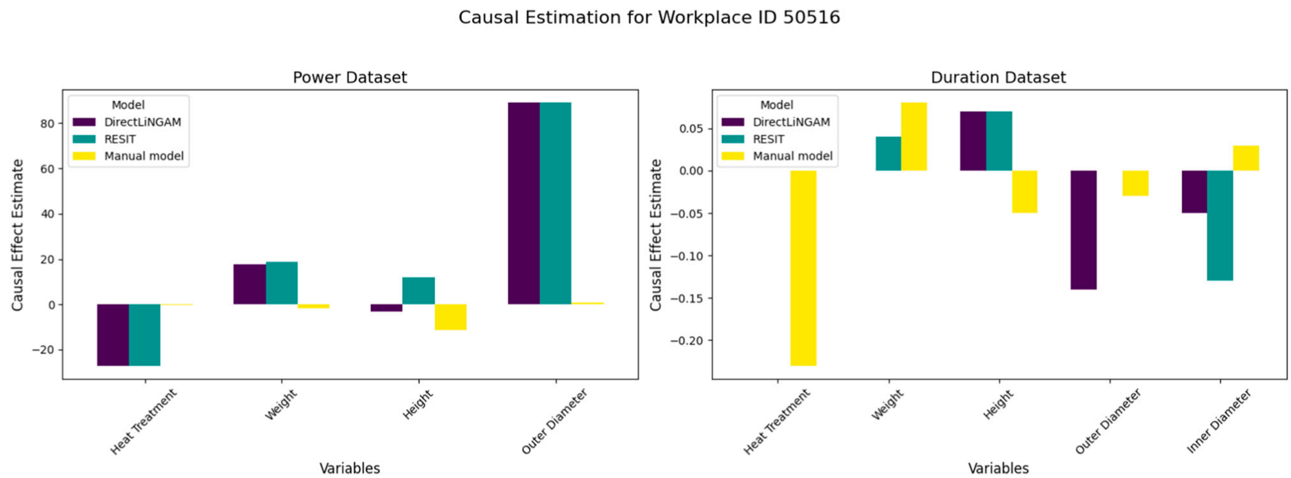


Figure A8. Causal estimation summary for workplace ID 50516 using DoWhy. The graph summarises the estimated causal effects of selected production features on energy use and process timing.