DEVELOPING AN INTEGRATED DATA-DRIVEN APPROACH TO OPTIMIZE PRODUCTION PROCESSES TOWARD SUSTAINABLE MANUFACTURING

Thesis for the Degree of Doctor of Philosophy (PhD)

MICHAEL MAIKO MATONYA

Supervisor: Dr. **ISTVÁN BUDAI**

University of Debrecen

Doctoral Council for Natural Sciences and Engineering

Doctoral School of Informatics

Debrecen, 2026

Declaration

Hereby I declare that I prepared this thesis within the Doctoral Council for Natural Sciences and Engineering, Doctoral School of Informatics, University of Debrecen in order to obtain a PhD Degree in Engineering at Debrecen University.
The results published in the thesis are not reported in any other PhD theses.

Debrecen, 2026

. . . .
signature of the candidate

Hereby I confirm that Michael Maiko Matonya candidate conducted his studies with my supervision within the Information Technology Systems and Networks with Industrial applications Doctoral Programme of the Doctoral School of Informatics of University of Debrecen between 2022 and 2026. The independent studies and research work of the candidate significantly contributed to the results published in the thesis. I also declare that the results published in the thesis are not reported in any other theses.
I support the acceptance of the thesis.

Debrecen, 2026

. . . .
signature of the supervisor

DEVELOPING AN INTEGRATED DATA-DRIVEN APPROACH TO OPTIMIZE PRODUCTION PROCESSES TOWARD SUSTAINABLE MANUFACTURING

Dissertation submitted in partial fulfilment of the requirements for the doctoral (PhD) degree in Engineering

Written by **Michael Maiko Matonya** certified Industrial Engineer Manager

Prepared in the framework of the Information Technology Systems and Networks with Industrial applications Doctoral Program of the Doctoral school of Informatics of University of Debrecen

Dissertation supervisor: Dr. **ISTVÁN BUDAI**

The official opponents of the dissertation:

Prof. Dr. Domicián MÁTÉ
Dr. habil. Mátyás ANDÓ

The evaluation board:

chairperson: Dr. ....
members: Dr. ....
Dr. ....
Dr. ....
Dr. ....

The date and venue of the dissertation defence: .... 20...

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

AHP	Analytical Hierarchy Process
BPMN	Business Process Model and Notation
CE	Carbon Emissions
CI	Consistency Index
CPS	Cyber-Physical Systems
CR	Consistency Ratio
CRITIC	Criteria Importance Through Intercriteria Correlation
CT	Cycle Time
dLCA	Dynamic Life Cycle Assessment
DR	Defect Rate
DRL	Deep Reinforcement Learning
DT	Downtime
EC	Energy Consumption
ERP	Enterprise Resource Planning
ES	Environmental Sustainability
FAHP	Fuzzy Analytical Hierarchy Process
GHG	Greenhouse Gas
GHGI	Greenhouse Gas Intensity
GRA	Grey Relational Analysis
GRA–TOPSIS	Grey Relational Analysis–TOPSIS Hybrid Ranking
GRC	Grey Relational Coefficient
GRG	Grey Relational Grade
IIoT	Industrial Internet of Things
IPSMF	Integrated Process-based Smart Manufacturing Framework
IT	Idle Time
KPI	Key Performance Indicator
LAM	Large Action Models

LCA	Life Cycle Assessment
LCI	Life Cycle Inventory
LCIA	Life Cycle Impact Assessment
LT	Lead Time
MCDM	Multi-Criteria Decision Making
MES	Manufacturing Execution System
MOO	Multi-Objective Optimisation
MOOM-OEP	Multi-Objective Optimisation Model for Operational and Environmental Performance
MOPSO	Multi-Objective Particle Swarm Optimisation
MOSA	Multi-Objective Simulated Annealing
NSGA-II	Non-dominated Sorting Genetic Algorithm II
OCEA	Object-Centric Event Analysis
OCEL	Object-Centric Event Log
OCPM	Object-Centric Process Mining
OEE	Overall Equipment Effectiveness
OP	Operational Performance
PM	Process Mining
PM-LCA	Process Mining–Life Cycle Assessment Integrated Dataset
RI	Random Consistency Index
RR	Recycling Rate
RU	Resource Utilisation
SCADA	Supervisory Control and Data Acquisition
TFN	Triangular Fuzzy Number
TP	Throughput
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
WC	Water Consumption
WG	Waste Generation

Symbols

L	Event log
σ_i	Trace (case) in an event log
e_j	Event within trace σ_i
E	Finite set of events

$\mathcal{A}$	Activity alphabet
$\mathcal{C}$	Set of case identifiers
O	Universe of objects (OCEL)
$\mathcal{T}_O$	Finite set of object types
$\mathcal{R}$	Raw manufacturing database
$\mathbf{x}_j$	Feature vector of production unit j
$\mathcal{L}$	Classical event log tuple
U_i	Resource utilisation of machine i
$T_{\text{active},i}$	Active processing time of resource i
$T_{\text{total},i}$	Total available time of resource i
$P_i(t)$	Instantaneous power demand of machine i at time t
E_e	Activity-level energy consumption for event e
$\varepsilon_m(s)$	State-specific power profile of machine m in state s
EF_i	Emission factor for activity i
Q_i	Detected defect count for activity i
δ_i	Waste conversion factor for activity i
η_{res}	Material/resource efficiency
$\mathbf{f}_e$	Environmental foreground vector for event e
$\mathcal{Q}_c$	Dataset completeness metric
$\mathcal{Q}_{\text{con}}$	Dataset consistency metric
Δt^*	Minimum temporal granularity across all data streams
X_{ijk}	Processing time decision variable for activity k, machine j, line i
Y_{ijk}	Binary machine assignment decision variable
Z_{ijk}	Resource allocation decision variable
C_{ij}	Cloud computational load of machine j in line i
$f_1(x)$	Operational efficiency objective function
$g_2(x)$	Environmental impact objective function
$w_1, \ldots, w_7$	Objective weighting factors
w_i	AHP-derived priority weight for criterion i (crisp baseline; see $w_j^\star$ for the hybrid weight used in this work)
$\mathbf{A}$	AHP pairwise comparison matrix (crisp baseline; see $\widetilde{\mathbf{A}}$ for the fuzzy form)

a_{ij}	Pairwise preference ratio of criterion i over j (crisp baseline; see $\widetilde{a}_{ij}$ for the fuzzy form)
$\lambda_{\max}$	Principal eigenvalue of the AHP comparison matrix (consistency verification)
P^*	Pareto-optimal front
d_i	NSGA-II crowding distance for solution i
A^+	TOPSIS ideal solution
A^-	TOPSIS negative-ideal solution
S_i^+	Euclidean separation of solution i from the ideal solution
S_i^-	Euclidean separation of solution i from the negative-ideal solution
C_i	TOPSIS relative closeness coefficient for solution i (crisp baseline)

Hybrid FAHP–Entropy Weighting

$\widetilde{a}_{ij} = (l_{ij}, m_{ij}, u_{ij})$	Triangular fuzzy pairwise comparison entry (lower, modal, upper)
$\widetilde{\mathbf{A}}$	Fuzzy pairwise comparison matrix on TFN entries
$\mu_{\widetilde{a}_{ij}}(x)$	TFN membership function
$\widetilde{S}_i$	Fuzzy synthetic extent value of criterion i (Chang's extent analysis)
$V(\widetilde{S}_i \geq \widetilde{S}_k)$	Degree of possibility that $\widetilde{S}_i$ dominates $\widetilde{S}_k$
$d'(C_i)$	Minimum degree of possibility of criterion C_i over all other criteria
w_j^{FAHP}, $\mathbf{w}^{\mathrm{FAHP}}$	FAHP-derived subjective weight (scalar/vector form)
r_{ij}	Min–max normalised entry of the Pareto decision matrix $\mathbf{X}$
p_{ij}	Normalised probability mass of criterion j in solution i
H_j	Shannon entropy of criterion j, normalised to $[0, 1]$
w_j^{Ent}, $\mathbf{w}^{\mathrm{Ent}}$	Shannon Entropy objective weight (scalar/vector form)
λ	FAHP–Entropy balance coefficient, $\lambda \in [0, 1]$ (reference $\lambda = 0.5$)
$w_j^\star$, $\mathbf{w}^\star$	Hybrid composite weight $\lambda \mathbf{w}^{\mathrm{FAHP}} + (1 - \lambda)\mathbf{w}^{\mathrm{Ent}}$

GRA–TOPSIS Ranking

v_{ij}	Weighted normalised value of solution i on criterion j

v_j^+, v_j^-	Positive and negative ideal reference values for criterion j
$\Delta_{ij}^{\pm}$	Deviation $\lvert v_{ij} - v_j^{\pm}\rvert$ of solution i from the positive/negative ideal
$\Delta_{\min}^{\pm}$, $\Delta_{\max}^{\pm}$	Global minimum and maximum of $\Delta_{ij}^{\pm}$ over all (i, j)
ζ	Grey distinguishing coefficient, $\zeta \in (0, 1]$ (reference $\zeta = 0.5$)
$\gamma_{ij}^{\pm}$	Grey relational coefficient of solution i on criterion j
Γ_i^+, Γ_i^-	Grey relational grade of solution i w.r.t. the positive/negative ideal
C_i^{TOPSIS}	TOPSIS closeness component of solution i
C_i^{GRA}	Grey relational closeness component of solution i
β	TOPSIS–GRA fusion coefficient, $\beta \in [0, 1]$ (reference $\beta = 0.5$)
$C_i^{\star}$	Combined GRA–TOPSIS closeness, $\beta C_i^{\mathrm{TOPSIS}} + (1 - \beta) C_i^{\mathrm{GRA}}$
$x^{\star}$	Recommended configuration $\arg\max_i C_i^{\star}$

Abstract

Background. Manufacturing generates abundant production data, yet operational efficiency, environmental sustainability, and operator well-being are typically analysed in isolation. Static LCA inventories and aggregate KPIs cannot close this gap.

Objective. This work develops an integrated, data-driven framework that couples object-centric process mining, dynamic LCA, multi-objective optimisation, and a hybrid decision layer (Fuzzy AHP and Shannon-Entropy weighting with GRA–TOPSIS ranking) to resolve the efficiency,sustainability and ergonomics trade-off.

Methods. The framework is organised in four stages: (i) automated conversion of raw sensor records into event logs; (ii) process discovery that exposes bottlenecks and rework loops; (iii) dynamic impact attribution that separates productive from idle emissions using a time-resolved Hungarian-grid emission factor $EF(t)$, set to 0.275 kg CO_2/kWh during the evening peak window $[16{:}00, 20{:}00)$ and 0.200 kg CO_2/kWh off-peak, together with a composite ergonomic risk index; and (iv) multi-objective optimisation followed by a hybrid weighting and ranking layer that selects the preferred operating point from the Pareto front.

Results. Applied to a tube manufacturing case, the framework identified bottleneck stations responsible for nearly half of daily emissions, with about one fifth originating from idle states, signatures undetectable by static LCA. The selected operating point delivered a **+5.64 percentage-point** efficiency gain (87.2% → 92.84%) and a **−17.01%** CO_2 reduction (863.2 → 716.3 kg CO_2/day), while ergonomic risk remained within the nominal band. Sensitivity analysis confirmed the stability of the top-ranked solutions.

Conclusions. The framework provides a replicable Industry 5.0 methodology for jointly optimising operational, environmental, and social criteria on live event-level data, with extensions toward multi-product scheduling and closed-loop ergonomic feedback. For the TL209 line, the rank-1 configuration P_5 is recommended as the preferred operating point and should be staged into production through the closed-loop MES write-back of the FAHP–Entropy/GRA–TOPSIS selection.

Acknowledgements

I thank Almighty God for His grace and divine providence, which sustained and guided me throughout this journey. I am sincerely grateful to my supervisor, Prof. Dr. István Budai, for his continuous support, patience, and motivation during my PhD study. His immense knowledge and experience have encouraged me throughout my studies.

I thank the Doctoral School of Informatics at the University of Debrecen for providing the environment and resources for this work. Special thanks go to Dr. Dr. Gergely Kocsis, Secretary of the Doctoral School, and Prof. Dr. Sándor Baran, Head of the Doctoral School, for their leadership and support. I thank Prof. Dr. János Sztrik, former Program Leader, along with all professors of the department, for their insights and guidance. Special appreciation goes to Dr. Zoltán Gál, current Program Leader, for his guidance in shaping my research direction.

I thank the library staff at the University of Debrecen for their assistance. My thanks go to the engineering and operations teams at the case study facilities for providing the essential data and domain expertise that made this research possible. I thank Prof. Tumaini Shabani Gurumo, Rector of the Dar es Salaam Maritime Institute (DMI), and the staff members of DMI for their support. I am grateful to my colleague Nyanga Honda and Dr. James Kachungwa for their support. Finally, I thank my family and friends for their endless love, understanding, and encouragement.

LLM Usage Disclosure

AI tools (Writefull in Overleaf) were used only for language refinement and LaTeX syntax correction. All scientific contributions and results are entirely original.

Chapter 1: Introduction

1.1 Research Background

The emergence of Industry 4.0 has enabled the collection of high-resolution production data through cyber-physical systems and Industrial Internet of Things (IIoT) technologies. Traditional optimization approaches, such as Lean Manufacturing and Six Sigma, focus primarily on operational efficiency, often treating environmental performance as a secondary concern[1, 2]. In contrast, Life Cycle Assessment (LCA) quantifies environmental burdens using industry-average static inventories that do not reflect the real-time variability of individual shop-floor decisions: a spatiotemporal dynamic LCA study found temporal variation of 23–38% and spatial variation of up to 76% in global warming potential for identical production designs evaluated under different conditions [3–5]. Process Mining and IIoT technologies offer a path for data-driven optimisation [6], yet existing applications remain confined to operational conformance without coupling insights with environmental performance: empirical evidence shows that process delays increase emissions by 16.7% and rework raises waste generation by 41.7% relative to the standard process flow [7, 8], costs that a conformance-only analysis would never surface. Currently, there is no integrated data-driven framework that simultaneously optimizes manufacturing processes and dynamically quantifies the environmental impacts of the life cycle, leaving practitioners unable to balance the environmental cost of efficiency gains against the operational cost of environmental improvements [9].

Industry 4.0 has transformed operations by integrating CPS and IIoT into data-rich paradigms where high-speed operational data enable evidence-based optimization [10, 11]. However, standard indicators such as overall equipment effectiveness (OEE) lack the event-level resolution needed to identify causes. Environmental impacts such as energy consumption vary continuously with machine state and processing speed, creating a coupling where higher throughput often increases energy intensity [12, 13]. Without a framework to model these joint dynamics, interventions risk producing trade-offs that undermine long-term sustainability goals [9].

Running parallel to this technological consolidation, the European Commission has articulated *Industry 5.0* as the successor paradigm that re-orients the Industry 4.0 automation agenda around three complementary pillars: human-centricity, sustainability, and resilience [14–16]. Where Industry 4.0 is defined by mass automation, IoT, and big data, Industry 5.0 restores the shop-floor operator to the analytical frame by pairing advanced machines and collaborative robots (cobots) with the people who operate them [17, 18]. This shift has a direct methodological consequence for sustainability assessment in manufacturing: the environmental pillar that conventionally dominates green-manufacturing literature (energy, carbon, waste) must be extended with a *social* pillar that captures operator well-being, ergonomic load, and fatigue risk through wearable and vision-based sensing [19]. The framework developed in this dissertation therefore treats operational efficiency, environmental burden, *and* operator ergonomic risk as three jointly optimisable pillars, aligning the TL209 case study with the Industry 5.0 human-centric agenda rather than the purely technology-driven Industry 4.0 reading.

To address these challenges, this research couples Object-Centric Process Mining (OCPM) on the OCEL 2.0 standard with a dynamic Life Cycle Assessment (dLCA) module that attributes impacts to event-level shop-floor observations under a time-resolved Hungarian-grid emission factor, and complements them with an Ergonomic Risk Index (ERI) that brings the operator into the analytical frame in line with the Industry 5.0 social pillar [6, 19–22]. The resulting three-pillar trade-off, operational efficiency, carbon burden, and ergonomic load, is resolved by a constrained NSGA-II solver that returns a non-dominated front, on top of which a hybrid decision layer combining Fuzzy AHP and Shannon-Entropy weighting with GRA–TOPSIS ranking selects the preferred operating point across eight efficiency, environmental, and ergonomic criteria [23–26]. By executing the full pipeline on empirically recovered process data rather than on idealised reference models, the methodology closes the Reality–Model gap between as-designed assumptions and as-executed shop-floor behaviour, and it does so within a cloud-resident architecture that writes the rank-1 configuration back to the MES for closed-loop validation [11, 27].

1.2 Research Objectives

1.2.1 General Objective

The main objective is to develop an integrated data-driven approach to optimize production processes toward sustainable manufacturing.

1.2.2 Specific Objectives

Three specific objectives guide the work:

(i) Evaluate existing data-driven methodologies in sustainable manufacturing, identify their strengths and limitations, and determine the critical control parameters needed to align production efficiency with sustainability goals.
(ii) To develop an integrated data-driven approach to optimizing production processes towards sustainable manufacturing.
(iii) Validate the proposed framework against industrial manufacturing systems, demonstrating measurable improvements in sustainability and operational robustness.

1.2.3 Research Questions

These objectives are addressed through three corresponding research questions:

(i) What are the limitations of current data-driven approaches in sustainable manufacturing, and which process parameters are most critical for jointly optimizing production efficiency and environmental performance?
(ii) How can real-time process mining, dynamic life cycle assessment, and multi-objective optimization be integrated into a unified, cloud-based framework that captures the dynamic coupling between operational and environmental objectives?
(iii) How significantly does the proposed framework enhance both process performance and sustainability outcomes when validated against empirical industrial data?

1.3 Scientific Contributions

The dissertation delivers four contributions, each aligned with a formal thesis in Chapter 7.

- Identification of the integration gap between object-centric process mining, dynamic LCA, multi-objective optimisation, and MCDM, and determination of the four critical control parameters $\{IT, EC, CT, DR\}$ that link efficiency, environmental burden, and ergonomic load on a single analytical substrate.
- An end-to-end closed-loop framework, IPSMF-MOOM-OEP, coupling OCEL 2.0 process mining, dynamic LCA with a time-resolved Hungarian-grid emission factor, feasibility-constrained NSGA-II, and an Industry 5.0 Ergonomic Risk Index, optimising operational, environmental, and human-centric criteria on live event-level data.
- A hybrid decision layer combining Fuzzy AHP and Shannon Entropy weighting with a GRA–TOPSIS ranking across an eight-criterion matrix, producing an auditable single-recommendation output with traceable fuzzy consistency.
- Empirical validation on the TL209 line (86,466 events, nine stations): $+5.64$ pp efficiency ($87.2\% \rightarrow 92.84\%$) and -17.01% CO_2 ($863.2 \rightarrow 716.3$ kg CO_2/day) at rank-1, with rank-1 stability confirmed by a four-tier sensitivity analysis.

1.4 Dissertation Organization

The dissertation is organised in eight chapters. Chapter 2 reviews the literature on process mining, dynamic LCA, multi-objective optimisation, and Industry 5.0 ergonomics. Chapter 3 consolidates the dataset description, covering the source taxonomy, quality metrics, TL209 site, acquisition pipeline, sensor suite, and numerical baseline, so that subsequent chapters reference it from one canonical location. Chapter 4 performs the gap analysis and states the refined research questions, objectives, and the four control parameters $\{IT, EC, CT, DR\}$. Chapter 5 formalises the IPSMF-MOOM-OEP framework as a four-stage pipeline (OCEL 2.0 discovery, dynamic LCA with Ergonomic Risk Index, NSGA-II, and FAHP, Entropy and

GRA–TOPSIS ranking). Chapter 6 reports the TL209 validation: discovery accuracy, baseline emissions, the Pareto front, the rank-1 selection, and the sensitivity analysis. Chapter 7 presents the three formal theses in the Debrecen Claim–Method–Result–Novelty pattern. Chapter 8 states the limitations and the future research programme.

Chapter 2: Literature Review

Optimising production processes while reducing environmental impact is not a problem that can be solved by any single methodology. It requires an understanding of how industrial data is structured and collected, which performance indicators best capture both operational and environmental behaviour, and which analytical methods are capable of recovering process reality from raw event data and translating it into sound decisions. Research across these areas has advanced considerably in recent years, yet the fields of process mining, life cycle assessment, and multi-objective optimisation have largely progressed in isolation, each addressing part of the problem without engaging the others. The literature reviewed in this chapter covers these intersecting fields, examining what has been established, where recognised limitations remain, and why an integrated data-driven approach is needed to bridge the gaps that existing work has left open.

The formal definitions of the industrial dataset taxonomy, the progressive hierarchy of data representations, the quality metrics $\mathcal{Q}_c$, $\mathcal{Q}_{\text{con}}$, and Δt^*, the TL209 case-study site, the on-site acquisition pipeline, the Industry 5.0 Minimum Viable Sensor Suite, and the numerical baseline are developed in Chapter 3. The present chapter covers the production and sustainability metrics, the multi-objective optimisation literature, and the multi-criteria decision-making background that situate the framework's analytical contribution.

2.1 Production Metrics

Understanding the actual manufacturing behavior requires formalizing the variables that constitute the Reality-Model Gap [6]. To successfully conduct multi-parameter optimization on complex manufacturing lines, fundamental variables and performance metrics must be identified and continuously monitored. The transition from traditional manufacturing to Industry 4.0 requires a paradigm shift in how performance is measured, moving from isolated indicators to interconnected metrics that jointly capture operational and environmental performance.

The primary productivity measure is Throughput (TP), representing the volume of accepted units successfully processed per hour. It is closely related to Cycle Time (CT), which is the continuous processing time required to complete a discrete production activity at a given machine. Lead Time (LT), on the other hand, encompasses the total time from the initiation of a production order to its completion, including all processing, waiting, and transport times. Overall Equipment Effectiveness (OEE) serves as a comprehensive metric that combines availability, performance, and quality to evaluate how effectively a manufacturing operation is utilized compared to its full potential [28].

Non-productive times are equally critical. Idle Time (IT) refers to the queueing or waiting duration caused by stochastic machine behaviors, blockages, or starvation across the line. Downtime (DT) includes all periods when a machine is scheduled for production but is not operating due to failures, maintenance, or setups. Resource Utilization (RU) measures the percentage of available time that machines or personnel are actively engaged in productive tasks. The Defect Rate (DR) is a critical quality variable indicating the ratio of rejected components arising from mechanical drift or material variance.

Table 2.1: Key operational performance metrics in data-driven manufacturing systems.

Metric	Description	Category	References
Throughput (TP)	Units successfully processed per unit time.	Productivity	[28, 29]
Cycle Time (CT)	Processing time per activity at a single machine.	Time	[30, 31]
Lead Time (LT)	Elapsed time from order release to completion.	Time	[29, 30]
OEE	Composite index: Availability × Performance × Quality.	Efficiency	[28, 31]
Idle Time (IT)	Non-productive waiting due to starvation or blockage.	Loss	[30, 32]
Downtime (DT)	Machine unavailability due to failure, maintenance, or setup.	Loss	[29, 31]
Resource Utilisation (RU)	Fraction of available time spent on productive tasks.	Efficiency	[28, 31]
Defect Rate (DR)	Ratio of non-conforming units to total units produced.	Quality	[29, 30]

2.2 Sustainability Metrics

Operational efficiency alone is insufficient for evaluating modern production systems. Every unit produced carries an environmental cost that aggregate output figures do not capture: energy drawn from the grid, emissions released to the atmosphere, water extracted for cooling and processing, and material discarded as scrap or waste [33, 34]. Quantifying these costs requires a distinct set of sustainability metrics that operate alongside operational indicators and that are sensitive to variations occurring at the level of individual machines, process steps, and product batches rather than annual facility totals [12, 35].

Five assessment domains organise the sustainability metrics relevant to manufacturing: energy performance, atmospheric emissions, material and water consumption, waste generation and circularity, and the social / human-centric pillar introduced by the Industry 5.0 paradigm [14, 17, 19]. Figure 2.1 illustrates this taxonomy. Within energy performance, Total Energy Consumption (EC) measures the aggregated power demand across all machine states, while Energy Intensity (EI) normalises this demand per unit of output to enable fair comparison across different production volumes [13, 33]. The Renewable Energy Ratio tracks the proportion of consumed energy drawn from non-fossil sources, an indicator of increasing relevance as manufacturers respond to decarbonisation targets [2].

Within the emissions domain, Carbon Emissions (CE) are computed by multiplying time-resolved energy consumption by a dynamic grid emission factor that reflects the carbon intensity of the electricity supply at each moment of production [21, 34]. Greenhouse Gas Intensity ($GHGI$) normalises total emissions per unit of output, while air pollutant indicators covering nitrogen oxides, sulphur oxides, and particulate matter address local atmospheric impacts beyond global warming [4, 36].

Material and water metrics capture resource flows through the production system. Material Efficiency (η_{res}) expresses the ratio of saleable output mass to total material input, directly quantifying yield losses caused by defects and scrap [35, 37]. Water Consumption (WC) accounts for water used in cooling, cleaning, and process chemistry, which can be a dominant environmental burden in thermally intensive operations [34]. Specific Resource Consumption aggregates multiple input streams into a single normalised indicator per unit of output.

Waste and circularity metrics close the environmental accounting picture. Waste Generation (WG) records the cumulative mass of rejected or non-viable material that leaves the system as waste rather than product [35]. Recycling Rate (RR) captures the proportion of process waste that re-enters productive use, and Hazardous Waste (HW) separately tracks regulated streams that carry disposal obligations beyond standard solid waste [4, 36].

The social (human-centric) pillar, aligned with the Industry 5.0 reframing of sustainability [14, 18], captures the cumulative occupational load placed on the operators stationed at line-paced workstations. The three component indicators adopted here are the Rapid Upper Limb Assessment (RULA) [38] for awkward-posture frequency, the Occupational Repetitive Actions index (OCRA) [39] for technical-action repetition, and the revised NIOSH lifting index (NIOSH-LI) [40] for manual-handling load. The three are combined into a single Ergonomic Risk Index (ERI) following the equal-weight composition rule of Ciccarelli et al. [19]; the formal definition of the composite and its real-time sensor instrumentation are developed in Chapter 5, section 5.1.5. Together, the resulting fifteen indicators provide the quantitative basis for the dynamic three-pillar attribution that the integrated framework developed in this dissertation applies at the event level.

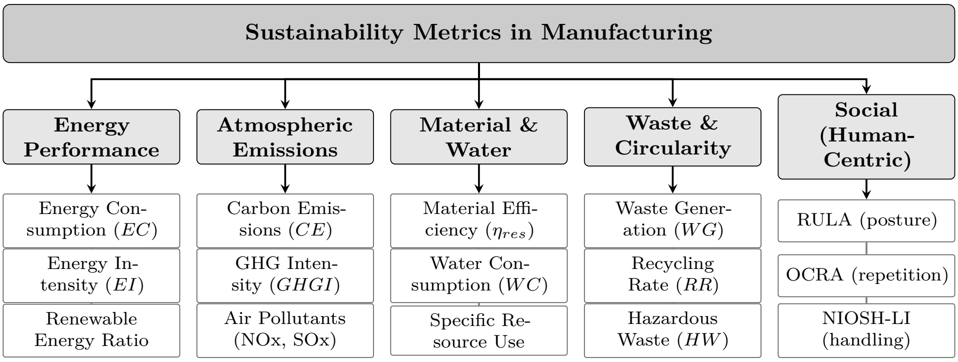


Figure 2.1: Taxonomy of sustainability metrics applicable to manufacturing systems, organised by assessment domain [4, 33–35]; the fifth column (Social, human-centric) is the Industry 5.0 extension adopted in this dissertation, with the three component indicators RULA, OCRA, and NIOSH-LI combined into the composite Ergonomic Risk Index (ERI) per Ciccarelli et al. [19].

2.3 Industry 5.0 and Human-Centric Sustainability Assessment

The taxonomy of Figure 2.1 is complete with respect to the Industry 4.0 sustainability agenda but is incomplete once the assessment frame is broadened to the Industry 5.0 paradigm. The European Commission's *Industry 5.0* policy brief [14] and the parallel conceptual consolidation by Xu et al. [15] and Leng et al. [16] restate the industrial modernisation agenda around three pillars: human-centricity, sustainability, and resilience. Where Industry 4.0 was defined by mass automation, the Internet of Things, and big data, Industry 5.0 re-humanises the factory floor by having human operators and advanced machines (including collaborative robots, or cobots) work side by side [17, 18, 41]. The methodological consequence for manufacturing-sustainability research is that the operational and environmental pillars studied by the literature covered in sections 2.1 and 2.2 must be extended with a *social (human-*

centric) pillar that captures operator well-being, ergonomic load, and cumulative fatigue risk.

2.3.1 The Social-Sustainability Pillar

In the Industry 5.0 frame, social sustainability is not a supplementary narrative concern but a measurable analytical object that competes directly against throughput and environmental objectives for the same decision-variable budget. Higher line speeds reduce cycle time and, at favourable points in the operating regime, also reduce per-unit energy intensity, but they simultaneously raise repetitive-motion exposure, awkward-posture frequency, and manual-handling load at the operators stationed at the bottleneck cells. Ciccarelli et al. [19] formalise this three-pillar view through a digital-twin workstation in which ergonomic risk is computed in real time alongside cycle time and energy consumption, and demonstrate that the three signals are mutually non-dominated over realistic assembly scenarios. The empirical lesson is structural: any optimisation framework that stops at two pillars under-reports the cost of throughput gains by systematically omitting the human operator.

2.3.2 Ergonomic Risk as a Composite Index

Three mature scalar indices span the dominant mechanical-load pathways on a discrete manufacturing line. RULA [38] scores upper-limb posture on a 1–7 scale from joint-angle observations; OCRA [39] quantifies cumulative repetition exposure as the ratio of technical actions to a force/posture/recovery-adjusted threshold; and the revised NIOSH lifting equation [40] yields a lifting index from the horizontal, vertical, angular, and frequency components of a manual-handling task. The three are complementary and their weighted composition into a single Ergonomic Risk Index (ERI) has been adopted in recent digital-twin implementations [19] because it captures posture, repetition, and load dimensions that any single index cannot. The dissertation adopts this three-component ERI as the scalar representing the social-sustainability pillar; the sensor suite, composition rule, and calibration are detailed in Chapters 3–5.

2.3.3 Minimum Viable Sensor Suite for Quantitative Social Sustainability

A key deployment concern for any Industry 5.0 extension of an Industry 4.0 pipeline is the additional instrumentation cost that the social pillar imposes. Ciccarelli et al. [19] and the parallel reviews of Industry 5.0 enabling technologies [17, 18] converge on a compact minimum viable sensor suite (MVSS) sufficient to populate RULA, OCRA, and NIOSH in real time: body-worn inertial measurement units (IMUs) for joint-angle tracking, surface electromyography (EMG) electrodes for muscle-activation and fatigue monitoring, and instrumented load cells or force-plate mats at the manual-handling stations. Optionally, a vision-based pose-estimation camera can replace or supplement the IMU channels when the workstation geometry permits occlusion-free line of sight. This MVSS is a one-time capital addition to the existing MES, IIoT and ERP stack and does not require modification of the production control system; the sensor streams feed the ergonomic-index calculators alongside the existing OCEL and power-meter streams. The detailed specification of the MVSS, its calibration against the three reference indices, and its integration into the optimisation objective function g_2 are presented in Chapter 5, section 5.1.5.

2.4 Object-Centric Event Analysis

Traditional process mining assumes that each event belongs to a single isolated case. In complex manufacturing settings, multiple entities interact concurrently, including raw materials, active machinery, tools, and diverse production orders [42]. Object-Centric Event Analysis (OCEA) represents a paradigm shift in how process data is structured and analyzed. Unlike traditional event logs that force a single case notion (often leading to convergence and divergence problems), OCEA allows an event to be related to multiple objects of different types simultaneously.

The Object-Centric Event Log (OCEL) standard addresses this structural limitation by linking events to multiple objects, mitigating the distortion caused by forcing case-centric perspectives [43]. In an OCEL, the complex multi-entity interactions inherent in modern manufacturing are preserved. For instance, a single "assembly" event might simultaneously involve a specific machine, a batch of raw materials, an operator, and a production order. By capturing these event-object relationships natively, OCEA provides a more accurate multi-entity view of the operational reality.

Object-Centric Process Mining (OCPM) builds upon OCEA to discover, conform, and enhance processes involving multiple interacting objects [44]. Classical process mining struggles with complexity and concurrency when multiple cases interact, often resulting in "spaghetti" models that are difficult to interpret. OCPM handles this by constructing object-centric behavioral models that explicitly represent the lifecycle of different object types and their synchronization points. This capability is crucial for identifying bottlenecks and inefficiencies that span across different process dimensions.

The main features of OCEA include multi-object representation, which eliminates the need for data flattening; event correlation, which explicitly links activities across different object lifecycles; scalability, enabling the analysis of massive industrial datasets; and flexibility, allowing analysts to dynamically shift their perspective depending on the analytical goal. Within these main features, several sub-features provide deeper analytical capabilities. Object lifecycle tracking monitors the state changes of individual entities from inception to completion. Relationship modeling captures the structural and temporal dependencies between different object types. Cross-object dependencies highlight how delays or variations in one object's lifecycle propagate to others, facilitating root-cause analysis in interconnected systems [45].

Table 2.2: Comparison of Traditional Process Mining and Object-Centric Process Mining.

Feature	**Traditional PM**	**Object-Centric PM**	**References**
Case Notion	Single, isolated case identifier.	Multiple interacting objects per event.	[6, 42]
Data Structure	Flattened event logs (XES format).	Relational event logs (OCEL standard).	[20, 43]
Concurrency	Prone to convergence and divergence artefacts.	Natively models multi-object synchronisation.	[42, 46]
Model Complexity	Spaghetti models in complex processes.	Structured, multi-dimensional process graphs.	[31, 47]
Analytical Scope	Fixed to the chosen case notion.	Dynamic perspective shifting across object types.	[43, 48]
Manufacturing Fit	Suitable for single-product workflows.	Designed for multi-entity shop-floor environments.	[30, 31]

Recent advances in process discovery demonstrate that object-centric graphs and machine learning capabilities can operate directly on these richer relational structures [47]. However, while OCEL improves the validity of discovered structural process models, it still requires integration with external empirical tools to account for instantaneous sustainability phenomena like power fluctuation or carbon emissions. The fusion of OCEA with real-time sensor data provides event-level visibility into both operational and environmental performance.

2.5 Life Cycle Assessment Limits

Life Cycle Assessment (LCA) provides the fundamental accounting principles for evaluating the environmental footprints of products, processes, and services [49]. In a general context, LCA is a standardized methodological framework designed to quantify the environmental impacts associated with all stages of a product's life, from raw material extraction through materials processing, manufacture, distribution, use, repair and maintenance, and disposal or recycling. The main components of an LCA include the Goal and Scope Definition, which outlines the study's boundaries and objectives; Inventory Analysis (LCI), which involves compiling data on energy and resource inputs and environmental releases; Impact Assessment (LCIA), which evaluates the potential environmental impacts of the inventory data; and Interpretation, where the findings are evaluated in relation to the goal and scope to reach conclusions and recommendations [4].

In manufacturing, LCA is extensively applied for energy analysis, emissions tracking, and overall sustainability evaluation. It provides a structured approach to identify environmental hotspots within production lines and evaluate the trade-offs associated with different operational strategies. The literature presents diverse applications of LCA. For instance, some authors have used LCA to evaluate the environmental benefits of adopting renewable energy sources in manufacturing, while others have focused on comparing the life cycle impacts of different material selection strategies [50].

To formalize the environmental evaluation, mathematical formulations are integrated into the LCA framework. The total Energy Consumption (E_{total}) of a manufacturing process can be modeled as the sum of the energy consumed during productive states (E_{prod}), idle states (E_{idle}), and setup phases (E_{setup}):

$$E_{total} = \sum_{i=1}^{n} (P_{prod,i} \cdot t_{prod,i} + P_{idle,i} \cdot t_{idle,i} + P_{setup,i} \cdot t_{setup,i}) \tag{2.1}$$

where P represents the power demand and t represents the duration for each state at machine i. The corresponding Carbon Emissions (CE) are calculated by multiplying the energy consumption by a dynamic emission factor ($EF(t)$), which accounts for the temporal variations in the carbon intensity of the energy grid:

$$CE = \int_0^T E_{total}(t) \cdot EF(t) dt \tag{2.2}$$

Resource efficiency (η_{res}) can be quantified as the ratio of useful product output (M_{out}) to the total material and energy inputs (M_{in}, E_{in}):

$$\eta_{res} = \frac{M_{out}}{M_{in} + \alpha E_{in}} \tag{2.3}$$

where α is a conversion factor aligning material and energy units.

A widely recognized limitation in classical LCA application is its reliance on historical, retrospective, or aggregated inputs that fail to capture real-time behavior in fast-paced Industry 4.0 lines [51]. Industrial application restrictions dictate that environmental indicators should remain temporally dynamic and directly associated with specific product batches or machine pulses [34]. Dynamic LCA (dLCA) overcomes these barriers by aligning environmental calculations with actual process observations instead of static inventory averages. This real-time visibility provides the necessary conditions for high-resolution carbon tracking that can act as a reliable input for multi-parameter industrial systems [33].

The fundamental deficiency of static LCA in Industry 4.0 environments is a temporal mismatch between environmental-data collection and manufacturing-decision granularity. Annual inventory averages smooth out the shift-level fluctuations characteristic of discrete manufacturing lines: a tube-forming machine draws different power in its productive, idle, and setup states, and the distribution across these states varies between shifts as order volumes, operator experience, and ambient temperature differ [4]. Collapsing these into a single annual average produces an estimate that may be accurate at year-end yet unrepresentative of any individual shift, batch, or product instance, and it cannot assign a defensible emission quantity to a single event because the dynamic context is not recorded. Bottleneck identification requires knowing that a station is simultaneously slow and carbon-intensive during that slowness, a diagnostic combination that only event-level attribution can confirm.

The enabling conditions for dynamic LCA are precisely those provided by modern IIoT-equipped production environments. Real-time sensor feeds delivering instantaneous power measurements at sub-second resolution, timestamped MES event records that bound the start and end of each process activity, and process-specific emission coefficients derived from metered grid data together allow an environmental inventory to be compiled on a per-event basis rather than per annual average [12, 21]. Under this architecture, the carbon footprint of each individual production event e is computed as the product of its measured energy consumption E_e (Equation 3.11) and the grid emission factor $EF(t_e)$ prevailing at the moment of execution:

$$CE_e = E_e \cdot EF(t_e), \tag{2.4}$$

where t_e is the event timestamp and $EF(t_e)$ is a time-varying coefficient that reflects the marginal carbon intensity of electricity generation at that instant. Walker and Harvie [34] demonstrate that using dynamic emission factors, rather than a single annual average, can alter per-product carbon attribution by up to 30% in grid-intensive manufacturing, a discrepancy large enough to change the environmental ranking of competing process configurations. Aggregating these per-event footprints across all events in a case trace yields a product-specific LCA that inherits the full temporal and process-pathway resolution of the underlying event log.

Table 2.3: Comparison of Static LCA, Dynamic LCA, and PM-LCA on key methodological dimensions relevant to manufacturing process optimisation.

Dimension	**Static LCA**	**Dynamic LCA**	**PM-LCA**
Inventory source	Annual average database records (e.g. Ecoinvent)	Time-resolved metered energy and grid emission factors	Per-event sensor and MES records linked to OCEL traces
Temporal resolution	Annual or batch-period aggregates	Shift-level or hourly intervals	Individual process-event level (seconds to minutes)
Attribution granularity	Product category or facility average	Production order or batch	Individual event, machine, and process path
Suitability for real-time optimisation	Low: updated at fixed reporting intervals	Medium: enables shift-level feedback	High: supports event-triggered carbon attribution and bottleneck targeting

Table 2.3 positions these three paradigms against one another on four dimensions directly relevant to the framework developed in this dissertation. The PM-LCA column corresponds to the integrated approach adopted here, in which every row of the environmental foreground vector $\mathbf{f}_e$ (Equation 3.13) is populated from live sensor and event-log data rather than from background database averages, achieving the event-level attribution granularity that static and standard dynamic LCA cannot provide.

2.6 Optimization and Ranking

2.6.1 The NSGA-II Approach

Operating a modern Industry 4.0 "Smart Factory" entails balancing intrinsically conflicting parameters like production speed against environmental emissions. Multi-objective optimization is essential in these environments to identify solutions that offer the best possible trade-offs. Deb's Non-Dominated Sorting Genetic Algorithm II (NSGA-II) is highly regarded across industrial operations precisely because it performs evolutionary Pareto-based optimization while guaranteeing diverse exploration [52].

NSGA-II is an evolutionary algorithm designed to solve multi-objective optimization problems. The algorithm begins with Initialization, where a random population of potential solutions (chromosomes) is generated. This is followed by Non-dominated Sorting, where the population is sorted into different fronts based on Pareto dominance. Solutions in the first front are completely non-dominated by any other solutions in the current population. To maintain diversity among the solutions, NSGA-II employs a Crowding Distance metric, which estimates the density of solutions surrounding a particular individual. The algorithm then iteratively applies Selection, Crossover, and Mutation operators to generate new offspring populations, continually refining the Pareto frontier over successive generations.

In manufacturing, NSGA-II is widely applied to optimize scheduling, resource allocation, and process parameters, seeking to minimize cycle times while simultaneously reducing energy consumption and defect rates. Compared to traditional single-objective Genetic Algorithms, which require the aggregation of objectives using arbitrary weights, NSGA-II explicitly preserves the trade-off relationships between conflicting goals. When compared to other multi-objective techniques like Particle Swarm Optimization (MOPSO) or Simulated Annealing (MOSA), NSGA-II is often praised for its robustness and its ability to maintain a diverse set of solutions along the Pareto front [9].

NSGA-II rests on three interlocking properties that matter when the Pareto front is non-convex, as in energy-throughput-quality trade-offs [52]. Binary tournament selection applies Pareto-dominance rank as the primary criterion and crowding distance as a tiebreaker, spreading selection pressure across the front rather than concentrating it at one region. The crowding-distance metric preferentially retains solutions in sparse neighbourhoods, preventing the drift toward local clusters that afflicts algorithms without explicit diversity preservation [23]. The elitist $(P_t \cup Q_t) \to N$ survivor strategy guarantees that a higher-rank solution cannot be displaced by a lower-rank one across any generation, a property single-objective GAs achieve trivially but multi-objective variants must engineer explicitly [52]. Together these converge toward the true front from multiple directions, which weighted-sum methods cannot recover on concave segments.

Practical performance depends on hyperparameters calibrated to problem dimensionality [23]. For five- to ten-variable problems, population sizes of 100–200 maintain diversity without prohibitive per-generation cost; Simulated Binary Crossover probabilities of 0.8–0.9 are consistently recommended in the manufacturing literature [52]; mutation probability is typically set to $1/n_{\text{var}}$ so that each chromosome undergoes one mutation per generation on average; and termination may use a fixed generation count or a hypervolume-change threshold $\Delta\mathcal{H}$. The values adopted for TL209 are reported in Chapter 3, section 3.7.

Table 2.4: Comparison of multi-objective optimisation algorithms applicable to manufacturing process optimisation, evaluated across convergence behaviour, diversity preservation, scalability, and documented manufacturing applications.

Algorithm	**Strengths**	**Limitations**	**Manufacturing Application**	**References**
NSGA-II	Elitist archive preserves best solutions across generations; crowding-distance metric explicitly maintains Pareto-front spread; guarantees convergence to the true front for convex and non-convex objectives; well-established for 2–3 objective problems.	Computational cost scales as $O(MN^2)$ per generation (M objectives, N population); diversity degrades for $>$ 3 objectives; crowding distance becomes less discriminating in high-dimensional objective spaces.	Production scheduling, energy–throughput parameter tuning, sustainable process optimisation.	[23, 52]
MOPSO	Rapid convergence via velocity-based search; simple parameter set (inertia weight, cognitive/social coefficients); naturally suited to continuous decision variable spaces.	Prone to premature convergence on local Pareto regions when the front is discontinuous; archive management required to maintain diversity; performance is sensitive to inertia weight schedule.	Continuous process parameter optimisation, real-time machine speed tuning.	[9, 53]
MOSA	Acceptance of inferior solutions (Metropolis criterion) enables escape from local Pareto fronts; effective on non-smooth and multimodal objective landscapes.	Convergence speed is slow relative to population-based methods; cooling schedule and initial temperature require careful problem-specific tuning; generates one solution per run, necessitating multiple restarts to approximate the Pareto front.	Complex combinatorial scheduling, job-shop and flexible-flow-shop problems.	[54]
Weighted-Sum GA	Computationally inexpensive; straightforward single-objective solver applicable with standard GA libraries; intuitive for decision-makers with predefined preferences.	Pre-specified aggregation weights define a linear scalarisation that cannot discover solutions on non-convex Pareto segments; weight sensitivity means small changes in w_i can yield qualitatively different solutions; full trade-off frontier is unavailable for post-hoc inspection.	Problems with well-defined, stable management preferences; preliminary screening of objective compatibility.	[32]
MOEA/D	Decomposes multi-objective problem into scalar subproblems solved in parallel; scales efficiently to many-objective ($\geq$4) problems; neighbourhood-based mating preserves local front structure.	Performance depends on the choice of decomposition scalarisation (Tchebycheff vs. weighted sum vs. PBI); uniform reference vector distribution assumes a regular Pareto-front geometry, which may not hold in manufacturing.	Many-objective process design, supply chain network optimisation.	[28]

In studies of time-cost-environmental trade-offs, NSGA-II systematically yields an "Industrial Feasibility Barrier" a frontier of configurations beyond which no metric can be improved without sacrificing another [54].

A critical distinction between NSGA-II and single-objective approaches lies in the treatment of decision-maker preferences. Weighted-sum and penalty-function formulations require that preferences be encoded as aggregation weights *before* the search [32], which structurally excludes Pareto-optimal solutions lying on non-convex segments of the true front and forces weight selection without knowledge of the trade-off surface; penalty coefficients further introduce problem-specific convergence artefacts [54]. NSGA-II defers the preference decision to a post-optimisation ranking stage: the algorithm delivers the complete front first, and a structured decision-support layer (in this dissertation, the GRA–TOPSIS hybrid with FAHP–Entropy weights) then selects the preferred operating point. This separation preserves the full Pareto surface for inspection and anchors the final recommendation in auditable preferences rather than search-phase artefacts [32, 54].

2.6.2 GRA–TOPSIS and Multi-Criteria Prioritisation

A Pareto front is Pareto-optimal by definition but under-determined for execution: all points are equally non-dominated, so a post-optimisation ranking rule is required to select one operating configuration. The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) supplies this rule by mapping each alternative to a scalar closeness coefficient $C_i \in [0, 1]$ based on its Euclidean distances to the ideal and anti-ideal reference vectors [55]. The Euclidean-distance view, however, is invariant under permutations of criterion-wise residuals: two alternatives with identical D_i^+ and D_i^- receive identical ranks even when their weighted normalised vectors $\mathbf{v}_i$ differ criterion-by-criterion relative to A^+. To recover this point-wise information, the dissertation augments TOPSIS with Grey Relational Analysis (GRA) [56], which measures element-wise similarity of $\mathbf{v}_i$ to the reference vectors, yielding the GRA–TOPSIS rule of [57–59]. The hybrid weight derivation is given in section 2.6.3.

The TOPSIS procedure begins by constructing a decision matrix $X = [x_{ij}]$, where rows correspond to candidate solutions and columns to performance criteria. Each entry is normalised to remove dimensional heterogeneity, and the normalised matrix is then multiplied by a vector of criterion weights to form the weighted normalised matrix $V = [v_{ij}]$. The ideal solution A^+ is the vector of best attainable values across all criteria, and the negative-ideal solution A^- is the vector of worst attainable values. For each candidate i, the Euclidean separations S_i^+ and S_i^- from these two reference points are computed, and the relative closeness is $C_i = S_i^- / (S_i^+ + S_i^-)$. A solution with C_i approaching unity is simultaneously close to the ideal and distant from the worst case, placing it at the top of the ranking. The method has been applied extensively in production systems for supplier selection, facility layout evaluation, and the prioritisation of process configurations under conflicting operational and environmental objectives [26, 60].

Normalisation removes the dimensional heterogeneity of the decision matrix columns (here: efficiency in per cent, CO_2 in kg/day, energy in kWh/day). Vector normalisation divides each entry x_{ij} by the Euclidean norm of its column:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{kj}^2}}, \tag{2.5}$$

producing a dimensionless matrix $R = [r_{ij}]$ whose columns each lie on the unit sphere. An alternative approach is linear (min-max) normalisation, which maps each criterion column to the interval $[0, 1]$ by subtracting the column minimum and dividing by the column range. Although linear normalisation is computationally equivalent and equally common in the MCDM literature, vector normalisation is preferred in manufacturing MCDM applications because it preserves the proportional relationships between solutions within each criterion [28]: a solution that achieves twice the energy savings of another retains that two-to-one ratio after vector normalisation, whereas linear normalisation compresses the ratio toward unity when the column range is large relative to the differences among mid-range solutions. This proportionality preservation is consequential when the Pareto front contains solutions clustered in one region and isolated extreme points in another, a pattern common in energy-throughput trade-offs.

The closeness coefficient C_i can be given a precise geometric interpretation that makes TOPSIS results directly communicable to production engineers. After the weighted normalised matrix V is formed, the ideal solution A^+ and negative-ideal solution A^- are two reference vectors in the n-dimensional weighted criterion space. The Euclidean separations S_i^+ and S_i^- are the Euclidean distances from solution i to these two reference points, and the closeness coefficient

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \in [0, 1] \tag{2.6}$$

can be interpreted as the fractional position of solution i along the line segment connecting the worst-case and ideal reference vectors in that weighted space. A solution with $C_i = 1$ coincides with the ideal reference point, achieving simultaneously the best value on every weighted criterion; a solution with $C_i = 0$ coincides with the worst-case reference, meaning it is simultaneously as far from the ideal as possible. Solutions with intermediate C_i values represent a trade-off between proximity to the ideal and distance from the worst case, and the monotone $[0, 1]$ scale maps directly onto a percentage-of-ideal interpretation that is intuitive for operational engineers reviewing results on a production dashboard. This geometric framing also clarifies why TOPSIS can yield a different ranking from simple distance-to-ideal methods: by anchoring the denominator to the sum of both separations, it penalises proximity to the worst case as well as rewarding proximity to the ideal, thereby distinguishing solutions that are merely average from solutions that are actively balanced.

The Euclidean closeness in equation (2.6) captures *how far* solution i lies from the ideal and anti-ideal, but not *in what shape* its criterion profile bends relative to those references. Grey Relational Analysis closes that gap by quantifying profile similarity point-by-point [56, 57]. To emphasise the dual role, the TOPSIS component of the hybrid is henceforth written $C_i^{\text{TOPSIS}} \equiv C_i$ with

$$C_i^{\text{TOPSIS}} = \frac{S_i^-}{S_i^+ + S_i^-}, \tag{2.7}$$

where $S_i^{\pm}$ retain the Euclidean definitions established above.

For the grey-relational limb, the deviation of the weighted normalised value v_{ij} from the positive and negative ideals is formed criterion-by-criterion,

$$\Delta_{ij}^+ = \left| v_j^+ - v_{ij} \right|, \qquad \Delta_{ij}^- = \left| v_j^- - v_{ij} \right|, \tag{2.8}$$

and the global extrema $\Delta_{\text{min}}^{\pm} = \min_{i,j} \Delta_{ij}^{\pm}$ and $\Delta_{\text{max}}^{\pm} = \max_{i,j} \Delta_{ij}^{\pm}$ are computed over the entire Pareto set. The grey relational coefficient that measures local similarity of solution i on criterion j to the two references follows the standard definition of Deng [56],

$$\gamma_{ij}^{\pm} = \frac{\Delta_{\text{min}}^{\pm} + \zeta\, \Delta_{\text{max}}^{\pm}}{\Delta_{ij}^{\pm} + \zeta\, \Delta_{\text{max}}^{\pm}}, \qquad \zeta \in (0, 1], \tag{2.9}$$

in which ζ is the distinguishing coefficient (conventionally $\zeta = 0.5$) that controls how sharply a deviation penalises similarity. Aggregating across criteria with the same composite weights $w_j^{\star}$ used by the TOPSIS limb yields the grey relational grades

$$\Gamma_i^{\pm} \;=\; \sum_{j=1}^{n} w_j^{\star}\,\gamma_{ij}^{\pm}, \tag{2.10}$$

and the grey-relational closeness of solution i is

$$C_i^{\mathrm{GRA}} \;=\; \frac{\Gamma_i^{+}}{\Gamma_i^{+}+\Gamma_i^{-}} \in [0,1]. \tag{2.11}$$

Whereas C_i^{TOPSIS} rewards overall Euclidean proximity to the ideal, C_i^{GRA} rewards consistently ideal-like behaviour on *every* criterion, penalising solutions that average well only because a large advantage on one criterion masks poor performance on another.

The two rankings are combined into a single hybrid closeness through a convex fusion coefficient $\beta \in [0,1]$,

$$C_i^{\star} \;=\; \beta\, C_i^{\mathrm{TOPSIS}} + (1-\beta)\, C_i^{\mathrm{GRA}}, \tag{2.12}$$

and the recommended configuration is $x^{\star} = \arg\max_i C_i^{\star}$. Equation (2.12) recovers pure TOPSIS at $\beta = 1$ and pure GRA at $\beta = 0$; the reference value $\beta = 0.5$ used throughout this dissertation weights distance-to-ideal and profile-similarity equally, and its robustness is verified by the sensitivity sweep reported in Chapter 6. This convex blend is the formulation adopted by Shen et al. [58] and Zhou et al. [59] for sustainability-oriented manufacturing decisions, and its adoption here is motivated precisely by the shape-sensitivity argument developed above.

Applications of TOPSIS in manufacturing optimisation span a range of decision contexts that closely parallel the TL209 case. Palczewski and Sałabun [26] employ TOPSIS for green supplier selection, demonstrating that the method produces stable rankings when criteria weights are derived from structured expert elicitation and when the number of alternatives does not exceed approximately 20. Both conditions apply here. Behera et al. [28] apply TOPSIS to process parameter optimization in a discrete manufacturing environment, confirming that the method's sensitivity to weight perturbations is manageable when criteria are physically interpretable and when the weight vector is subjected to scenario analysis. Zizovic and Pamucar [60] demonstrate the suitability of TOPSIS for multi-objective scheduling in high-volume production, where the method's ability to handle five to twenty candidate solutions and three to ten criteria simultaneously makes it an efficient post-optimisation ranking instrument. These applications confirm that TOPSIS is well-matched to the Pareto-front ranking task in this framework: the NSGA-II output provides a solution set of manageable size, the criteria are physically measurable and interpretable, and the closeness coefficient provides a single-number summary that plant engineers can act upon.

TOPSIS does carry two documented methodological limitations. First, its rankings can be sensitive to the number of criteria included in the decision matrix: adding or removing a criterion changes all column norms and therefore all normalised values, potentially altering the relative ordering of solutions that were closely ranked under the original criterion set. Second, rank reversal a change in the relative ordering of two existing alternatives caused by the addition or removal of a third alternative is a known property of the method, arising because the ideal and negative-ideal reference vectors are redefined whenever the alternative set changes [26]. In the framework adopted in this dissertation, both limitations are mitigated by design: the criterion set is fixed by the three objectives of the optimisation (throughput, energy, emissions) and does not change between runs, and the alternative set is fixed by the NSGA-II Pareto front and is not modified during the ranking stage. The hybrid FAHP–Entropy weighting coupled with GRA–TOPSIS ranking (section 2.6.3) further addresses the weight-assignment subjectivity concern by blending a consistency-verified fuzzy pairwise comparison with a data-driven entropy contribution, ensuring that the weights entering Equation (2.6) are both reproducible and interpretable.

The critical sensitivity of TOPSIS lies in the weight vector. Small changes in the relative importance assigned to competing criteria can shift the top-ranked solution substantially, which is precisely why the four-tier sensitivity analysis presented in Chapter 6 sweeps the subjective–objective balance parameter λ, the grey distinguishing coefficient ζ, and the TOPSIS–GRA fusion coefficient β. The validity of that sweep, however, depends entirely on whether the weights themselves were derived through a rigorous and reproducible process. Assigning equal weights by default or choosing weights intuitively introduces an unjustifiable subjectivity that undermines the scientific credibility of the ranking.

2.6.3 The Fuzzy AHP, Shannon Entropy, and GRA–TOPSIS Hybrid Framework

Classical AHP–TOPSIS exhibits three structural gaps in this application. (i) *Linguistic*: the crisp nine-point Saaty scale cannot represent the vagueness of natural-language pairwise judgements without information loss [24, 61]. (ii) *Epistemological*: a consistency-verified crisp AHP weight vector is fully subjective and carries no information about the dispersion of the alternatives across the Pareto front. (iii) *Geometric*: the TOPSIS closeness coefficient depends only on Euclidean distances to A^+ and A^- and is therefore insensitive to element-wise correspondence between $\mathbf{v}_i$ and the reference vectors, allowing rank reversal on fronts containing clusters of near-equidistant alternatives with different criterion-wise composition [57, 58]. The hybrid addresses each gap with a dedicated component: Fuzzy AHP (FAHP) for (i), Shannon Entropy for (ii), and GRA fused with TOPSIS for (iii).

The FAHP replaces the crisp pairwise matrix of classical AHP with a fuzzy matrix whose entries are Triangular Fuzzy Numbers (TFNs), and extracts the criterion weight vector through Chang's extent analysis procedure [24]. The TFN formalism, originally proposed by van Laarhoven and Pedrycz [61] and refined by Buckley [62], admits the intrinsic vagueness of expert language and propagates it through the aggregation step rather than collapsing it prematurely onto an integer scale. For the TL209 decision problem, the FAHP layer produces a weight vector $\mathbf{w}^{\text{FAHP}}$ that reflects managerial priority while retaining a consistency-verified audit trail. Rank reversal, a theoretical concern for crisp AHP when alternatives change [63], is not operative here: the alternative set is fixed by the static NSGA-II Pareto front and is not modified between weight derivation and ranking.

The Shannon Entropy weighting [25] provides the objective counterpart. For each criterion column of the Pareto decision matrix, the normalised entropy $H_j \in [0, 1]$ is computed; the divergence $1 - H_j$ measures how much that criterion discriminates among alternatives, and the entropy weights $\mathbf{w}^{\text{Ent}}$ are obtained by normalising the divergence vector [64]. The resulting weights are a data-driven indicator of criterion-level information content that is entirely independent of the expert-elicited FAHP vector. The two are then combined through a convex blend $\mathbf{w}^\star = \lambda\, \mathbf{w}^{\text{FAHP}} + (1 - \lambda)\, \mathbf{w}^{\text{Ent}}$, where $\lambda \in [0, 1]$ controls the subjective–objective balance and is swept in the sensitivity analysis of Chapter 6. Unlike the CRITIC alternative discussed below, Shannon Entropy does not depend on the inter-criteria correlation structure of the decision matrix, which makes it more stable on the narrow, low-dimensional Pareto fronts typical of three-objective NSGA-II runs. Recent manufacturing MCDM studies [65] confirm that this FAHP and Entropy+TOPSIS architecture remains an active and current design pattern for multi-objective sustainability problems.

On the ranking side, GRA–TOPSIS replaces the single Euclidean view of classical TOPSIS with a fused similarity measure. The TOPSIS closeness C_i^{TOPSIS} retains the geometric-distance component, while the Grey Relational Grade C_i^{GRA}, computed in the spirit of Deng's original grey-systems formulation [56], captures the shape or trajectory similarity between the i-th alternative and the positive ideal reference. The two are fused through a balancing coefficient $\beta \in [0, 1]$ into a combined closeness $C_i^{\star} = \beta\, C_i^{\mathrm{TOPSIS}} + (1 - \beta)\, C_i^{\mathrm{GRA}}$, as originally proposed by Kuo et al. [57] and validated across a broad spectrum of MCDM applications by Shen et al. [58]. The grey-relational term further depends on a distinguishing coefficient ζ that tunes the resolution between strong and weak similarities; both β and ζ are included in the four-tier sensitivity sweep of Chapter 6.

A recent synergetic study by Zhou and Chen [59] deploys exactly the three-component architecture of this dissertation (intuitionistic fuzzy AHP for subjective weights, entropy for objective weights, and a distance/relation hybrid for ranking) in a data-fabric solution-selection context, and reports that the triad is markedly more stable than any of its constituents applied in isolation. This dissertation transfers the architecture to a manufacturing setting and adds a four-tier sweep over λ, ζ, and β; the recommended configuration remains rank-1 at the reference parameters $\lambda = \zeta = \beta = 0.5$ and across the full neighbourhood of plausible decision-maker preferences.

Table 2.5: Comparison of Multi-Criteria Decision Making methods applicable to post-optimisation solution selection in manufacturing settings, updated to reflect the Fuzzy AHP, Shannon Entropy, and GRA–TOPSIS hybrid adopted in this dissertation.

Method	**Strengths**	**Weaknesses**	**Weight source**	**Mfg. fit**
TOPSIS	Intuitive geometric logic; handles large alternative sets efficiently; computationally transparent	Requires externally supplied weights; sensitive to normalisation method choice	External input required	High
AHP (crisp)	Structured pairwise elicitation; consistency verification (CR); hierarchical decomposition	No representation of linguistic vagueness; purely subjective; susceptible to rank reversal with set changes	Expert judgement	High (weights only)
FAHP (Chang extent)	Preserves linguistic vagueness via TFNs; CR-verifiable; more robust to elicitation noise than crisp AHP	Slightly heavier elicitation; requires defuzzification at the aggregation step	Expert judgement (fuzzy)	Very high (weights)
Shannon Entropy	Fully objective, data-driven; complements subjective weights; stable on narrow Pareto fronts	Requires complete alternative set a priori; not meaningful as stand-alone ranking	Statistical (decision matrix)	High (supplementary)
GRA–TOPSIS	Fuses Euclidean distance with shape similarity; robust to rank-reversal clusters; ζ tunes discrimination	Two tunable parameters (β, ζ); requires construction of both ideal and reference series	External input required	Very high (ranking)
FAHP, Entropy and GRA–TOPSIS *(this work)*	Closes the linguistic, subjectivity, and geometric gaps simultaneously; three-parameter sensitivity sweep demonstrates robustness of the recommended configuration	Three-stage pipeline; requires a parameter budget for λ, ζ, β	Hybrid (FAHP $\oplus$ Entropy)	**Very high (adopted)**

Table 2.5 positions the FAHP, Entropy and GRA–TOPSIS hybrid of this dissertation against the five constituent or directly comparable MCDM methods that bear on its design. The manufacturing-fit rating reflects the practical deployability of each method in an industrial context where decision-makers require transparent, communicable reasoning and where the alternative set is generated by a preceding optimisation rather than enumerated by domain experts. The adopted hybrid receives the highest composite rating because it simultaneously satisfies (i) the linguistic-fidelity requirement via FAHP, (ii) the objective-grounding requirement via Shannon Entropy, and (iii) the rank-stability requirement via GRA–TOPSIS, while remaining fully auditable through the λ–ζ–β sensitivity sweep of Chapter 6.

Chapter 3: Dataset Description and Its Characteristics

This chapter establishes the empirical foundation of the dissertation. It describes the source taxonomy, the TL209 case-study site, the on-site acquisition pipeline, the Industry 5.0 sensor extension, and the resulting numerical inventory. The methodology in Chapter 5 and the analyses in Chapter 6 both take the dataset characterised here as their input; references to "the dataset" in those chapters resolve to one of the sections of the present chapter through the labels reproduced below.

3.1 Source Taxonomy and Conceptual Framework

The empirical foundation of smart manufacturing research rests on the systematic collection and analysis of high-resolution industrial datasets [6, 11]. Many production environments operating under Industry 4.0 frameworks generate data from a multiplicity of interconnected sources, including Enterprise Resource Planning (ERP) systems, Manufacturing Execution Systems (MES), Supervisory Control and Data Acquisition (SCADA) platforms, and dense networks of Industrial Internet of Things (IIoT) sensors embedded throughout the shop floor [10, 66]. Each of these layers captures a distinct facet of the production process: ERP records order management and scheduling decisions at the business tier; MES provides real-time machine-level production tracking; SCADA captures low-level signals, alarms, and actuation commands; and IIoT sensors deliver continuous streams of physical measurements such as temperature, vibration, pressure, and instantaneous power draw [67, 68]. Together, these heterogeneous sources form a multi-dimensional digital shadow of the manufacturing system and supply the raw material necessary for process mining, dynamic sustainability assessment, and multi-parameter optimisation [29, 31]. Depending on how this raw material is structured and analysed, four progressively richer data representations emerge: *(i)* tabular snapshot records used in conventional production-efficiency monitoring; *(ii)* classical event logs that introduce temporal process-awareness; *(iii)* object-centric event logs that capture multi-entity interactions; and *(iv)* integrated PM-LCA datasets that extend the operational record with environmental inventory data. Each level subsumes the previous one and demands a qualitatively different mathematical treatment, as formalised in the subsections that follow.

Before any analytical framework is imposed, the most primitive data atom generated by a manufacturing information system is the *raw operational record* an atomic observation drawn from a well-defined relational schema. For a system with k attribute domains $A_1, A_2, \ldots, A_k$, each raw record is a tuple

$$r \in A_1 \times A_2 \times \cdots \times A_k, \tag{3.1}$$

and the complete raw dataset accumulated over an observation window is the relation

$$\mathcal{R} \subseteq A_1 \times A_2 \times \cdots \times A_k, \qquad |\mathcal{R}| = M. \tag{3.2}$$

For a typical MES installation, the schema at minimum spans a timestamp domain $\mathbb{T}$, a finite machine-identifier set M, an operational state set $S = \{\text{running, idle, fault, setup}\}$, a part-identifier set P, and a measurement-value domain $V \subseteq \mathbb{R}$, so that each MES record takes the form [31, 67]

$$r_{\text{MES}} = (t,\ m,\ s,\ p,\ v) \in \mathbb{T} \times M \times S \times P \times V. \tag{3.3}$$

Correspondingly, a raw ERP record encodes the scheduling layer as [11, 66]

$$r_{\text{ERP}} = (\mathit{ord},\ \mathit{mat},\ q,\ t_{\text{due}},\ \mathit{status}) \in \mathcal{O} \times \mathit{Mat} \times \mathbb{R}_{>0} \times \mathbb{T} \times \Sigma, \tag{3.4}$$

where $\mathcal{O}$ is the order-identifier set, Mat the material catalogue, q the order quantity, t_{due} the due date, and Σ the finite status alphabet (e.g., {released, in-progress, completed, scrapped}). The union of all source-system relations,

$$\mathcal{R} = \mathcal{R}_{\text{MES}} \cup \mathcal{R}_{\text{ERP}} \cup \mathcal{R}_{\text{SCADA}} \cup \cdots, \tag{3.5}$$

constitutes the *raw manufacturing database* from which every higher-level analytical dataset is derived [11, 69]. In its native form $\mathcal{R}$ offers no direct analytical utility: the records carry no inherent notion of a production case (how to group records belonging to the same unit of work), no activity notion (how to interpret a state transition as a named event), and no guaranteed correspondence between contemporaneous observations from different source systems [6]. It is precisely this structural ambiguity that makes the choice of an analytical framing non-trivial: the same relation $\mathcal{R}$ can be projected into any of the four representations described below, and each projection preserves a different subset of the information it contains [6, 70].

Prior to the adoption of process mining methodologies, the dominant dataset archetype in production-efficiency research was the *structured tabular record*, assembled by periodically aggregating MES and ERP outputs into a flat, snapshot-based representation [11, 71]. In this setting, each production unit j a job, a batch, or a completed work order is encoded as a d-dimensional feature vector, and the complete efficiency dataset is the finite set

$$\mathcal{D}_{\text{eff}} = \{\mathbf{x}_j\}_{j=1}^{N}, \quad \mathbf{x}_j = \left(OEE_j,\ CT_j,\ DT_j,\ DR_j,\ EC_j,\ \dots\right)^\top \in \mathbb{R}^d, \tag{3.6}$$

where N is the number of observed production units in the recording window and each coordinate of $\mathbf{x}_j$ corresponds to a scalar key performance indicator such as Overall Equipment Effectiveness (OEE), Cycle Time (CT), Downtime (DT), Defect Rate (DR), or Energy Consumption (EC). The dataset $\mathcal{D}_{\text{eff}}$ supports regression analysis, clustering, and statistical process control [72, 73], and is the natural output of conventional MES reporting dashboards [71]. However, it suffers from a fundamental structural limitation: it discards the sequential execution order of production activities. Two units may share an identical vector $\mathbf{x}_j$ yet have followed entirely different operational paths through the factory, a distinction that $\mathcal{D}_{\text{eff}}$ cannot represent [6]. This inability to capture the *how* the sequence, concurrency, and waiting patterns of production motivates the introduction of the event log as the primary data type within process mining [70, 74].

The oldest and most studied data structure in process-oriented manufacturing research is the *classical event log*, which records the execution history of a process as a set of timestamped, case-attributed activity observations [6, 75]. Formally, a classical event log is defined as the tuple

$$\mathcal{L} = (E,\, \mathcal{A},\, \mathcal{C},\, \pi), \tag{3.7}$$

where E is a finite set of events, $\mathcal{A}$ is the activity alphabet, $\mathcal{C}$ is the set of case identifiers (e.g., production order numbers, product serial identifiers), and $\pi : E \to \mathcal{A} \times \mathcal{C} \times \mathbb{T}$ is a total function mapping each event to its activity label, owning case, and occurrence timestamp, with $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ denoting the temporal domain. A *trace* for case $c \in \mathcal{C}$ is the ordered sub-sequence

$$\sigma_c = \langle e_1,\, e_2,\, \ldots,\, e_{|\sigma_c|} \rangle, \quad \text{where } \pi_t(e_i) \leq \pi_t(e_{i+1})\ \forall\, i, \tag{3.8}$$

with π_t denoting the timestamp projection of π. Industrial event logs routinely contain hundreds of thousands of such traces, recording machine state transitions, operator actions, quality inspection checkpoints, and inter-station material handovers, thereby exposing the behavioural complexity intrinsic to production lines, including cycle-time variation, non-productive waiting, and rework loops [30, 31, 76].

Whilst classical event logs are well-suited to single-entity processes, modern manufacturing environments involve rich interactions among multiple co-occurring objects, such as raw-material batches, tooling assets, work orders, and human operators. Forcing such concurrent interactions into a single-case structure causes severe information loss, commonly manifesting as convergence and divergence artefacts in discovered process models [42, 43]. *Object-Centric Event Logs* (OCEL) address this structural limitation by natively mapping each event to a set of simultaneously involved objects. An OCEL is formally defined as the tuple

$$\mathcal{O} = (E,\, \mathcal{A},\, O,\, \mathcal{T}_O,\, \pi_e,\, \pi_o,\, \pi_{\text{type}}), \tag{3.9}$$

where O is the universe of objects, $\mathcal{T}_O$ is the finite set of object types (e.g., *raw material*, *machine*, *operator*, *work order*), $\pi_e : E \to \mathcal{A} \times \mathbb{T}$ assigns each event an activity label and a timestamp, $\pi_o : E \to \mathcal{P}(O)$ maps each event to a non-empty power-set of participating objects, and $\pi_{\text{type}} : O \to \mathcal{T}_O$ assigns a type to every object [43]. The multi-object condition $|\pi_o(e)| \geq 1$ is the defining distinction from the classical log: a single assembly event can simultaneously relate a machine, a material batch, and a production order, preserving the full relational structure of shop-floor interactions without data flattening [20]. This property is particularly consequential in high-throughput production lines, where a defect event on one object type may propagate causally across several co-participating objects, a dependency that would remain invisible in a case-centric representation [46, 48].

In parallel with event-based records, IIoT-equipped production lines generate continuous time-series data streams from physically embedded sensors [10, 77]. A sensor stream from instrument k is formally represented as the ordered sequence

$$\mathbf{s}_k = \left\{ (t_i,\, v_i^{(k)}) \right\}_{i=1}^{N_k}, \tag{3.10}$$

where $t_i \in \mathbb{T}$ is the observation timestamp and $v_i^{(k)} \in \mathbb{R}$ is the measured physical quantity. The inter-sample interval $\delta_k = t_{i+1} - t_i$ governs the temporal resolution of stream k, whilst the empirical mean $\bar{v}^{(k)} = N_k^{-1} \sum_{i=1}^{N_k} v_i^{(k)}$ and sample variance $\sigma_k^2 = (N_k - 1)^{-1} \sum_{i=1}^{N_k} (v_i^{(k)} - \bar{v}^{(k)})^2$ characterise its statistical properties. These streams are indispensable for condition monitoring, anomaly detection, and predictive maintenance, enabling early intervention before unplanned downtime materialises [35, 68, 78]. When aligned temporally with event log timestamps, these sensor streams enable fine-grained energy attribution to individual activities, a requirement that is central to high-resolution sustainability accounting [79].

This capability is formalised within the *Process Mining–Life Cycle Assessment (PM-LCA) integrated dataset*, a multi-dimensional data structure designed to bridge *execution data* what happened on the shop floor and *environmental inventory data* what resources were consumed and what emissions were generated at each process step [8, 36]. Unlike a static LCA inventory that records average industry values [80], a PM-LCA dataset maps measured operational observations directly onto quantitative environmental outcomes, enabling a dynamic, per-product footprint calculation. Each machine $m \in M$ is characterised by a state-specific power profile $\varepsilon_m : S_m \to \mathbb{R}_{\geq 0}$, where $S_m = \{\text{productive, idle, setup}\}$ is its operational state space [12]. The activity-level energy consumption for event e of duration $\Delta t_e = \pi_t(e_{\text{end}}) - \pi_t(e_{\text{start}})$ executing at state $s \in S_m$ is

$$E_e = \varepsilon_m(s) \cdot \Delta t_e, \tag{3.11}$$

from which the idle-time energy waste accumulated by machine m across all cases is

$$E_{\mathrm{idle},m} = \varepsilon_m(\mathrm{idle}) \cdot \sum_{c \in \mathcal{C}} \mathrm{IT}_m^{(c)}, \tag{3.12}$$

where $\mathrm{IT}_m^{(c)}$ is the idle duration of machine m during case c, computed directly from the event-log timestamps of equation (3.7). This explicit linkage between waiting time and wasted energy expenditure means that operational bottlenecks identified by process mining become simultaneously identifiable as environmental inefficiencies [8, 36]. Each process step is further augmented by an *environmental foreground vector*

$$\mathbf{f}_e = \left(m_{\mathrm{raw}},\ E_e,\ w_{\mathrm{water}},\ c_{\mathrm{CO}_2},\ w_{\mathrm{scrap}}\right)^\top \in \mathbb{R}^5_{\geq 0}, \tag{3.13}$$

recording raw-material input, activity energy, water use, CO_2-equivalent emissions, and scrap generation for that individual event [80]. Multiplying $\mathbf{f}_e$ by background emission factors from life cycle inventory databases yields a per-product, path-specific environmental footprint that reflects the actual route a batch took through the factory rather than a static average [13, 81]. This granular attribution enables the identification of *green bottlenecks* process steps that may be time-efficient but are disproportionately carbon-intensive whose detection is impossible without event-level alignment of execution and environmental data [12, 36].

Beyond event logs, sensor streams, and PM-LCA records, industrial datasets include relational and tabular records originating from ERP and MES layers. These records are characterised by the three canonical dimensions of big data: *Volume* (the scale of records accumulated per shift), *Velocity* (the arrival rate of new observations), and *Variety* (the heterogeneity of formats, schemas, and encoding conventions across systems) [66, 82]. The challenge of schema heterogeneity is well illustrated by benchmark datasets drawn from multi-modal industrial processes, where physical sensor readings, process event records, and quality outcome labels originate from entirely disparate acquisition systems and must be harmonised before any unified analysis is feasible [83]. In high-throughput production settings, the combined event, sensor, relational, and foreground environmental streams can accumulate to hundreds of gigabytes per shift, demanding scalable ingestion pipelines and semantic alignment procedures before meaningful analysis can proceed [5, 79].

The fitness of any dataset for process optimisation depends critically on three measurable quality dimensions [6, 84]. *Completeness* quantifies the proportion of expected observations that were actually recorded:

$$\mathcal{Q}_c = 1 - \frac{|E_{\text{missing}}|}{|E_{\text{expected}}|} \in [0, 1], \tag{3.14}$$

where $E_{\text{missing}} = E_{\text{expected}} \setminus E_{\text{observed}}$; a value of $\mathcal{Q}_c = 1$ corresponds to a gap-free log. *Consistency* measures the fraction of records free from cross-source logical or temporal conflicts (e.g., misaligned timestamps between ERP order records and MES machine-start events):

$$\mathcal{Q}_{\text{con}} = \frac{\big|\{e \in E : \neg\,\text{conflict}(e)\}\big|}{|E|} \in [0, 1]. \tag{3.15}$$

Granularity is characterised by the finest achievable temporal resolution across all data sources:

$$\Delta t^* = \min_k \min_i \left\{ t_{i+1}^{(k)} - t_i^{(k)} \right\}. \tag{3.16}$$

High granularity (small Δt^*) enables micro-level machine-state analysis and per-activity energy attribution, whereas coarse granularity constrains analysis to macro-level production tracking and aggregated sustainability estimates [36]. In practice, achieving $\mathcal{Q}_c \approx 1$, $\mathcal{Q}_{\text{con}} \approx 1$, and a sufficiently small Δt^* simultaneously across all data layers is a non-trivial integration challenge, yet it is a necessary prerequisite for constructing reliable analytical models and deriving actionable operational and environmental insights [5, 85].

The remainder of this chapter binds the abstractions established above to the specific manufacturing site that supplies the empirical material of the dissertation. Section 3.2 introduces the TL209 line and its physical and functional architecture; section 3.3 specifies the on-site acquisition pipeline and the cross-source consistency procedure that implements equation (3.15) in practice; section 3.4 describes the Industry 5.0 sensor extension required to populate the social-sustainability pillar in real time; and section 3.5 reports the numerical baseline that the methodology of Chapter 5 subsequently consumes.

3.2 TL209 Case-Study Site and Physical Architecture

The empirical work reported in this dissertation is conducted on a single high-throughput tube manufacturing line, designated TL209, located at the partner facility in Debrecen. The line operates at a nameplate rate of 240 tubes/min and comprises four primary process machines (M1–M4) instrumented with six industrial sensors (S1–S6) and two Work In Process (WIP) accumulators positioned at the inter-machine buffers. Within the four-machine sequence, four bottleneck stations, denoted S202–S205 (extruder, capping machine, printing unit, and packaging machine), are the principal subjects of the diagnostic, optimisation, and ranking analyses presented in Chapters 5 and 6. The physical layout of the line is shown in Figure 3.1, the functional flow and sensing architecture in Figure 3.2, and the symbolic key for the functional diagram in Figure 3.3.

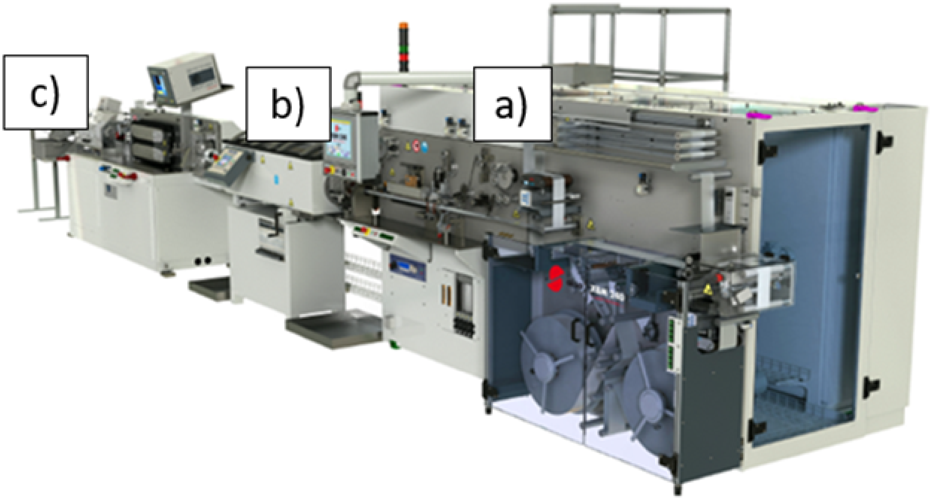


Figure 3.1: Physical layout illustration of the TL209 tube manufacturing line (Stations 202–209).

The physical layout in Figure 3.1 is complemented by the functional flow mapping shown in Figure 3.2. This mapping illustrates the logical sequence of the four primary machines (M1–M4), the location of the six industrial sensors (S1–S6) used for event generation, and the position of the Work In Process (WIP) accumulators. The structural key for these functional symbols is provided in Figure 3.3.

With the physical and functional architecture established, the next section

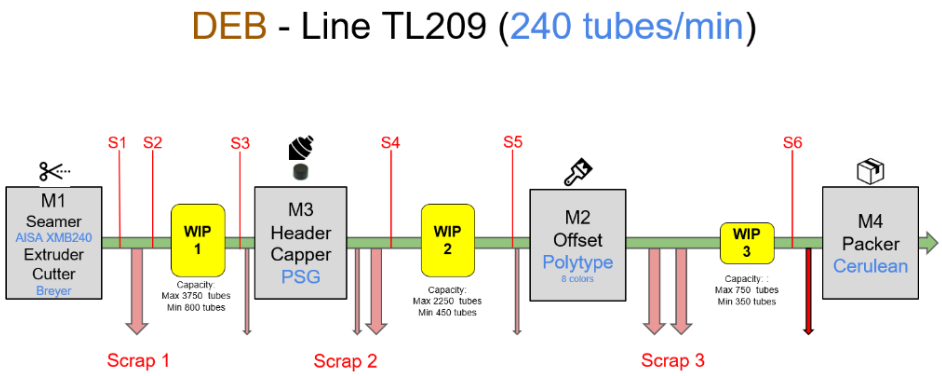


Figure 3.2: Functional flow diagram of the DEB – Line TL209 (240 tubes/min), showing machine sequence and sensing architecture (S1–S6).

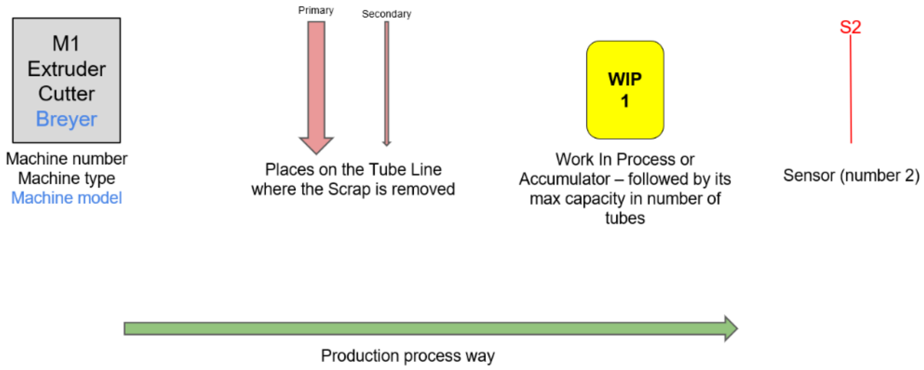


Figure 3.3: Key/Legend for the TL209 functional diagram, identifying machine identifiers, scrap removal points, and Work In Process (WIP) capacity parameters.

describes the on-site acquisition pipeline that converts the telemetry generated by the four machines and six sensors of Figure 3.2 into the OCEL 2.0 event log that the methodology of Chapter 5 subsequently consumes.

3.3 On-Site Acquisition Pipeline and Quality Assurance

The framework ingests data from three systems: the MES records timestamped state changes and process sequences; six IIoT sensors (S1-S6) capture physical measurements and power draw at 1 Hz; and the ERP system provides operational context through production schedules and material specifications. Integrating these independent sources requires temporal alignment and schema harmonization before the analytical pipeline begins.

The process begins with the continuous collection of relevant data from the manufacturing environment (S1, S2). This involves the following:

- Event Log Generation: Capturing timestamped events from MES or machine controllers, detailing activities performed, the associated case identifier (e.g., production order ID) and the resource/machine involved. This forms the basis for process mining. Following the formal tuple definition of the classical event log introduced in equation (3.7), the log is represented operationally as:

  $$L = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}, \quad \sigma_i = (e_1, e_2, \ldots, e_{m_i}) \tag{3.17}$$

  where L is the event log, σ_i is a trace (case) in the sense of equation (3.8), and e_j are events within the trace. In practice, a state-change detection algorithm converts the continuous time-series into a structured event log by generating one event record each time a machine transitions between operational states (e.g., running → idle, idle → fault) or a continuous sensor reading crosses a predefined operational threshold. The resulting log is structured in conformance with the Object-Centric Event Log (OCEL) standard, forming the direct input to the process mining module of Chapter 5 [20, 43].

- Sensor Data Collection: Gathering real-time readings from sensors (e.g., power meters, temperature sensors). Dedicated power meters installed on the main machine stations log energy consumption continuously. This telemetry is aligned temporally with event log timestamps to enable activity-level energy attribution in the dynamic life cycle assessment step of the methodology. The 1 Hz sampling rate provides a temporal resolution of $\Delta t^* = 1$ s, which exceeds the minimum resolution required for machine-state-specific power profiling in high-speed tube manufacturing contexts [12].

- Contextual Data Integration: Accessing relevant data from ERP or planning systems (e.g., schedules, material specifications, emission factors).

Prior to entering the process mining pipeline, the integrated dataset undergoes a data quality assurance pass. The completeness metric $\mathcal{Q}_c$ and the consistency metric $\mathcal{Q}_{\text{con}}$ introduced in section 3.1 are computed over the combined MES IIoT ERP dataset. Records exhibiting cross-source timestamp conflicts exceeding 60 seconds between the MES event timestamp and the corresponding IIoT sensor record are flagged as temporally inconsistent and removed from the analytical sample before any downstream computation is performed. This threshold reflects the maximum state-change dwell time that can be unambiguously attributed to a single operational state at the 1 Hz sensor sampling resolution; at 1 Hz, a 60-second window accumulates 60 consecutive readings per machine state, providing a statistically stable basis for energy attribution rather than relying on a single instantaneous measurement.

The pipeline described above produces the operational and environmental record that the framework of Chapter 5 consumes. To populate the social-sustainability pillar introduced in section 2.3–2.3.1 of Chapter 2 with the same activity-instance granularity, the MES, IIoT and ERP stack is augmented with a compact wearable-and-embedded sensor extension whose minimum specification is described next.

3.4 Industry 5.0 Sensor Extension: Minimum Viable Sensor Suite

Populating the Ergonomic Risk Index defined in Chapter 5, equation (5.11) in real time requires a compact extension of the MES and IIoT stack that the acquisition pipeline of section 3.3 already ingests. Following [17–19], the Minimum Viable Sensor Suite (MVSS) adopted for the Industry 5.0 extension of the framework is:

- **Inertial measurement units (IMUs).** Body-worn 9-DoF IMUs (one per upper limb, one trunk) sampled at ≥ 50 Hz supply the joint-angle stream from which the RULA action score is computed at the current pose. This channel replaces the manual observation sheet of the original RULA protocol with a continuous sensor feed.

- **Surface electromyography (EMG).** A two-channel sEMG patch on the dominant forearm and trapezius supplies the muscle-activation envelope from which the OCRA frequency and force multipliers are estimated, together with a shift-level fatigue signal correlated with sustained isometric load.

- **Instrumented load cells.** A floor-embedded load-cell mat (or the instrumented tool handle where applicable) supplies the vertical-force and centreof-pressure streams required by the revised NIOSH equation for the lifting index.

- **Optional vision channel.** A single RGB or RGB-D camera with a modern 2-D pose-estimation backbone can supplement or substitute for the IMU channel in workstations where occlusion-free line of sight is available, reducing the operator-worn hardware at the cost of a vision-processing pipeline on the cloud platform.

All four streams are timestamp-aligned with the OCEL event log through the same 60 s cross-source consistency filter defined in section 3.3, so that each ergonomic-index value is attributed to exactly one event (i, j, k) and the activity-instance granularity of the optimisation objectives in Chapter 5 is preserved. The MVSS is additive with respect to the existing MES, IIoT, and ERP stack: no modification of the production control system is required, and the ergonomic-index calculators run on the cloud platform alongside the existing process-mining and dLCA modules.

With the operational, environmental, and social acquisition stacks specified, the next section reports the numerical inventory that results from applying the pipeline of section 3.3–3.4 to the eight-week observation window on the TL209 line. This inventory constitutes the empirical input to every diagnostic, optimisation, and ranking computation in Chapter 6.

3.5 Numerical Inventory and Baseline Snapshot

3.5.1 Hourly Telemetry Record Snapshot

The acquisition pipeline of section 3.3 persists the four operational and environmental streams as an hourly per-station telemetry record covering the four bottleneck stations of TL209 over the 2021–2024 observation horizon. Each row is a single $(\text{date}, \text{hour}, \text{station})$ tuple and carries thirteen attributes: the calendar key $(\texttt{date}, \texttt{hour})$, the station identifier $\texttt{stn} \in \{S202, \ldots, S205\}$, the Boolean evening-peak flag `peak` that tracks the $[16{:}00, 20{:}00)$ window of the Hungarian grid (section 3.6), the state-resident minutes $(\texttt{run\_min}, \texttt{idle\_min})$ whose sum is bounded by the 60 s consistency filter of section 3.3, the hourly energy draw `ec_kwh`, the produced and rejected unit counts $(\texttt{units}, \texttt{defects})$, the defect-derived waste mass `wg_kg`, the grid emission factor `ef` applied to that hour under equation (5.6), the attributed CO_2 mass $\texttt{ce\_kg} = \texttt{ec\_kwh} \cdot \texttt{ef}$, and the composite ergonomic risk index `eri_hour` defined in section 5.1.5. The complete record contains 140,256 row-hours (four stations × 24 hours × 1,461 calendar days of the 2021–2024 window) and is the raw input to the OCEL transformation reported in section 3.5.2. Table 3.1 shows a representative eight-row extract that contrasts an off-peak hour (hour 0, $\varepsilon_{\text{off}} = 0.200$) with the evening-peak hour 17 ($\varepsilon_{\text{peak}} = 0.275$) on a single calendar day.

Table 3.1: Representative extract from the hourly per-station telemetry record of line TL209 for 2022-06-15, showing the four bottleneck stations S202–S205 in one off-peak hour (hour 0) and the evening-peak hour 17. Column semantics follow section 3.1: `run`/`idle` in minutes per hour, `ec` in kWh, `wg` and `ce` in kg, `ef` in kg CO_2/kWh, `eri` dimensionless. The step from $\texttt{ef} = 0.200$ to $\texttt{ef} = 0.275$ across the peak boundary is the hourly manifestation of the two-window grid factor of section 3.6.

date	**hr**	**stn**	**peak**	**run**	**idle**	**ec**	**units**	**def**	**wg**	**ef**	**ce**	**eri**
2022-06-15	0	S202	off	11.79	48.21	9.701	2.300	0.064	0.372	0.200	1.940	0.084
2022-06-15	0	S203	off	11.19	48.81	9.541	2.477	0.058	0.338	0.200	1.908	0.094
2022-06-15	0	S204	off	10.55	49.45	8.638	1.993	0.070	0.393	0.200	1.728	0.071
2022-06-15	0	S205	off	11.80	48.20	9.572	2.437	0.069	0.373	0.200	1.914	0.092
2022-06-15	17	S202	peak	49.62	10.38	40.816	9.678	0.269	1.565	0.275	11.224	0.354
2022-06-15	17	S203	peak	46.13	13.87	39.327	10.210	0.237	1.393	0.275	10.815	0.386
2022-06-15	17	S204	peak	44.24	15.76	36.216	8.357	0.293	1.648	0.275	9.959	0.296
2022-06-15	17	S205	peak	44.28	15.72	35.922	9.144	0.259	1.400	0.275	9.879	0.345

Three invariants of the record are visible in Table 3.1. First, the state-resident minutes satisfy $\texttt{run} + \texttt{idle} = 60$ s per hour, which is the closure condition enforced by the consistency filter of section 3.3. Second, the attributed carbon $\texttt{ce} = \texttt{ec} \cdot \texttt{ef}$ holds to four decimal places on every row, so the record is consistent with the dLCA attribution equation (5.5) at the hourly granularity. Third, the energy draw at hour 17 is approximately 4× the draw at hour 0, reflecting the intra-day run/idle shift rather than any parameter change at the machines themselves, whereas `ef` changes discretely at the peak boundary; the two effects compound in `ce`, producing the $\approx 6\times$ carbon differential per hour that motivates the time-resolved attribution framework adopted throughout the dissertation.

3.5.2 Event Log and Data Transformation

The framework was validated using historical and real-time data collected from sensors and MES logs within the tube manufacturing facility described in section 3.2.

The data encompassed operational parameters (cycle times, throughput, idle times, defect rates) and environmental metrics (energy consumption, derived emissions, waste generation).

Raw sensor data (32,257 rows of hourly snapshots across 123 columns) were transformed into a granular event log (S1) by the state-change detection algorithm of section 3.3. An event was generated each time a machine's status sensor changed state (e.g., from *idle* to *running*); for continuous sensors, an event was recorded when the reading crossed a pre-defined threshold. This produced a structured log of 391,127 discrete events (S2), each containing a CaseID, Activity, Timestamp, Resource, and Event Type; a representative extract is shown in Table 3.2.

Table 3.2: Event log sample (first 10 events).

CaseID	**Activity**	**Timestamp**	**Resource**	**Event Type**	**Value**
Station_207	Sensor_6_Start	1/4/2021 6:00	S6_207	Start	1
Station_205	Sensor_4_Start	1/4/2021 7:00	S4_205	Start	1
Station_207	Sensor_6_Complete	1/4/2021 7:00	S6_207	Complete	0
Station_203	Sensor_4_Start	1/4/2021 8:00	S4_203	Start	1
Station_207	Sensor_7_Start	1/4/2021 8:00	S7_207	Start	1
Station_202	Sensor_7_Start	1/4/2021 9:00	S7_202	Start	107
Station_203	Sensor_4_Complete	1/4/2021 9:00	S4_203	Complete	0
Station_205	Sensor_1_Start	1/4/2021 9:00	S1_205	Start	78
Station_205	Sensor_2_Start	1/4/2021 9:00	S2_205	Start	2
Station_205	Sensor_4_Complete	1/4/2021 9:00	S4_205	Complete	0

3.5.3 Baseline Machine Performance and Operational Constraints

Table 3.3 presents the eight key performance indicators recorded at each bottleneck station during the pre-optimisation observation period. These values define the starting point (P_{base}) for all subsequent optimisation and comparison work. The system totals are: 4,015 kWh/day energy, 863.2 kg CO_2/day emissions, and 146.5 kg/day waste; the system-average Ergonomic Risk Index is $\overline{ERI} = 0.450$ (the Industry 5.0 social-sustainability pillar introduced in Chapter 5). The carbon total is reported under the Hungarian LCA grid intensity of Anita et al. [22] ($\varepsilon_{\text{peak}} = 0.275$, $\varepsilon_{\text{off}} = 0.200$ kg CO_2/kWh, evening peak window [16:00, 20:00)); the corresponding site-average intensity is 0.215 kg CO_2/kWh, within the 0.185–0.275 envelope of the reference study.

Table 3.3: Baseline operational and environmental performance of the bottleneck stations (TL209). The final column (ERI) is the Industry 5.0 social-sustainability addition (Chapter 5, equation (5.11)); per-station values are computed as $ERI_{\text{base},i} \propto (1/CT_i)(1 + DR_i/100)$ and anchored so that $\overline{ERI} = 0.450$.

Machine / Station	**CT (s)**	**IT (s)**	**TP (units/day)**	**DR (%)**	**EC (kWh/day)**	**CE (kg CO_2/day)**	**WG (kg/day)**	**ERI (–)**
Extruder (S202)	750	95	115	2.8	485	104.3	18.5	0.428
Capping Machine (S203)	636	82	135	2.3	520	111.8	19.8	0.502
Printing Unit (S204)	804	124	108	3.5	468	100.6	21.2	0.402
Packaging Machine (S205)	684	98	126	2.6	495	106.4	20.5	0.468
System total			**484**		**4,015**	**863.2**	**146.5**	**0.450**†

† Arithmetic mean across S202–S205 (not a sum); anchors the Ciccarelli-2025 RULA, OCRA and NIOSH composite to $\overline{ERI} = 0.450$.

Table 3.4 defines the operational and environmental boundary conditions enforced throughout the optimisation. Every solution on the Pareto front satisfies all constraints listed here.

Table 3.4: Operational and environmental constraints applied during optimisation.

Constraint group	**Parameter**	**Value**
Operational	Machine utilisation limit	95%
	Minimum throughput	94 units/day
	Maximum idle time	15%
	Station availability	98%
Environmental	Maximum energy consumption (EC)	600 kWh/day per machine
	Maximum carbon emissions (CE)	300 kg CO_2/day per machine
	Maximum waste generation (WG)	24 kg/day per machine
Data Management	Data transfer latency	250 ms
	Sensor sampling rate	1 Hz minimum
	Historical retention	5 years

The three tabulated environmental limits implement the admissibility constraints in equations (5.18)–(5.20). Energy and carbon match the central framework paper [27] and the feasibility filter for the Pareto front (Section 6.5.1): 600 kWh/day and 300 kg CO_2/day per machine, reflecting worst-case scenarios observed at the bottleneck segment with operational margin relative to the peak daily baselines in Table 3.3 (520 kWh/day and 260 kg CO_2/day, respectively). The waste cap extends the same proportional uplift as for energy (600/520) to the highest observed bottleneck waste generation (21.2 kg/day at the printing unit), i.e. $21.2 \times (600/520) \approx 24.5$ kg/day, rounded to 24 kg/day per machine for the dissertation implementation tabulated here.

3.6 Hungarian Grid Emission Factors

The dynamic life-cycle assessment module of Chapter 5 attributes energy draw to carbon emissions through the two-window Hungarian grid intensity of Anita et al. [22]: an off-peak factor of $\varepsilon_{\text{off}} = 0.200$ kg CO_2/kWh for the twenty hours outside the evening-peak window, and a peak factor of $\varepsilon_{\text{peak}} = 0.275$ kg CO_2/kWh during the four-hour evening peak [16:00, 20:00). The 37.5% difference between these two factors means that the same energy draw at Station 202 generates substantially different CO_2 attributions depending solely on the clock time of the event. A static LCA model using the site-averaged annual intensity of $\bar{EF} = 863.2/4015 \approx 0.215$ kg CO_2/kWh would systematically under-report evening-peak emissions by approximately 28% and over-report all other hours by approximately 7.5%; the under-reporting error is structurally larger because the peak window coincides with the end-of-shift ramp that concentrates line-pacing decisions and the irreducible stand-by load on the bottleneck cells. Table 3.5 summarises the published Hungarian grid-EF record used to parameterise this window.

Table 3.5: Published Hungarian grid emission-factor (EF) sources, most recent to oldest, together with the envelope adopted for the dLCA attribution of equation (5.6). The Anita et al. (2025) LCA envelope is the one retained for the baseline calibration of Table 3.3 and for all Pareto-front evaluations in Chapter 6 (Thesis III).

Source	**Year**	**Value (kg CO_2/kWh)**	**Resolution**	**Basis / notes**
Anita et al. [22]†	2024	0.140–0.240 (direct) 0.185–0.275 (LCA)	Hourly, with 16:00–20:00 peak	Operational + LCA; ENTSO-E hourly flow-tracing; peer-reviewed (WEVJ 16(4):240)
MAVIR RED III series [86]	2023	0.19–0.26	Hourly, attributional	Hungarian TSO, Directive 2018/2001 compliance series
EEA country intensity [87]	2022	0.181	Annual, location-based	European Environment Agency; single-value baseline
Ember monthly review [88]	2023	0.20	Monthly, operational	Flow-tracing against imports; fuel-mix decomposition

† Adopted as the dLCA parameterisation. The two-level piecewise form of equation (5.6) anchors $\varepsilon_{\text{peak}}$ and ε_{off} to the upper and lower bounds of the LCA envelope in [22], so that the carbon total 863.2 kg CO_2/day is consistent with the Hungarian grid record and the ±30% sensitivity band examined in Chapter 6 encloses every other entry in the table.

3.7 NSGA-II Optimisation Configuration

To ensure the reported Pareto front meets the required academic rigour, the NSGA-II solver applied in Chapter 6 is executed using the population and computation parameters specified in Table 3.6. These settings are chosen so that the search space is explored with enough density to identify the non-dominated configurations (P_1–P_{18}) while maintaining the 95% utilisation constraint mandatory for industrial feasibility; the values are fixed input constants to the analysis and therefore belong with the dataset description rather than with the analytical results that they produce.

Table 3.6: NSGA-II computational hyperparameters and justification.

Parameter	**Specification**	**Justification**
Solver Algorithm	NSGA-II	Standard for multi-objective non-dominated sorting logic.
Population Size (N)	100 individuals	Provides global coverage across the 9-station decision space.
Max Generations (G)	200 generations	Achieves convergence stability of $\pm 10^{-4}$.
Crossover Rate (P_c)	0.9 (SBX)	Required for global search across bottleneck configurations.
Mutation Rate (P_m)	0.1 (Polynomial)	Prevents premature convergence near the 95% utilisation limit.
Selection Logic	Tournament	Based on Pareto Rank and Crowding Distance diversity.

The numerical inventory consolidated in this section, comprising 32,257 raw sensor records, 391,127 OCEL state-change events, the per-station baseline of Table 3.3, and the feasibility envelope of Table 3.4, is the empirical input to every diagnostic, optimisation, and ranking computation reported in Chapter 6. The criterion-weight vector that consumes these data inside the NSGA-II objectives and the GRA–TOPSIS ranking is derived in Chapter 6 (section 6.1) from the same baseline through the hybrid Fuzzy AHP and Shannon Entropy procedure formalised in Chapter 5.

Chapter 4: Critical Review and Research Gap Analysis

Chapter 2 surveyed the main bodies of knowledge relevant to this dissertation. This chapter converts that survey into a critical analytical position by identifying what remains methodologically unresolved when operational efficiency and environmental sustainability must be improved simultaneously.

4.1 Critical Synthesis of the Reviewed Studies

Process mining is well established for reconstructing actual workflows from event logs, identifying bottlenecks, and evaluating conformance between observed and normative behaviour [6, 44, 89]. Its critical limitation, however, is that even when resource utilisation is included, the analysis remains focused on time, frequency, and compliance: carbon emissions, energy intensity, and waste generation are rarely attached to process instances with the same granularity as operational metrics [8, 31]. Process mining can reveal where delays occur, but not how those delays co-vary with environmental burdens in real time.

Life Cycle Assessment provides the internationally standardised basis for quantifying environmental burdens across processes and products [4, 49]. Its limitation is temporal and operational rather than conceptual: most LCA applications in manufacturing remain static and retrospective, relying on averages and batch-level aggregates that are poorly matched to fast-changing production conditions [50, 51]. Conventional LCA may identify a hotspot at the process-stage level but cannot explain how a specific machine setting or delay pattern altered environmental impact during actual execution.

Multi-objective optimisation is the correct mathematical response once efficiency and sustainability are treated as simultaneously important but conflicting goals, with NSGA-II providing strong evidence that Pareto-based search outperforms single-objective methods across manufacturing trade-offs [9, 52, 54]. The limitation lies in the upstream data architecture: objectives are often computed from historical or scenario-based inputs rather than live process measurements [90], and Pareto generation alone is insufficient for deployment without a systematic post-Pareto ranking step to translate a frontier of options into a single implementable recommendation [55, 91].

The literature therefore reveals a consistent pattern: process mining explains operational behaviour but not environmental impact at instance level; LCA explains environmental impact but not at event-level operational resolution; and MOO can balance conflicting objectives only if fed by timely, trustworthy inputs. What remains unaddressed is a closed-loop integration of all three in a validated industrial setting [8, 36].

4.2 Refined Research Questions

1. How effectively can process-mining-based workflow recovery represent the actual operational behaviour of a complex multi-stage manufacturing system?
2. How can operational and environmental metrics be integrated into a unified optimisation model without losing the specificity of either domain?
3. To what extent can the integrated framework generate and rank feasible trade-off solutions for simultaneous efficiency and sustainability improvement?

Chapter 5: Proposed Mathematical Method and Framework

This chapter presents the **Integrated Process and sustainability mining framework, Multi-Objective Optimization Model for Operational and Environmental Performance** (**IPSMF-MOOM-OEP**). This methodology integrates process mining, dynamic life cycle assessment (dLCA), and multi-objective optimisation to enhance TL209 production line performance [27]. The framework is demonstrated using a representative high-throughput week from the TL209 manufacturing line to generate step-by-step computations and visualisations [29, 30].

5.1 Framework Architecture

The framework utilizes a cloud platform to manage the high computational demands of the analytical pipeline. Concurrently executing process discovery, dynamic LCA attribution, and NSGA-II populations requires elastic resource scaling that exceeds the capacity of local edge controllers. This cloud-based approach ensures near real-time processing without latency between data acquisition and analytical output [10, 11].

The framework adopts a three-layer cloud architecture. The *data layer* ingests ERP, MES event streams and IIoT sensor telemetry from TL209 via a message queuing service, buffering incoming records and enforcing the OCEL 2.0 schema [20, 45] together with the 60 s consistency filter of equation (3.15) before passing them to the processing layer. The *processing layer* hosts the four analytical modules: the object-centric process mining engine (Inductive Miner with performance and conformance analysis on OCEL 2.0 logs), the dynamic LCA inventory compiler driven by the two-window Hungarian grid factor of Anita et al. [22], the Ergonomic Risk Index compiler that fuses the RULA, OCRA, and NIOSH-LI channels of the minimum viable sensor suite, and the feasibility-constrained NSGA-II optimisation solver together with the hybrid Fuzzy AHP and Shannon Entropy weight-derivation routines feeding a GRA–TOPSIS ranking stage. The *presentation layer* delivers the rank-1 recommendation, the full Pareto-front table, and operational monitoring dashboards to production engineers, and writes the top-ranked configuration parameters back to a monitoring interface accessible to the MES. This three-layer decomposition mirrors the cloud manufacturing reference architectures documented in the literature, which consistently distinguish data infrastructure, analytical computation, and user-facing service tiers as functionally distinct concerns [11].

The overall architecture, depicted conceptually in Figure 5.1, establishes a systematic flow from real-time data acquisition to optimised parameter recommendations. The cloud platform is crucial for the following:

- Data Ingestion and Storage: Handling high-speed, high-volume data streams from numerous sensors and manufacturing systems (ERP, MES) across potentially many production lines [92].
- Scalable Computation: Providing on-demand computational resources required for demanding tasks such as process discovery algorithms (e.g., Inductive Miner), real-time LCA calculations, and complex multi-objective optimisation solvers (NSGA-II) [93]. The efficient utilisation and optimisation of these cloud resources are themselves critical aspects, often addressed through exploratory data analysis and machine learning techniques [94].
- Integration and Centralisation: Acting as a central hub to integrate diverse data sources and make processed information and optimisation results accessible to relevant stakeholders and downstream control systems.

5.1.1 Research and Modelling Assumptions

The mathematical formulation that follows in Steps 1–7 rests on a set of research and modelling assumptions that bound the scope of the framework and are stated explicitly here. These assumptions are the conditions under which the MCDM layer (Fuzzy AHP and Shannon Entropy for weighting, and GRA–TOPSIS for ranking) and the NSGA-II solver produce results that are internally consistent with the observed TL209 operational regime. They apply jointly to Chapters 5 and 6.

- **A1 – Event-log completeness.** The OCEL 2.0 event log L recovered by the ingestion layer is assumed to capture every machine state transition over the

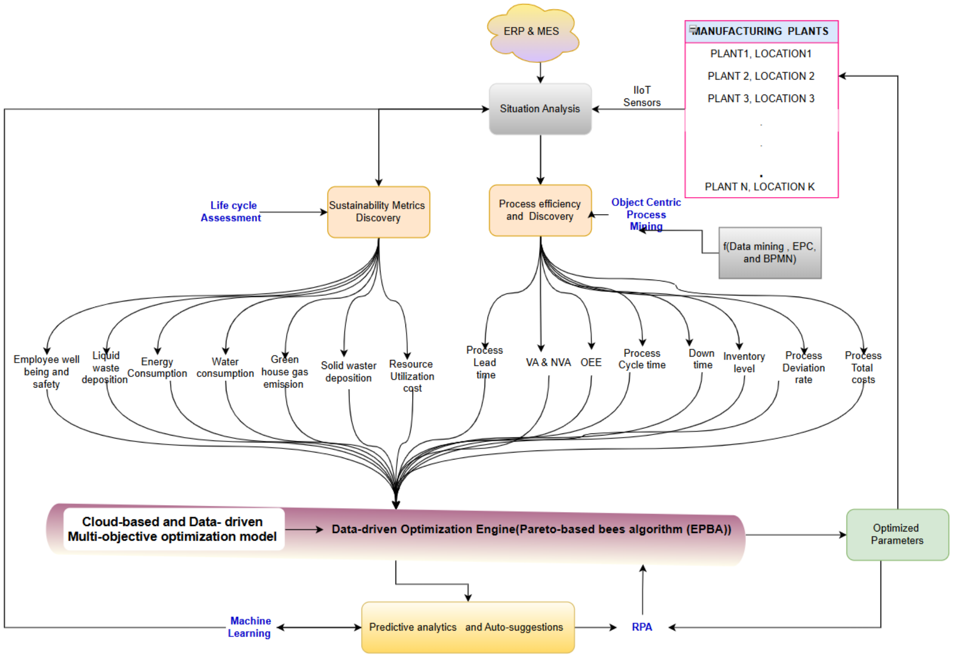


Figure 5.1: Architecture of the cloud-based IPSMF-MOOM-OEP framework.

observation window. Records excluded under the 60 s cross-source consistency filter (Step 1) are treated as missing-at-random with respect to the decision variables.

- **A2 – Stationary process regime.** The underlying process structure (routing graph, rework topology, scrap-diversion logic) is assumed to be stationary across the eight-week observation horizon. Short-horizon disruptions (tool changes, unplanned stoppages, supply interruptions) are absorbed into the idle-time and defect-rate distributions rather than modelled as exogenous shocks.

- **A3 – Machine-state power additivity.** Station-level energy consumption is assumed to be the time integral of the machine-state-specific power profile $\varepsilon_m(s)$ over the event duration (equation (5.3)); power cross-coupling between adjacent stations via the shared bus bar is treated as a bounded second-order effect absorbed into the ±2% measurement uncertainty of the EC signal.

- **A4 – Hybrid foreground–background inventory and time-resolved grid emission factor.** The dLCA inventory follows the hybrid foreground–background convention of ISO 14044:2006, section 4.3.1 [95]: foreground process data – activity-level power telemetry $P_i(t)$, event timings, and machine states – are primary data collected on site at the case-study facility over the eight-week observation window, whereas the grid emission factor $EF(t)$ is a secondary, time-matched, literature-calibrated coefficient adopted by the researcher from the published European grid-carbon literature because the facility itself does not meter grid-mix composition. For computational tractability $EF(t)$ is represented by the two-window piecewise form of equation (5.6), which captures the structural peak/off-peak split of the Hungarian grid mix; finer hourly variation is treated as a bounded residual absorbed into the ±30% sensitivity band of $\varepsilon_{\text{peak}}$ and ε_{off} reported in Chapter 6.

- **A5 – Decision-maker linguistic consistency.** The expert pairwise judgements feeding the Fuzzy AHP pairwise matrix (Step 5b) are assumed to be linguistically consistent in the sense of Chang [24]: the triangular-fuzzy-number spread $(m-1, m, m+1)$ captures the decision maker's uncertainty around each crisp judgement m, and the resulting pairwise matrix satisfies the fuzzy consistency ratio $\widetilde{CR} < 0.10$.

- **A6 – Objective-agnostic NSGA-II and post-hoc MCDM on a fixed non-dominated set.** The 18-solution non-dominated front returned by NSGA-II at generation $G_{\max} = 200$ is assumed to approximate the true Pareto frontier within hypervolume tolerance $\tau_{HV} \leq 1\%$. The NSGA-II evolution is objective-agnostic in the MCDM sense: the vector objectives (f_1, g_2) enter the non-domination sort without reference to the hybrid FAHP and Entropy weight vector $\mathbf{w}^{\star}$, so that the Pareto front is a structural property of the decision variables and the dLCA inventory rather than of managerial preference. The MCDM layer (FAHP pairwise extent analysis, Shannon-entropy weighting, and GRA–TOPSIS ranking) is applied to this discrete front as a fixed alternative set; no re-evolution of the population is performed during the ranking stage. Consequently, alternative weight vectors arising from the λ-sweep or the β-sensitivity analysis reorder the ranking of the 18 solutions but leave the non-dominated front itself invariant.

Assumptions A1–A4 scope the data pipeline and the dLCA inventory; A5–A6 scope the decision-making layer. The four-tier sensitivity analysis of Section 6.6.5 and the

λ-ζ landscape reported in Chapter 6 quantify the sensitivity of the recommended configuration to the most load-bearing of these assumptions, namely A5 (fuzzy-judgement spread) and A6 (Pareto-front stability).

5.1.2 Step 1: Data Acquisition and Monitoring

Step 1 of the framework consumes the on-site acquisition pipeline described in Chapter 3, section 3.3, together with the Industry 5.0 sensor extension of section 3.4. To avoid duplication, the MES, six IIoT sensors and ERP instrumentation stack, the 60 s cross-source consistency filter that implements equation (3.15), and the state-change detection rule that converts continuous telemetry into the OCEL 2.0 event log L of equation (3.17) are specified in Chapter 3 and are not restated here. Step 1 takes the resulting OCEL 2.0 log L, the activity-level power telemetry $P_i(t)$, and the ERP context records as input and forwards them unchanged to Steps 2 (process mining) and 3 (dynamic LCA inventory) of the methodology.

5.1.3 Step 2: Process Mining for Operational Insight

The choice of discovery algorithm for this step is consequential because different algorithms make different guarantees about the structural properties of the models they produce. The Alpha Miner, while foundational, is known to fail on event logs containing short loops and non-free-choice constructs, both of which are present in TL209's rework topology (Station 205 → 204). The Heuristic Miner and Fuzzy Miner can handle noise-laden logs but produce models that are not guaranteed to be sound workflow nets: they may contain deadlocks or livelocks that render the discovered model unexecutable and therefore analytically unreliable. In contrast to these alternatives, the Inductive Miner is selected here because it guarantees the soundness of every discovered process model by construction, ensures the absence of deadlocks and livelocks, and produces hierarchically decomposed process trees that are directly interpretable by production engineers without post-processing. Critically, the Inductive Miner detects concurrency and loop structures from the event log without introducing false-positive noise edges, making it well-suited to the TL209 context in which cross-transitions (e.g., Station 202 → 205) and rework loops coexist within a generally sequential flow [6, 31].

The event logs collected (L) are fed into the process mining module (S3, S4, S5)

[31]:

- Process Discovery: The Inductive Miner algorithm generates a process model (e.g., Petri Net or BPMN) that represents the actual workflow (see Figure 6.2 in Section 6.2.2).
- Performance Analysis: Operational KPIs are calculated from the event data:
  - Cycle Time (CT): measures the elapsed time from the start to the completion of a single production order at one station. Its variability across stations directly reveals scheduling imbalances: a station whose CT distribution has a high coefficient of variation is a candidate source of flow instability regardless of whether it is the absolute bottleneck.
  - Waiting Time/Idle Time (IT): measures non-productive time caused by upstream blockage or downstream starvation. The accumulation of IT at bottleneck stations is directly correlated with idle-state energy waste, as formalised in the idle energy equation in Chapter 2: a station that waits for upstream supply continues to draw standby power without contributing to throughput. Bottlenecks identified via the following:

$$\text{Bottleneck}_{ij} = \arg\max_{i,j} \text{Waiting Time}_{ij} \tag{5.1}$$

  - Throughput (TP): measures completed production orders per unit time. Improving throughput requires targeted intervention at the constraining station rather than uniform parameter adjustment across the line.
  - Resource Utilisation (U_i) for resource i:

$$U_i = \frac{T_{\text{active},i}}{T_{\text{total},i}} \tag{5.2}$$

    where $T_{\text{active},i}$ is active time and $T_{\text{total},i}$ is total time. U_i identifies stations operating at near-capacity utilisation, which carry a high risk of cascading overload, as well as stations with available slack that could absorb additional work orders without degrading flow.
- Conformance Checking (Optional): Comparing logs against normative models [96].

The Petri Net produced by the Inductive Miner in this step performs a structural function beyond performance diagnosis: it provides the process-structural context that Step 3 requires to attribute environmental impacts to specific activity types, and defines the topology of the decision variable space that Step 4 uses to construct the optimisation problem. The discovered model is therefore not a visualisation artefact but an analytical input to the subsequent pipeline stages [27].

5.1.4 Step 3: Dynamic Life Cycle Assessment (LCA) Inventory

The dynamic LCA module applied in this step departs from conventional static LCA in a theoretically significant way. Static LCA applies industry-average emission factors uniformly across an entire production run, assigning the same environmental burden per unit to every tube regardless of whether it was produced during a fully loaded running state or a partial-capacity idle state. The dynamic LCA module instead uses machine-state-specific power profiles $\varepsilon_m(s)$, defined in Chapter 2, to attribute energy consumption to each individual event recorded in the OCEL. Because each event in the log carries a precise start and end timestamp, the power meter telemetry from the 1 Hz sensor stream can be integrated over exactly the duration of that event, yielding an energy figure that reflects the actual operational state of the machine during that specific activity instance. This allows the framework to distinguish, for example, between energy consumed during productive tube processing and energy wasted during idle periods at bottleneck stations, producing an instance-level environmental inventory rather than an annual aggregate total [12, 21]. The empirical significance of this distinction is substantial: a spatiotemporal dynamic LCA study found temporal variation of 23–38% and spatial variation of up to 76% in global warming potential for identical production designs evaluated under different conditions [3], confirming that a static industry-average figure cannot represent the environmental burden of a line that operates across multiple machine states within a single production shift.

The framework performs a dynamic LCA inventory analysis using real-time data to quantify environmental burdens at the process-instance level (S6, S7). This step moves beyond traditional static LCA by linking live data to environmental impacts, forming a basis for subsequent optimisation. The analysis includes the following:

- Energy Consumption (EC): Calculated by integrating power readings $P_i(t)$ over the event duration (extending the machine-state energy formulation of equations (3.11)–(3.12) in Chapter 2 to the full sensor-integrated form):

$$EC_{\text{instance/period}} = \sum_i \int_{t_{\text{start},i}}^{t_{\text{end},i}} P_i(t)\, dt \tag{5.3}$$

Equation (5.3) retains the time-additive form of delivered electrical energy: one kilowatt-hour is one kilowatt-hour regardless of the clock, which is the correct accounting unit for sub-metering, billing compliance, and the operational KPI reported to plant management. The *value* of that kilowatt-hour to the environmental objective, however, does vary with the clock, because the Hungarian grid mix shifts diurnally between a nuclear-and-hydro-heavy off-peak regime and a fossil-and-import-heavy peak regime. Two complementary time-resolved factors therefore act on the same $P_i(t)$ stream: [12, 92]

$$EC^{\star}_{\text{instance/period}} = \sum_i \int_{t_{\text{start},i}}^{t_{\text{end},i}} P_i(t)\, \pi(t)\, dt, \tag{5.4}$$

where $\pi(t)$ is a piecewise-constant site-specific primary-energy and time-of-use weighting factor taking the value π_{peak} during the Hungarian grid evening-peak window (16:00–20:00) and π_{off} outside this window, parameterised from the same Hungarian grid-mix and tariff schedule that calibrates $EF(t)$. Setting $\pi(t) \equiv 1$ recovers the physical metering formulation of equation (5.3); selecting a primary-energy factor (e.g. $\pi_{\text{peak}} = 2.5$, $\pi_{\text{off}} = 2.1$ following the EN 15603 / EU conversion convention [97]) yields a primary-energy consumption metric that is dimensionally consistent with the carbon integral of equation (5.5); selecting the retail tariff schedule yields a peak-shaving cost surrogate. The environmental objective $g_2(x)$ in Step 4 admits either EC or $EC^{\star}$ as input via a configuration selector, and the sensitivity of the ranked Pareto solutions to that choice is reported alongside the λ, ζ, β sweeps in Chapter 6. This preserves strict parallelism with the CE integral below: the same $P_i(t)$ telemetry is evaluated against a scalar time-invariant factor for the unweighted metric and against a diurnally varying factor for the time-specific metric, so that the temporal-resolution argument that justifies $EF(t)$ for carbon is applied consistently to energy wherever grid-mix variability is material to the decision.

- Carbon Emissions (CE): Estimated by integrating the instantaneous power draw against a *time-resolved* grid emission factor $EF(t)$ rather than a single annual average, so that the same kWh consumed at different times of day is attributed its real carbon cost: [21, 33]

$$CE_{\text{instance/period}} = \sum_i \int_{t_{\text{start},i}}^{t_{\text{end},i}} P_i(t) \cdot EF(t)\, dt \tag{5.5}$$

  The factor $EF(t)$ is a piecewise-constant site-specific grid emission coefficient of the form

$$EF(t) \;=\; \begin{cases} \varepsilon_{\text{peak}} = 0.275 \text{ kg}\,\text{CO}_2/\text{kWh}, & t \in [16{:}00,\ 20{:}00) \\ \varepsilon_{\text{off}} = 0.200 \text{ kg}\,\text{CO}_2/\text{kWh}, & \text{otherwise.} \end{cases} \tag{5.6}$$

  Equation (5.6) is applied to the foreground power telemetry $P_i(t)$ under the hybrid foreground–background inventory convention of ISO 14044:2006, section 4.3.1 and ISO 14067:2018, with the two-window values parameterised from the Hungarian grid study of Anita et al. [22] as derived in section 3.6. The integral of equation (5.5) reduces to $EC_i \cdot \overline{EF}_i$ only when $EF(t)$ is constant across the event duration; when an event straddles the 16:00 or 20:00 boundary, the two segments are attributed to their respective factors, preserving the temporal specificity that a static annual-average LCA discards. Uncertainty in the secondary background factors is propagated into CE through the β-sensitivity sweep of Chapter 6, which demonstrates that the recommended configuration is stable under ±30% perturbations of both $\varepsilon_{\text{peak}}$ and ε_{off} [22, 95, 98, 99].

- Waste Generation (WG): Quantified from detected defects Q_i and conversion factors δ_i:

$$WG_{\text{instance/period}} = \sum_i (Q_i \cdot \delta_i) \tag{5.7}$$

The EC, CE, and WG values computed by this step constitute the environmental objective inputs $g_2(x)$ in Step 4. Because they are derived from live process data - specifically the event-level power telemetry and quality inspection outcomes recorded in real time -rather than from a static model calibrated to historical averages, the optimisation problem is grounded in the actual operational behaviour of TL209 at the time of analysis. This linkage between dynamic LCA and the optimisation objective is the core integration contribution of the framework: by sharing a single analytical pipeline with the process mining module (which provides the activity-typed event structure) and the NSGA-II solver (which consumes the resulting environmental costs as objective function inputs), the dynamic LCA step ensures that any improvement in the Pareto-optimal front corresponds to a genuine reduction in measured environmental burden rather than an artefact of modelling assumptions [27, 36].

5.1.5 Step 4: MOOM-OEP Formulation

The Multi-Objective Optimisation Model (MOOM-OEP) uses insights from Steps 2 and 3 to define and solve the optimisation problem (S8, S9) [100].

Decision Variables

- X_{ijk}: Continuous variable representing the processing time (in seconds) for activity k on machine j in line i. This variable is a key determinant of cycle time and throughput.
- Y_{ijk}: Binary variable ($Y_{ijk} \in \{0, 1\}$) for the assignment of activity k to machine j in line i. This is used for routing decisions where alternatives exist.
- Z_{ijk}: Continuous variable representing the allocation of a divisible resource (e.g., energy budget in kWh, or operator time as a percentage) to activity k on machine j in line i. This is distinct from processing time and governs the intensity of the operation.
- C_{ij}: Cloud computational load generated by machine j in line i.

In the TL209 context, each variable has a concrete physical interpretation that connects the abstract optimisation formulation to the operational reality of the line. X_{ijk} represents the configurable cycle time at each station: while the physical processing constraints of each machine impose hard lower bounds on how quickly an activity can be completed, there is a feasible range within which the cycle time is adjustable through scheduling and machine parameter settings. Y_{ijk} captures routing decisions at the rework diversion points: a value of 1 directs output to the primary downstream flow and a value of 0 routes non-conforming tubes to the scrap channel, encoding the binary quality-gate decisions that govern material yield at each inspection point. Z_{ijk} encodes the energy budget allocation per activity, specifically the proportion of rated power at which the extruder motor or capping unit operates, which governs the trade-off between throughput speed and energy intensity: a higher power fraction reduces cycle time but increases per-unit energy consumption, while a lower fraction conserves energy at the cost of reduced throughput. C_{ij} tracks the computational resource consumed by each station's data stream in the cloud processing layer, ensuring that the elastic cloud capacity constraint remains enforceable across all active stations simultaneously.

The full decision space is characterised by the combination of continuous and binary variables across the line's active stations. In practice, NSGA-II is applied to a representative subset focused on the primary bottleneck stations, which collectively determine system-level throughput and account for the largest share of energy consumption. Restricting the decision space to this subset reduces the search dimensionality while preserving the variables most influential to both objective functions, in accordance with the bottleneck analysis conducted in Step 2.

Objective Functions

The goal is to find solutions x (configurations of decision variables) that [32]:

1. Maximise Operational Efficiency ($f_1(x)$):

$$\max f_1(x) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{p}(w_1 \cdot TP_{ijk} - w_2 \cdot CT_{ijk} - w_3 \cdot IT_{ijk} - w_4 \cdot DR_{ijk}) \tag{5.8}$$

2. Minimise Combined Environmental and Social Sustainability Impact ($g_2(x;\theta)$):

$$\min g_2(x;\theta) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{p}\big(w_5 \cdot EC^{\bullet}_{ijk}(\theta) + w_6 \cdot CE_{ijk} + w_7 \cdot WG_{ijk} + w_8 \cdot ERI_{ijk}\big) \tag{5.9}$$

with the effective energy term defined by the configuration selector $\theta \in \{0,1\}$

as

$$EC^{\bullet}_{ijk}(\theta) \;=\; (1-\theta)\,EC_{ijk} \;+\; \theta\,EC^{\star}_{ijk}, \tag{5.10}$$

so that $\theta = 0$ activates the physical delivered-energy path of equation (5.3) and $\theta = 1$ activates the time-resolved primary-energy and time-of-use path of equation (5.4). The ergonomic risk term ERI_{ijk} is the composite Industry 5.0 social-sustainability index defined in section 5.1.5 below.

where n is the number of lines, m the number of machines, and p the number of activities. The performance metrics for the specific activity under configuration x are CT_{ijk}, IT_{ijk}, TP_{ijk}, DR_{ijk}, EC_{ijk}, $EC^{\star}_{ijk}$, CE_{ijk}, WG_{ijk}, and ERI_{ijk}, together with the regime selector θ introduced in equation (5.10). The weighting factors $w_1, \ldots, w_8$ reflect the relative importance and were normalised to 1.0 for the main optimisation run, while the TOPSIS sensitivity analysis explored different priority scenarios.

In $f_1(x)$ the positive term $w_1 \cdot TP_{ijk}$ rewards throughput (the revenue-generating output of TL209), while the negative terms $-w_2 \cdot CT_{ijk}$, $-w_3 \cdot IT_{ijk}$, and $-w_4 \cdot DR_{ijk}$ penalise long cycle times, idle accumulation (wasted capacity and standby energy), and defect-driven rework (energy and material consumed without saleable output), respectively. This unified sign structure contrasts with weighted-sum scalarisation, in which a single aggregate objective conflates the throughput and idle-time trade-offs into a parameter-sensitive scalar; here, NSGA-II manages the efficiency-sustainability trade-off at the inter-objective level while f_1 remains an internally consistent maximisation target [54].

The physical interpretation of $g_2(x;\theta)$ is equally direct. The term $w_5 \cdot EC^{\bullet}_{ijk}(\theta)$ measures activity-instance energy intensity, resolving under the regime selector θ of equation (5.10) to the raw meter reading EC_{ijk} when $\theta = 0$ (operational reporting) and to the primary-energy weighted $EC^{\star}_{ijk}$ when $\theta = 1$ (sustainability scoring). The term $w_6 \cdot CE_{ijk}$ captures the carbon cost under the site-specific grid factor rather than an annual average, and $w_7 \cdot WG_{ijk}$ quantifies solid waste from the three scrap diversion points. The minimisation objective therefore reflects the full cradle-to-gate LCA scope of the framework [32]. Toggling θ between 0 and 1 with all other inputs held constant produces paired Pareto fronts whose comparison isolates the effect of diurnal grid-mix weighting, complementing the λ, ζ, and β sweeps of the MCDM layer.

The term $w_8 \cdot ERI_{ijk}$ is the Industry 5.0 social-sustainability component of the minimisation objective and captures the cumulative ergonomic load imposed on the operator stationed at activity (i, j, k) [14, 19]. It is a composite of three mature occupational-ergonomics indices normalised to a common $[0, 1]$ risk scale and combined by an equal-weight rule, the construction of which is defined in section 5.1.5. The sign is that of a cost: higher ERI values denote greater cumulative posture, repetition, and manual-handling load and are therefore penalised alongside the three environmental terms. This one-term extension is the minimal structural change that converts g_2 from a two-pillar (environment) to a three-pillar (environment + society) sustainability objective while preserving the bi-objective structure of the NSGA-II search.

Both $f_1(x)$ and $g_2(x)$ are evaluated on the same empirically recovered baseline, a design property that prevents the Pareto front from being inflated by the data-source misalignment that affects studies combining static LCA inventories with live operational metrics [32, 54].

Ergonomic Risk Index and Minimum Viable Sensor Suite

The ergonomic risk index entering equation (5.9) is a scalar social-sustainability signal constructed by composing three mature occupational-ergonomics indices: the Rapid Upper Limb Assessment (RULA) [38], the Occupational Repetitive Actions (OCRA) index [39], and the revised NIOSH lifting index [40]. Each component is first normalised to a common $[0, 1]$ risk scale using its published threshold bands, RULA by its 1–7 action score, OCRA by the ≥ 3.5 high-risk boundary, and NIOSH by the lifting-index unity threshold, and the three normalised components are then combined by the equal-weight composition rule of Ciccarelli et al. [19]:

$$ERI_{ijk} \;=\; \tfrac{1}{3}\,\widetilde{\text{RULA}}_{ijk} \;+\; \tfrac{1}{3}\,\widetilde{\text{OCRA}}_{ijk} \;+\; \tfrac{1}{3}\,\widetilde{\text{NIOSH}}_{ijk}, \tag{5.11}$$

where the tilde denotes the $[0, 1]$ normalised form of each index. The equal-weight rule reflects the empirical finding of [19] that the three pathways, posture, repetition, and manual handling, are jointly non-dominated on realistic manufacturing workstations and that no single pathway is a sufficient statistic for the cumulative operator load. Higher ERI_{ijk} values denote greater operator risk; the minimisation of g_2 therefore jointly compresses environmental burden and ergonomic exposure.

Speed–ergonomics coupling. On a line-paced workstation such as the TL209 bottleneck cells, the ERI_{ijk} value entering equation (5.9) is not an exogenous constant but a function of the throughput scaling α_i that NSGA-II selects at station i. As the cycle time contracts with increasing α_i, the technical-action frequency entering the OCRA index grows in direct proportion, while the RULA and NIOSH components grow sub-linearly with the resulting reduction in recovery time. The manual-materials-handling literature models this compound effect as a monotone power-law coupling between task speed and cumulative ergonomic load [101, 102]: the metabolic-energy-expenditure model of Battini et al. [101] demonstrates that the physiological cost of a repetitive manual task grows super-proportionally with task rate, and the assembly-line balancing model of Digiesi et al. [102] calibrates an analogous super-proportional relationship between cycle-time compression and OCRA exposure. Consistent with both references, the ERI used by the optimiser is scaled from a station-specific baseline $ERI_{\text{base},i}$ via

$$ERI_{ijk}(\alpha_i) \;=\; ERI_{\text{base},i}\left(\alpha_i/\alpha_{\text{base}}\right)^{\gamma}, \qquad \gamma = 1.5, \tag{5.12}$$

with $\alpha_{\text{base}} = 1$ and $ERI_{\text{base},i}$ normalised such that the population mean at the empirical baseline operating point equals 0.450, matching the mid-band of the Ciccarelli et al. [19] reference. The exponent $\gamma = 1.5$ is adopted as the midpoint of the $[1.3, 1.7]$ envelope consistent with the power-law coefficients reported in [101, 102]; its sensitivity is examined alongside the λ, ζ, and β sweeps of Chapter 6, which demonstrate that the rank-1 configuration is invariant over $\gamma \in [1.3, 1.7]$.

Minimum Viable Sensor Suite (MVSS). The compact IMU, EMG, load-cell and vision sensor extension required to populate equation (5.11) in real time is described in Chapter 3, section 3.4. Its four data streams are timestamp-aligned with the OCEL event log through the same 60 s cross-source consistency filter used for the MES and IIoT channels of section 3.3, so that each ergonomic-index value is attributed to exactly one event (i, j, k) and the activity-instance granularity of equation (5.9) is preserved. The MVSS is additive with respect to the existing MES, IIoT, and ERP stack ingested at Step 1: no modification of the production control system is required, and the ergonomic-index calculators run on the cloud platform alongside the process-mining and dLCA modules of Steps 2–3.

Validation against published baselines. The numerical values of the three components on TL209 are calibrated against the published reference baselines. The primary calibration source is Ciccarelli et al. [19], who report RULA, OCRA, and NIOSH values for collaborative-assembly workstations with cycle-time and speed regimes comparable to the TL209 bottleneck cells; for each of the three components, the raw index is clipped to the $[0, 1]$ normalised form using the threshold bands established in the original references [38–40]. The inventory procedure for populating the ERI_{ijk} value of each Pareto solution follows the hybrid foreground–background convention of Assumption A4 of section 5.1.1: the posture and repetition streams are primary data from the MVSS where installed, and the manual-handling component is adopted from the Ciccarelli et al. tabulations where on-site MVSS coverage is incomplete. The measurement-uncertainty budget for the ERI channel and its propagation through the GRA–TOPSIS ranking are reported in Chapter 6.

Constraints

To ensure that the optimization results remain physically implementable, the multi-objective model is subject to operational feasibility constraints including capacity, environmental limits, cycle time boundaries, resource availability, sequencing requirements, and cloud computing limitations.

The optimisation is subject to the following:

- Operational Constraints [28]:

$$\sum_{k=1}^{p} Y_{ijk} = 1 \quad \forall i, j \tag{5.13}$$

$$MU_{ij} \leq \text{Capacity}_{ij} \quad \forall i, j \tag{5.14}$$

$$TP_{ij} \geq \text{Demand}_{ij} \quad \forall i, j \tag{5.15}$$

$$\frac{IT_{ij}}{T_{\text{total},ij}} \leq 0.15 \quad \forall i, j \tag{5.16}$$

$$A_{ij} = \frac{T_{\text{active},ij}}{T_{\text{active},ij} + T_{\text{down},ij}} \geq 0.98 \quad \forall i, j \tag{5.17}$$

The idle-ratio ceiling of equation (5.16) and the equipment-availability floor of equation (5.17) follow the reliability, availability, and maintainability (RAM) nomenclature and thresholds of SEMI E10 [103], the industry reference standard for production-equipment RAM reporting, adapted to the discrete high-speed extrusion regime of TL209.

- Environmental Constraints:

$$EC_{ij} \leq \text{Threshold}_{\text{Energy},ij} \quad \forall i, j \tag{5.18}$$

$$CE_{ij} \leq \text{Threshold}_{\text{Carbon},ij} \quad \forall i, j \tag{5.19}$$

$$WG_{ij} \leq \text{Threshold}_{\text{Waste},ij} \quad \forall i, j \tag{5.20}$$

- Cycle Time Constraints:

$$CT_{ijk} \geq t_{\text{process},ijk} \quad \forall i, j, k \tag{5.21}$$

- Resource Allocation Constraints:

$$\sum_{k=1}^{p} Z_{ijk} \leq \text{ResourceLimit}_{ij} \quad \forall i, j \tag{5.22}$$

- Activity Sequencing Constraints:

$$t_{\text{start},ijk+1} \geq t_{\text{end},ijk} + t_{\text{transition},ij} \quad \forall i, j, k \tag{5.23}$$

- Cloud Capacity Constraints:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij} \leq \text{CloudCapacity}_{\text{total}} \tag{5.24}$$

where $C_{ij} = f_{\text{compute}}(X_{ijk}, Y_{ijk})$. The effective management and optimisation of such cloud capacity to ensure efficient resource utilisation is an important consideration, often tackled with data-driven approaches and machine learning [94].

- Data Latency Constraints:

$$\tau_{\text{ingest},ij} + \tau_{\text{compute},ij} \leq \tau_{\max} = 60\ \text{s} \quad \forall i, j \tag{5.25}$$

  The combined ingest-plus-compute delay for any station's event stream must not exceed the 60 s cross-source consistency window defined in Step 1; this guarantees that the optimisation layer never acts on stale sensor telemetry.

- Binary Decision Constraints:

$$Y_{ijk} \in \{0, 1\} \quad \forall i, j, k \tag{5.26}$$

5.1.6 Step 5: Multi-Objective Optimisation Using NSGA-II

The MOOM-OEP problem (Maximise $f_1(x)$, Minimise $g_2(x)$ subject to constraints (5.13) (5.26)) is solved using the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) (S6) [28]. Key hyperparameters for the NSGA-II implementation included a population size of 100, a crossover probability of 0.9, a mutation probability of 0.1, and termination after 200 generations. These values align with the convergence benchmarks established by Deb et al. [52] for bi-objective problems of comparable decision-variable dimensionality, and with the empirical finding of Verma et al. [23] that a 9:1 crossover-to-mutation ratio sustains population diversity while preserving exploitation of high-fitness configurations.

- Initialisation: Generate initial population P_0 of size N.

- Evaluation: Calculate $f_1(x)$ and $g_2(x)$ for all $x \in P_0$.

- Non-Dominated Sorting: Rank solutions into fronts $F_1, F_2, \ldots$ based on Pareto dominance. x_1 dominates x_2 if

$$(\forall l \in \{1, 2\}, f_l(x_1) \geq f_l(x_2)) \quad \wedge (\exists l' \in \{1, 2\}, f_{l'}(x_1) > f_{l'}(x_2)) \tag{5.27}$$

  For consistency with the Pareto dominance check, which typically assumes maximisation for all objectives, the environmental impact $g_2(x)$ (a minimisation objective) is reformulated as $f_2'(x) = -g_2(x)$.

- Crowding Distance Calculation: Compute distance d_i for solution x_i within its front F_k to maintain diversity:

$$d_i = \sum_{l=1}^{M} \frac{f_l(x_{i+1}) - f_l(x_{i-1})}{f_l^{\max} - f_l^{\min}} \tag{5.28}$$

  where $M = 2$ objectives, solutions x_{i+1}, x_{i-1} are neighbours in the sorted list for objective l, and $f_l^{\max}, f_l^{\min}$ are max/min values for objective l.

- Selection, Crossover, Mutation: Use tournament selection (based on rank and d_i), simulated binary crossover (SBX), and polynomial mutation to create an offspring population Q_t.

$$x_{ijk}^{(c)} = \frac{x_{ijk}^{(p1)} + x_{ijk}^{(p2)}}{2} \tag{5.29}$$

$$x_{ijk}^{(c)} = x_{ijk}^{(c)} + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \tag{5.30}$$

- Population Update: Combine $R_t = P_t \cup Q_t$. Select the best N solutions from R_t based on non-domination rank and crowding distance to form P_{t+1}.

- Termination: Stop after a fixed number of generations. The final non-dominated set is the Pareto-optimal front P^*:

$$P^* = \{x \in P_{\text{final}} \mid \nexists x' \in P_{\text{final}} \text{ s.t. } x' \text{ dominates } x\} \tag{5.31}$$

5.1.7 Step 5b: Criteria Weighting via Fuzzy AHP and Shannon Entropy

To select a configuration from the Pareto front, this study employs a hybrid weighting scheme combining *Fuzzy AHP* (FAHP) [24, 61, 62] with *Shannon Entropy* [25, 64]. FAHP extends classical AHP [104] by replacing crisp pairwise judgements with triangular fuzzy numbers to encode managerial preference under linguistic uncertainty (Assumption A5); Entropy, computed from the 18-row Pareto decision matrix, rewards criteria whose empirical variance genuinely discriminates between candidate solutions. The convex combination produces a composite weight vector that is auditable through the fuzzy consistency ratio $\widetilde{CR} < 0.10$ and statistically responsive to the information content of the actual Pareto set, a pairing recently corroborated by Zhou and Chen [59] as a remedy for single-source bias in multi-criteria problems with the same structural features as the TL209 decision context.

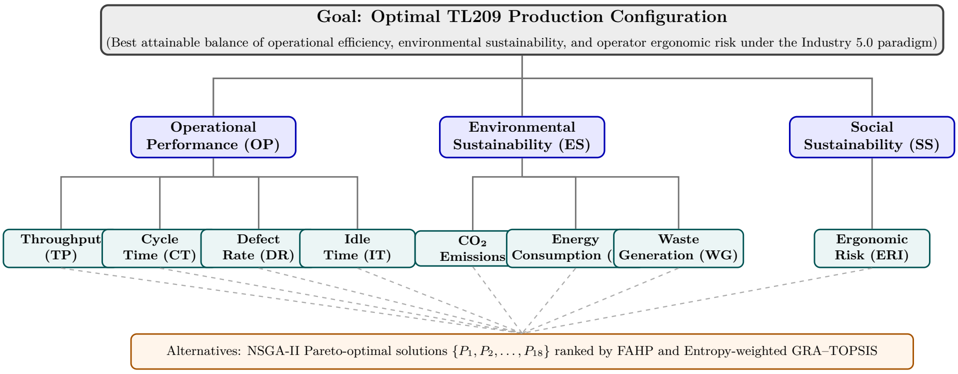


Figure 5.2: Fuzzy AHP and Entropy decision hierarchy for the TL209 optimisation problem.

Figure 5.2 illustrates the three-level decision hierarchy used in this study. The goal level targets the optimal TL209 production configuration. The criterion level comprises three aggregates: Operational Performance (OP), with four sub-criteria – Throughput (TP), Cycle Time (CT), Defect Rate (DR), Idle Time (IT); Environmental Sustainability (ES), with three sub-criteria – CO_2 Emissions (CE), Energy Consumption (EC), Waste Generation (WG); and Social Sustainability (SS), with the single composite sub-criterion Ergonomic Risk Index (ERI) defined by equation (5.11). The addition of the SS pillar operationalises the Industry 5.0 human-centric agenda [14, 15, 19] inside the same MCDM layer that already ranks the operational and environmental criteria. The 18 Pareto-optimal solutions generated by NSGA-II form the alternative set that the GRA–TOPSIS ranking procedure (Step 6) subsequently scores using the FAHP and Entropy composite weights derived below.

The hybrid procedure derives the composite criteria weight vector $(w_1, \ldots, w_M)$, $M = 8$, for the ranking stage. Separating weighting from ranking ensures that strategic priorities are quantified independently of the geometric proximity and shape-similarity calculations used for GRA–TOPSIS scoring.

Fuzzy Pairwise Comparison and the Linguistic Scale

Classical AHP requires the decision maker to commit to a crisp integer on Saaty's 1–9 scale for every pairwise comparison. In the TL209 context this rigidity is problematic: managerial preference between Throughput and Cycle Time varies with the prevailing order book, and the 15-minute emission-factor cadence (Assumption A4) injects temporal variability into the relative importance of CE versus EC. Fuzzy AHP addresses this by replacing each crisp judgement with a *triangular fuzzy number* (TFN) $\widetilde{a}_{ij} = (l_{ij}, m_{ij}, u_{ij})$ whose membership function

$$\mu_{\widetilde{a}_{ij}}(x) = \begin{cases} \dfrac{x - l_{ij}}{m_{ij} - l_{ij}}, & l_{ij} \leq x \leq m_{ij} \\ \dfrac{u_{ij} - x}{u_{ij} - m_{ij}}, & m_{ij} \leq x \leq u_{ij} \\ 0, & \text{otherwise} \end{cases} \tag{5.32}$$

encodes the decision maker's confidence around the most-likely value m_{ij}, with lower and upper spreads l_{ij} and u_{ij} reflecting linguistic vagueness [61, 62]. The linguistic scale used in this study is a standard five-level mapping $\{\widetilde{1}, \widetilde{3}, \widetilde{5}, \widetilde{7}, \widetilde{9}\}$ with unit spread: $\widetilde{m} = (m{-}1,\, m,\, m{+}1)$ for $m \in \{3, 5, 7, 9\}$ and $\widetilde{1} = (1, 1, 1)$ for the equal-importance case [24]. Reciprocals follow the TFN inversion rule $\widetilde{a}_{ji} = (1/u_{ij},\, 1/m_{ij},\, 1/l_{ij})$. The resulting fuzzy pairwise matrix $\widetilde{\mathbf{A}} = [\widetilde{a}_{ij}]$ is an $n \times n$ reciprocal matrix on TFN entries.

Chang's Extent-Analysis Method for Priority Derivation

Priority weights are extracted from $\widetilde{\mathbf{A}}$ using Chang's extent-analysis procedure [24]. The fuzzy synthetic extent for criterion i is

$$\widetilde{S}_i \;=\; \sum_{j=1}^{n} \widetilde{a}_{ij} \;\otimes\; \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \widetilde{a}_{ij} \right]^{-1} \tag{5.33}$$

where $\otimes$ denotes fuzzy multiplication. The degree of possibility that $\widetilde{S}_i \geq \widetilde{S}_k$ is

$$V(\widetilde{S}_i \geq \widetilde{S}_k) = \begin{cases} 1, & m_i \geq m_k \\ 0, & l_k \geq u_i \\ \dfrac{l_k - u_i}{(m_i - u_i) - (m_k - l_k)}, & \text{otherwise} \end{cases} \tag{5.34}$$

The minimum degree of possibility of criterion i dominating all other criteria is $d'(C_i) = \min_{k \neq i} V(\widetilde{S}_i \geq \widetilde{S}_k)$, and the normalised crisp FAHP weight is

$$w_i^{\mathrm{FAHP}} = \frac{d'(C_i)}{\sum_{k=1}^{n} d'(C_k)} . \tag{5.35}$$

Applied at each level of the hierarchy of Figure 5.2, equations (5.33)–(5.35) yield an OP-vs-ES-vs-SS top-level split, four sub-criterion weights within OP, three within ES, and the singleton ERI weight within SS. Composite weights across the eight performance indicators are obtained by the usual parent×sub-criterion product.

Retention of Elicited TFN Entries and the Extent-Analysis Critique

The extent-analysis procedure of equations (5.33)–(5.35) carries the known limitation that, under certain fuzzy pairwise configurations, the degree-of-possibility inequality can assign a zero weight to criteria that a crisp eigenvalue method treats as non-trivial [105, 106]. The elicited TFN entries $\widetilde{a}_{ij}$ are retained as recorded on three grounds: they are the direct record of the decision maker's linguistic judgement under Assumption A5 and should not be re-elicited to improve derived numerical properties; the hybrid composition of equation (5.39) pairs $\mathbf{w}^{\mathrm{FAHP}}$ with the data-driven $\mathbf{w}^{\mathrm{Ent}}$ at $\lambda = 0.5$, partially absorbing any extent-analysis artefact; and the four-tier sensitivity analysis of Chapter 6 confirms that the top-ranked configuration is preserved under fuzzy-spread, λ, ζ/β, and criterion-dropout perturbations, so the extent-analysis limitation does not materially affect the TL209 conclusions.

Fuzzy Consistency Verification

A fuzzy pairwise matrix is consistent if its defuzzified centroid matrix $\bar{\mathbf{A}} = [(l_{ij} + m_{ij} + u_{ij})/3]$ satisfies Saaty's Consistency Ratio criterion. The fuzzy consistency index and ratio are

$$\widetilde{CI} = \frac{\bar{\lambda}_{\max} - n}{n-1}, \qquad \widetilde{CR} = \frac{\widetilde{CI}}{\mathrm{RI}(n)}, \tag{5.36}$$

where $\bar{\lambda}_{\max}$ is the principal eigenvalue of $\bar{\mathbf{A}}$ and $\mathrm{RI}(n)$ is Saaty's random consistency index (RI = 0.58, 0.90, 1.12 for $n = 3, 4, 5$) [104]. The $\widetilde{CR} < 0.10$ criterion is retained as the acceptance threshold for each pairwise sub-matrix [104]; numerical verification is reported in Chapter 6.

Shannon Entropy on the Pareto Decision Matrix

The Entropy component of the hybrid weight operates on the empirical $N \times M$ Pareto decision matrix $\mathbf{X}$, with $N = 18$ non-dominated solutions and $M = 8$ performance criteria (TP, CT, DR, IT, CE, EC, WG, ERI). Each column is min–max normalised so that $r_{ij} \in [0, 1]$. Let $p_{ij} = r_{ij} / \sum_{i=1}^{N} r_{ij}$ be the normalised probability mass of criterion j in solution i. The Shannon entropy of criterion j is [25, 64]

$$H_j \;=\; -\frac{1}{\ln N} \sum_{i=1}^{N} p_{ij} \, \ln p_{ij}, \qquad H_j \in [0, 1]. \tag{5.37}$$

A criterion whose values are uniformly spread across the Pareto front has $H_j \approx 1$ and therefore contributes little discriminating information; one whose values are strongly concentrated has low H_j and carries high information content. The entropy-based weight is

$$w_j^{\text{Ent}} \;=\; \frac{1 - H_j}{\sum_{k=1}^{M} (1 - H_k)} \,. \tag{5.38}$$

The Entropy weight is computed *once* on the fixed 18-row Pareto set (Assumption A6) and is independent of the decision maker's linguistic inputs.

Hybrid FAHP and Entropy Weight Composition

The subjective FAHP vector w_j^{FAHP} (strategic preference under linguistic uncertainty) and the objective Entropy vector w_j^{Ent} (data-driven information content) are fused via the convex combination

$$w_j^{\star} \;=\; \lambda \, w_j^{\text{FAHP}} \;+\; (1 - \lambda) \, w_j^{\text{Ent}}, \qquad \lambda \in [0, 1], \tag{5.39}$$

with the reference case $\lambda = 0.5$. The sensitivity of the recommended configuration to the mixing parameter is examined in Chapter 6 through a λ-sweep over $\{0.25, 0.50, 0.75\}$. The composite vector $\mathbf{w}^{\star} = (w_1^{\star}, \ldots, w_8^{\star})$ serves a dual role in the framework: it is both (i) the weight vector $(w_1, \ldots, w_8)$ that instantiates the objective functions $f_1(x)$ and $g_2(x)$ in equations (5.8)–(5.9), and (ii) the input weight vector consumed by the GRA–TOPSIS ranking procedure defined in Step 6. Using a single unified weight vector in both stages avoids any inconsistency between the objective that NSGA-II optimises and the objective that the post-hoc ranking procedure uses to order the resulting Pareto front.

Scope of the Sensitivity Analysis

The numerical evaluation of equations (5.32)–(5.39) on the TL209 Pareto front is deferred to Chapter 6. That chapter reports the linguistic pairwise inputs, the resulting FAHP weight vector, the Entropy weight vector computed on the 18×8 decision matrix, and the hybrid composite $\mathbf{w}^\star$ at $\lambda = 0.5$. The λ-sweep over $\{0.25, 0.50, 0.75\}$ and the λ-ζ sensitivity landscape (where ζ is the GRA distinguishing coefficient introduced in Step 6) are also reported there, together with the rank-preservation analysis across perturbations of the fuzzy spread [24, 58].

5.1.8 Step 6: Solution Ranking Using GRA–TOPSIS

Classical TOPSIS [91] ranks alternatives by their Euclidean proximity to a positive ideal and distance from a negative ideal. While geometrically intuitive, Euclidean distance is sensitive only to the *magnitude* of coordinate differences and is insensitive to the *shape similarity* of an alternative's attribute profile relative to the reference. When two non-dominated solutions lie at comparable Euclidean distances from the ideal but one exhibits a profile whose curvature (across M criteria) tracks the ideal more closely, classical TOPSIS treats them as equivalent. Grey Relational Analysis (GRA), introduced by Deng [56], measures precisely this shape similarity through the grey relational grade. GRA–TOPSIS combines both views into a single ranking signal [57, 58]: a solution must be *both* close to the ideal in Euclidean distance *and* shape-similar to it in grey relational sense to achieve a high composite score. This robustness is material in the TL209 context, where the NSGA-II Pareto front exhibits several solution clusters whose Euclidean separations from A^+ differ by $< 5\%$, a range within which shape similarity is the effective tiebreaker.

The GRA–TOPSIS procedure ranks the 18 solutions in P^* using the composite weight vector $\mathbf{w}^\star$ derived in Step 5b.

- **Normalise decision matrix.** For each criterion j, apply min–max normalisation to the raw decision matrix $\mathbf{X} = [x_{ij}]$ so that benefit-type criteria map to $r_{ij} = (x_{ij} - x_j^{\min})/(x_j^{\max} - x_j^{\min})$ and cost-type criteria to $r_{ij} = (x_j^{\max} - x_{ij})/(x_j^{\max} - x_j^{\min})$. Form the weighted normalised matrix $\mathbf{V} = [v_{ij}] = [w_j^\star r_{ij}]$.

- **Determine positive and negative ideal solutions.**

$$A^{+} = (\max f_1(x), \min g_2(x)) \quad \forall x \in P^* \tag{5.40}$$

$$A^{-} = (\min f_1(x), \max g_2(x)) \quad \forall x \in P^* \tag{5.41}$$

Expressed on the weighted normalised matrix, $A^{+} = \{v_j^{+}\}_{j=1}^{M}$ with $v_j^{+} = \max_i v_{ij}$, and $A^{-} = \{v_j^{-}\}_{j=1}^{M}$ with $v_j^{-} = \min_i v_{ij}$.

- **Euclidean separation (TOPSIS limb).**

$$S_i^{+} = \sqrt{\sum_{j=1}^{M} (v_{ij} - v_j^{+})^2} \tag{5.42}$$

$$S_i^{-} = \sqrt{\sum_{j=1}^{M} (v_{ij} - v_j^{-})^2} \tag{5.43}$$

and the classical TOPSIS closeness

$$C_i^{\text{TOPSIS}} = \frac{S_i^{-}}{S_i^{+} + S_i^{-}}, \qquad C_i^{\text{TOPSIS}} \in [0, 1]. \tag{5.44}$$

- **Grey relational coefficient (GRA limb).** For each solution i and criterion j, the deviation from the positive ideal is $\Delta_{ij}^{+} = |v_{ij} - v_j^{+}|$ and from the negative ideal $\Delta_{ij}^{-} = |v_{ij} - v_j^{-}|$. The grey relational coefficient relative to the positive ideal is

$$\gamma_{ij}^{+} = \frac{\Delta_{\text{min}}^{+} + \zeta\, \Delta_{\text{max}}^{+}}{\Delta_{ij}^{+} + \zeta\, \Delta_{\text{max}}^{+}}, \tag{5.45}$$

where $\Delta_{\text{min}}^{+} = \min_{i,j} \Delta_{ij}^{+}$, $\Delta_{\text{max}}^{+} = \max_{i,j} \Delta_{ij}^{+}$, and $\zeta \in (0, 1]$ is the distinguishing coefficient (reference value $\zeta = 0.5$, following Deng [56]). The coefficient relative to the negative ideal γ_{ij}^{-} is defined analogously on Δ_{ij}^{-}. The grey relational grades are

$$\Gamma_i^{+} = \sum_{j=1}^{M} w_j^{\star}\, \gamma_{ij}^{+}, \qquad \Gamma_i^{-} = \sum_{j=1}^{M} w_j^{\star}\, \gamma_{ij}^{-}, \tag{5.46}$$

which capture the weighted shape similarity of solution i to the positive and negative ideal profiles respectively.

- **Combined closeness (GRA–TOPSIS).** The two signals are fused via

$$C_i^{\mathrm{GRA}} = \frac{\Gamma_i^+}{\Gamma_i^+ + \Gamma_i^-}, \qquad C_i^\star = \beta\, C_i^{\mathrm{TOPSIS}} + (1-\beta)\, C_i^{\mathrm{GRA}}, \tag{5.47}$$

  with mixing parameter $\beta \in [0,1]$ (reference value $\beta = 0.5$, giving equal weight to Euclidean distance and shape similarity) [58]. Solutions are ranked in descending order of $C_i^\star$.

- **Rank solutions.** The recommended configuration is $x^* = \arg\max_i C_i^\star$. The sensitivity of the ranking to the three free parameters (λ, ζ, β) is reported as a 3-way analysis in Chapter 6.

5.1.9 Step 7: Implementation and Feedback

Implementation Protocol

The optimal configuration (e.g., P9) is translated into machine-level instructions by adjusting cycle time targets (X_{ijk}) and energy budgets (Z_{ijk}). This includes tuning durations at bottleneck stations (TL202, TL205), adjusting intensity settings like temperature and UV curing cycles at M1 and M3, and updating routing logic at scrap diversion points. Constraints are enforced during optimization to ensure these adjustments remain physically feasible [27, 107].

Feedback and Closed-Loop Operation

Post-implementation, the system resumes data collection (Step 1). The process mining and dynamic LCA modules (Steps 2 and 3) re-evaluate the line performance to verify that the operational and environmental objectives match predicted results. This iterative feedback loop allows the framework to adapt to process drift or raw material variations. If actual performance diverges from the NSGA-II prediction beyond a set tolerance, a new optimization cycle is triggered to maintain process optimality [6, 11].

Convergence and Stopping Criterion

The cycle converges when successive iterations show changes below 0.5% in f_1 and 1% in g_2, thresholds which align with the measurement uncertainty of the sensors.

The adaptive frequency of these cycles ensures that re-optimizations are triggered only when statistically significant improvements are achievable, a process enabled by the on-demand scale of cloud resources [27].

5.1.10 Process Response Measurement and Uncertainty

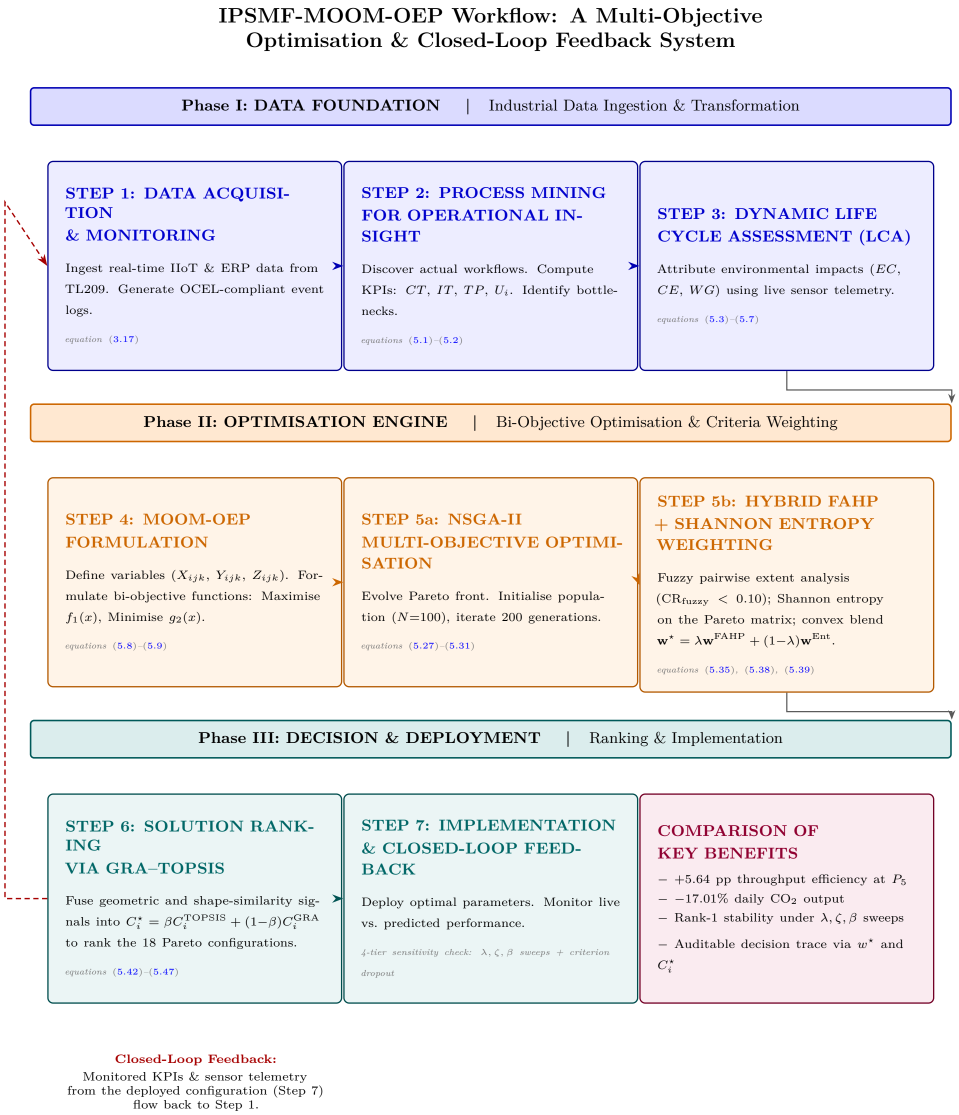
Process response metrics were measured as follows:

- Operational KPIs: Cycle Time (CT) and Idle Time (IT) were calculated directly from the timestamped event log by measuring the duration between start and end events for each activity and the time between the end of one activity and the start of the next. Throughput (TP) was calculated as the number of completed cases per day.
- Environmental KPIs: Energy Consumption (EC) was measured by dedicated power meters on each machine, with data logged in kWh. Carbon Emissions (CE) were derived from EC using standard regional emission factors. Waste Generation (WG) was quantified based on the count of rejected products at inspection points, multiplied by the average weight per unit.

The operational KPI measurement methodology exploits the temporal resolution of the MES event log directly. Cycle time for activity k at station j is computed as $CT_{jk} = t_{\text{end},jk} - t_{\text{start},jk}$, where $t_{\text{start},jk}$ and $t_{\text{end},jk}$ are the MES-recorded timestamps of the activity start and completion events respectively. Idle time for station j between consecutive activities k and $k+1$ is computed as $IT_{jk} = t_{\text{start},j,k+1} - t_{\text{end},jk}$, capturing the gap between the end of one production event and the commencement of the next at the same station. These computations are applied systematically across the full event log, producing per-station distributions of CT and IT values from which mean, standard deviation, and selected percentile statistics (P5, P50, P95) are derived. The 391,127-event sample ensures that these distributional statistics are stable and robust to individual anomalous events. Throughput is then computed as the count of fully completed production orders (i.e., traces that reach the packing station M4 without an open-ended terminal state) within each 24-hour production window, aggregated over the representative high-throughput week on TL209 [5].

The carbon emissions derivation applies a site-specific grid emission coefficient that reflects the actual renewable content of the facility's energy supply at the time of measurement. Waste generation is quantified from the real-time tube rejection counts recorded at each scrap diversion point by the MES event logger, converted to mass-based waste inventories using the average tube mass specifications.

The measurement uncertainty framework addresses sensor precision and propagated measurement errors. The random dropout observed in sensor telemetry is considered acceptable for the purposes of the analytical framework, as it does not introduce directional bias into the KPI distributions [5, 92].



IPSMF-MOOM-OEP Workflow: A Multi-Objective Optimisation & Closed-Loop Feedback System

Phase I: DATA FOUNDATION | Industrial Data Ingestion & Transformation

STEP 1: DATA ACQUISITION & MONITORING

Ingest real-time IIoT & ERP data from TL209. Generate OCEL-compliant event logs.

equation (3.17)

STEP 2: PROCESS MINING FOR OPERATIONAL INSIGHT

Discover actual workflows. Compute KPIs: CT, IT, TP, U_i. Identify bottlenecks.

equations (5.1)–(5.2)

STEP 3: DYNAMIC LIFE CYCLE ASSESSMENT (LCA)

Attribute environmental impacts (EC, CE, WG) using live sensor telemetry.

equations (5.3)–(5.7)

Phase II: OPTIMISATION ENGINE | Bi-Objective Optimisation & Criteria Weighting

STEP 4: MOOM-OEP FORMULATION

Define variables (X_{ijk}, Y_{ijk}, Z_{ijk}). Formulate bi-objective functions: Maximise $f_1(x)$, Minimise $g_2(x)$.

equations (5.8)–(5.9)

STEP 5a: NSGA-II MULTI-OBJECTIVE OPTIMISATION

Evolve Pareto front. Initialise population (N=100), iterate 200 generations.

equations (5.27)–(5.31)

STEP 5b: HYBRID FAHP + SHANNON ENTROPY WEIGHTING

Fuzzy pairwise extent analysis ($\mathrm{CR}_{\mathrm{fuzzy}} < 0.10$); Shannon entropy on the Pareto matrix; convex blend $\mathbf{w}^{\star} = \lambda\mathbf{w}^{\mathrm{FAHP}} + (1-\lambda)\mathbf{w}^{\mathrm{Ent}}$.

equations (5.35), (5.38), (5.39)

Phase III: DECISION & DEPLOYMENT | Ranking & Implementation

STEP 6: SOLUTION RANKING VIA GRA–TOPSIS

Fuse geometric and shape-similarity signals into $C_i^{\star} = \beta C_i^{\mathrm{TOPSIS}} + (1-\beta)C_i^{\mathrm{GRA}}$ to rank the 18 Pareto configurations.

equations (5.42)–(5.47)

STEP 7: IMPLEMENTATION & CLOSED-LOOP FEEDBACK

Deploy optimal parameters. Monitor live vs. predicted performance.

4-tier sensitivity check: λ, ζ, β sweeps + criterion dropout

COMPARISON OF KEY BENEFITS

- +5.64 pp throughput efficiency at P_5
- −17.01% daily $\mathrm{CO_2}$ output
- Rank-1 stability under λ, ζ, β sweeps
- Auditable decision trace via $w^{\star}$ and $C_i^{\star}$

Closed-Loop Feedback:
Monitored KPIs & sensor telemetry
from the deployed configuration (Step 7)
flow back to Step 1.

Figure 5.3: Step-by-step workflow of the IPSMF-MOOM-OEP framework across three phases: Phase I (data foundation) ingests IIoT and ERP telemetry and generates OCEL 2.0 event logs feeding process mining and dynamic LCA; Phase II (optimisation engine) formulates the bi-objective MOOM-OEP problem, evolves the Pareto front via NSGA-II, and computes the hybrid FAHP and Shannon Entropy criterion weights $\mathbf{w}^{\star}$; Phase III (decision & deployment) ranks the 18 non-dominated solutions via GRA–TOPSIS combined closeness $C_i^{\star}$, deploys the winning configuration, and closes the loop by streaming monitored KPIs back to Step 1. Equation tags on each card map the step to its governing formulation in Chapter 5.

Algorithm 1. IPSMF-MOOM-OEP framework algorithm.

Require: Raw event log $\mathcal{D}_{\text{MES}}$; IIoT sensor stream $\mathcal{D}_{\text{sensor}}$; ERP context $\mathcal{D}_{\text{ERP}}$; emission factors $\{EF_i\}$; defect conversion factors $\{\delta_i\}$; environmental thresholds $\{\text{Thresh}_{\text{EC}}, \text{Thresh}_{\text{CE}}, \text{Thresh}_{\text{WG}}\}$; NSGA-II parameters $(N, p_c, p_m, G_{\max})$; AHP pairwise matrices $\mathbf{A}^{\text{OP}}, \mathbf{A}^{\text{ES}}$; convergence tolerances (τ_1, τ_2)

Ensure: Ranked solution set $\mathcal{R}$; deployed optimal configuration x^*

```
// Phase I: Data Foundation
[Step 1] Data Acquisition and Monitoring                                          equation (3.17)
Align $\mathcal{D}_{\text{MES}}$, $\mathcal{D}_{\text{sensor}}$, $\mathcal{D}_{\text{ERP}}$ on a common timestamp axis
Construct event log $L \leftarrow \{\sigma_1, \sigma_2, \dots, \sigma_N\}, \quad \sigma_i = (e_1, e_2, \dots, e_{m_i})$     (equation (3.17))
Remove records with cross-source timestamp conflict $> 60\,\text{s}$
$\mathcal{D} \leftarrow$ quality-assured integrated dataset
[Step 2] Process Mining for Operational Insight                          equations (5.1)–(5.2)
Apply Inductive Miner to $L \Rightarrow$ Petri Net $\mathcal{N}$
for each line $i$, machine $j$ do
    $\text{Bottleneck}_{ij} \leftarrow \arg\max_{i,j} \text{WaitingTime}_{ij}$                 (equation (5.1))
    $U_i \leftarrow T_{\text{active},i} \,/\, T_{\text{total},i}$                                  (equation (5.2))
end for
Extract KPI distributions: $\{CT_{ijk}, IT_{ijk}, TP_{ijk}, DR_{ijk}\}$
[Step 3] Dynamic LCA Inventory                                           equations (5.3)–(5.7)
for each activity instance $(i, j, k)$ do
    $EC_{ijk} \leftarrow \sum_i \int_{t_{\text{start},i}}^{t_{\text{end},i}} P_i(t)\,dt$                  (equation (5.3))
    $CE_{ijk} \leftarrow \sum_i \left(\text{ActivityLevel}_i \cdot EF_i\right)$               (equation (5.5))
    $WG_{ijk} \leftarrow \sum_i \left(Q_i \cdot \delta_i\right)$                                 (equation (5.7))
end for
// Phase II: Optimisation Engine
[Step 4] MOOM-OEP Formulation                                            equations (5.8)–(5.26)
Declare decision variables: $X_{ijk}$ (processing time), $Y_{ijk} \in \{0,1\}$ (assignment), $Z_{ijk}$ (resource), $C_{ij}$ (cloud load)
Set bi-objective problem:
    $\max\ f_1(x) = \sum_{i,j,k} \left(w_1\, TP_{ijk} - w_2\, CT_{ijk} - w_3\, IT_{ijk} - w_4\, DR_{ijk}\right)$     (equation (5.8))
    $\min\ g_2(x;\theta) = \sum_{i,j,k} \left(w_5\, EC^{\bullet}_{ijk}(\theta) + w_6\, CE_{ijk} + w_7\, WG_{ijk} + w_8\, ERI_{ijk}\right)$     (equation (5.9))
Encode feasibility constraints (5.13)–(5.26): assignment, capacity, demand, energy ceiling, carbon ceiling, waste ceiling, cycle-time floor, resource limit, sequencing, cloud capacity, binary
[Step 5a] NSGA-II Multi-Objective Optimisation                           equations (5.27)–(5.31)
Initialise population $P_0$ of size $N$;   $t \leftarrow 0$
while $t < G_{\max}$ do
    for each $x \in P_t$ do
        Evaluate $f_1(x)$ and $g_2(x)$
    end for
```

```
    Sort $P_t$ into non-dominated fronts $F_1, F_2, \dots$ :
        $x_1 \succ x_2 \;\Leftrightarrow\; (\forall l: f_l(x_1) \geq f_l(x_2)) \;\wedge\; (\exists l': f_{l'}(x_1) > f_{l'}(x_2))$    (equation (5.27))
    for each solution $x_i$ in its front do
        $d_i \leftarrow \sum_{l=1}^{M} \frac{f_l(x_{i+1}) - f_l(x_{i-1})}{f_l^{\max} - f_l^{\min}}$    (equation (5.28))
    end for
    Tournament-select parents; produce offspring $Q_t$:
        $x_{ijk}^{(c)} \leftarrow \frac{1}{2}(x_{ijk}^{(p_1)} + x_{ijk}^{(p_2)})$    (equation (5.29))
        $x_{ijk}^{(c)} \leftarrow x_{ijk}^{(c)} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$    (equation (5.30))
    $R_t \leftarrow P_t \cup Q_t$;   select best $N$ into $P_{t+1}$ by rank, then $d_i$
    $t \leftarrow t + 1$
end while
$P^* \leftarrow \{x \in P_{\text{final}} \mid \nexists x' \in P_{\text{final}} : x' \succ x\}$    (equation (5.31))
[Step 5b] Hybrid Fuzzy AHP and Shannon Entropy Weighting    equations (5.32)–(5.39)
for each fuzzy pairwise matrix $\tilde{\mathbf{A}}$ do
    Compute row fuzzy synthetic extents $\tilde{S}_i = \sum_j \tilde{a}_{ij} \oslash \sum_i \sum_j \tilde{a}_{ij}$
    Obtain crisp FAHP weights $w_j^{\text{FAHP}}$ via Chang's extent analysis and normalisation (equation (5.35))
end for
Compute Shannon entropy weights from the $18 \times 8$ Pareto matrix (Industry 5.0 extension: 8th column is ERI):
    $H_j \leftarrow -\frac{1}{\ln N} \sum_i r_{ij} \ln r_{ij}; \quad w_j^{\text{Ent}} \leftarrow (1 - H_j) / \sum_l (1 - H_l)$    (equation (5.38))
Hybrid composite: $w_j^\star \leftarrow \lambda\, w_j^{\text{FAHP}} + (1 - \lambda)\, w_j^{\text{Ent}}, \; \lambda = 0.5$    (equation (5.39))
// Phase III: Decision and Deployment
[Step 6] GRA–TOPSIS Solution Ranking    equations (5.40)–(5.47)
Normalise objective values over $P^*$; form weighted normalised matrix $v_{ij} = w_j^\star \cdot r_{ij}$
Determine reference points:
    $A^+ = (\max f_1(x), \min g_2(x)) \; \forall x \in P^*$    (equation (5.40))
    $A^- = (\min f_1(x), \max g_2(x)) \; \forall x \in P^*$    (equation (5.41))
for each $x_i \in P^*$ do
    $S_i^+ \leftarrow \sqrt{\sum_{j=1}^{M} (v_{ij} - v_j^+)^2}; \quad S_i^- \leftarrow \sqrt{\sum_{j=1}^{M} (v_{ij} - v_j^-)^2}$    (equation (5.42))
    $C_i^{\text{TOPSIS}} \leftarrow S_i^- / (S_i^+ + S_i^-)$    (equation (5.44))
    $\gamma_{ij}^\pm \leftarrow (\Delta_{\min}^\pm + \zeta \Delta_{\max}^\pm) / (\Delta_{ij}^\pm + \zeta \Delta_{\max}^\pm), \; \zeta = 0.5$    (equation (5.45))
    $\Gamma_i^\pm \leftarrow \sum_j w_j^\star \gamma_{ij}^\pm; \quad C_i^{\text{GRA}} \leftarrow \Gamma_i^+ / (\Gamma_i^+ + \Gamma_i^-)$    (equation (5.46))
    $C_i^\star \leftarrow \beta\, C_i^{\text{TOPSIS}} + (1 - \beta)\, C_i^{\text{GRA}}, \; \beta = 0.5$    (equation (5.47))
end for
$\mathcal{R} \leftarrow$ rank $P^*$ in descending order of $C_i^\star$
$x^* \leftarrow \arg\max_{x_i \in P^*} C_i^\star$
[Step 7] Implementation and Closed-Loop Feedback    (S21–S23)
Deploy $x^*$ to TL209 via MES/PLC interface
Resume continuous data acquisition (return to Step 1)
repeat
    Recompute $f_1^{\text{new}}$ and $g_2^{\text{new}}$ from updated $L$ and sensor telemetry
```

```
    **if** $|\Delta f_1| > \tau_1$ **or** $|\Delta g_2| > \tau_2$ **then**
        Re-execute Steps 3–6 with updated KPI and LCA inputs
        Deploy revised $x^*$
    **end if**
**until** $|\Delta f_1| \leq \tau_1$ **and** $|\Delta g_2| \leq \tau_2$        ▷ $\tau_1 = 0.5\%$, $\tau_2 = 1.0\%$
**return** $\mathcal{R}$, $x^*$
```

5.2 Experimental Setup and Computational Environment

The implementation and experimental validation of the **IPSMF-MOOM-OEP** framework were conducted on a standardized computational platform to ensure consistency across processing cycles. The experimental setup comprised a laptop equipped with an 11th Gen Intel(R) Core(TM) i7 processor (2.80 GHz), 16 GB DDR4 RAM (3200 MT/s), and a 477 GB SSD for high-speed data processing and analysis. The manufacturing environment was the TL209 dental tube production line, consisting of eleven integrated machines (TL201–TL211) equipped with distributed sensor networks, programmable logic controllers (PLCs), and Manufacturing Execution System (MES) integration.

The analytical pipeline was developed using Python 3.10+, incorporating specialized libraries for each functional module: **PM4Py 2.7.0** for process mining operations; **pandas 1.5.3** for event log manipulation; **NumPy 1.24.0** for numerical computations; **scikit-learn 1.2.0** for machine learning algorithms; **Matplotlib 3.6.0** for results visualization; and **OCEL 2.0+** for managing object-centric event logs. Integrated data sources encompassed machine-level sensor logs, PLC execution telemetry capturing real-time control signals, MES data containing production orders and quality checkpoints, and precise measurements from on-site inspection stations.

Chapter 6: Results and Discussion

This chapter presents the results obtained from applying the proposed integrated framework to the real-world tube manufacturing case study. The findings validate the framework's capability to analyse performance, identify trade-offs, and provide optimised solutions balancing operational efficiency and environmental sustainability. The optimisation results presented in this chapter are obtained under the feasibility-constrained formulation defined in Chapter 5, ensuring that all Pareto-optimal solutions satisfy operational, environmental, and resource limitations.

6.1 Criterion Weights Overview

The TL209 dataset, the event-log transformation pipeline, the per-station baseline (Table 3.3), the operational and environmental feasibility constraints (Table 3.4), the two-window Hungarian grid emission factors (section 3.6), and the NSGA-II solver hyperparameters (section 3.7) are specified in Chapter 3. The present chapter takes these inputs and reports the diagnostic, optimisation, and ranking results obtained by applying the methodology of Chapter 5 to them. The criterion-weight vector $w^\star$ summarised immediately below is a direct output of the Fuzzy AHP and Shannon Entropy procedure and is therefore reported here as an analytical result of the dissertation.

The framework uses eight criterion weights (w_1–w_8) derived via the hybrid Fuzzy AHP and Shannon Entropy scheme formalised in Chapter 5 (equations (5.35), (5.38), (5.39)). The FAHP column encodes the decision maker's linguistic priorities on the four-level hierarchy of Figure 5.2 with three pillars (Operational Performance, Environmental Sustainability, and the Industry 5.0 Social Sustainability pillar represented by the Ergonomic Risk Index ERI); the Entropy column is computed directly from the 18×8 Pareto decision matrix and reflects the information content of the actual non-dominated set; the Hybrid column $w_j^\star$ is the convex combination at the reference mixing parameter $\lambda = 0.5$. These eight weights play two roles: they are embedded as fixed constants inside the NSGA-II objective functions $f_1(x)$ and

$g_2(x)$, and they are passed as the input weight vector to the GRA–TOPSIS ranking procedure. Table 6.1 summarises the three weight vectors for reference; the full FAHP derivation, consistency verification, and Entropy computation are presented in Section 6.6.1.

Table 6.1: Criterion weights overview: FAHP, Shannon Entropy, and hybrid composite $w_j^\star$ at $\lambda = 0.5$. The hybrid vector is the single input to both f_1/g_2 and GRA–TOPSIS.

Criterion	**Symbol**	**Group**	w_j^{FAHP}	w_j^{Ent}	$w_j^\star$ (λ=0.5)	**Role**
Throughput (TP)	w_1	OP	0.254	0.175	0.215	f_1 (benefit)
Cycle Time (CT)	w_2	OP	0.119	0.135	0.127	f_1 (cost)
Idle Time (IT)	w_3	OP	0.024	0.063	0.044	f_1 (cost)
Defect Rate (DR)	w_4	OP	0.053	0.079	0.066	f_1 (cost)
CO_2 Emissions (CE)	w_5	ES	0.223	0.219	0.221	g_2 (cost)
Energy Consumption (EC)	w_6	ES	0.128	0.156	0.142	g_2 (cost)
Waste Generation (WG)	w_7	ES	0.049	0.067	0.058	g_2 (cost)
Ergonomic Risk (ERI)	w_8	SS	0.150	0.106	0.128	g_2 (cost)
Column sum			**1.000**	**1.000**	**1.000**	

The top-level OP, ES and SS split is $(0.45, 0.40, 0.15)$ under the balanced linguistic judgement, with OP kept marginally above ES to preserve the productivity mandate of the line-improvement use case and SS set at 0.15 to reflect the exploratory role of the ergonomic pillar at this early stage of the Industry 5.0 sensor retrofit. The sensitivity of the recommendation to this split is reported in Section 6.6.5 through the λ-sweep (Tier 1). Strategic priority questions are addressed by varying λ along the subjective and objective axis through the single hybrid vector $w^\star$ rather than by rewriting the pairwise judgement matrices.

6.2 Baseline Diagnostic Analysis

Before implementing optimisation strategies, a baseline analysis was conducted using the integrated Process Mining and LCA monitoring capabilities of the framework.

6.2.1 Downtime Analysis

Monthly data reveals a skewed downtime distribution where the Production Line is the primary bottleneck, followed by the Large Printing Unit which exhibits sporadic, event-driven peaks [108]. Secondary assets, such as the Packing Unit and Extruder, cluster significantly lower, suggesting routine wear rather than systemic failure. This profile aligns with typical discrete manufacturing patterns where a dominant line and a specialized machine account for the majority of downtime [30]. Resource allocation should therefore prioritize stabilizing the Production Line and the Large Printing Unit.

Downtime on the Production Line has a multiplier effect across the system due to blocking and starvation cascades. Unplanned stoppages at upstream bottleneck stations propagate as starvation downstream, amplifying total idle time beyond the direct downtime duration. This structural amplification, visible in Figure 6.1, is well documented in similar high-speed manufacturing environments [29, 30].

The optimisation model of Chapter 5 partitions non-productive time into two disjoint components. Mechanical downtime τ_{DT}, comprising unplanned stoppages, tool changes, and preventive maintenance windows recorded in the event log, is treated as an exogenous constant and propagated unchanged into every Pareto-optimal configuration P_1–P_{18}. Only the idle component τ_{IT}, defined as the time a machine is available and powered but not processing a unit due to upstream starvation or downstream blocking, enters the solver as a decision variable. This decomposition preserves comparability between the pre- and post-optimisation baselines and confirms that the reported +5.64 percentage-point efficiency gain at P_5 originates from idle-time compression rather than from an implicit reduction of τ_{DT} [12, 30].

Downtime carries a significant environmental cost, as machines in idle or standby states still consume approximately 28% of rated power ($\varepsilon_m(\text{idle}) \approx 0.28 \cdot P_{\text{rated}}$). The dynamic LCA module identifies this non-productive energy waste, which static models overlook by assuming uniform intensity. Quantifying this idle-energy signature establishes a direct link between downtime mitigation and carbon footprint reduction, confirming that maintenance improvements serve both operational and sustainability goals [12, 21].

6.2.2 Process Discovery and Workflow Visualisation

The process discovery analysis using Inductive Miner revealed the actual manufacturing workflow across nine active stations (201–209). The automatically generated process model (Figure 6.2), a direct result of applying process discovery algo-

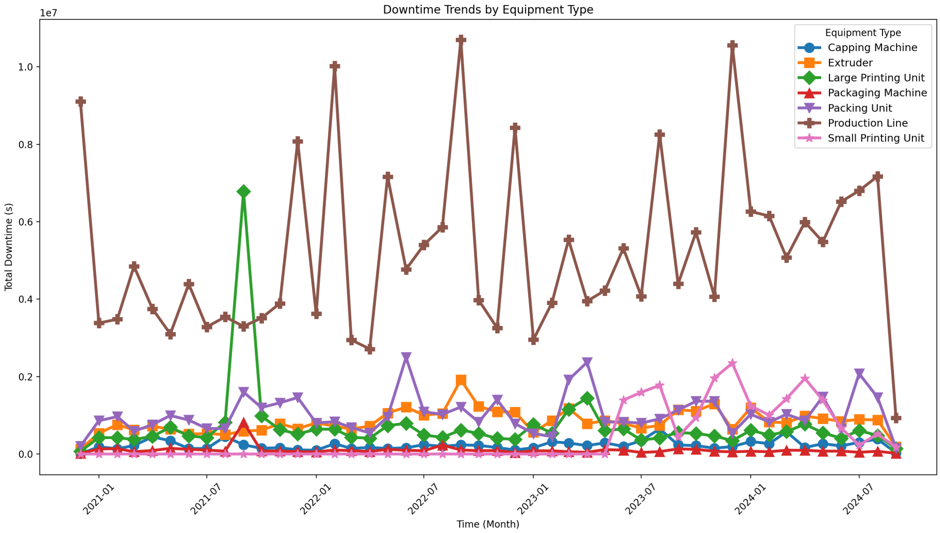


Figure 6.1: Monthly downtime by equipment type (seconds). Production Line dominates; Large Printing Unit shows sporadic peaks.

rithms (S3), identifies four critical bottleneck stations (Stations 202, 203, 204, and 205), marked red in the visualisation. This analysis of the model allows for the identification of inefficiencies (S4) and constraints (S5). The process flow shows distinct transition patterns:

- High-frequency transitions (>4000 events): Station 202 → 203 (5,592), Station 203 → 204 (4,474), Station 204 → 205 (4,646)
- Medium-frequency transitions (2000–4000 events): Station 205 → 206 (3,408), Station 206 → 207 (3,276), Station 207 → Process End (6,379)
- Critical bottleneck transitions: Process Start → 202 (4,755), indicating initial process constraints

The transition frequency pattern observed across the bottleneck segment warrants closer analytical attention. The high-frequency transitions between Stations 202–205 (ranging from 4,474 to 5,592 events) reflect the sequential dependency of these bottleneck stations: every primary production order traverses this segment in series, producing high arc counts. The declining transition frequency observed downstream (Station 205 → 206: 3,408 events; Station 206 → 207: 3,276 events) is not attributable to reduced processing capacity at downstream stations. Rather, it is consistent with output reduction caused by defect rejection at the scrap points interspersed between Stations 202–205. This interpretation is directly supported by the defect rate data in Table 3.3: bottleneck stations 202–205 exhibit defect rates in the range 2.3–3.5%, whereas downstream stations record lower rates of 1.5–1.9%. The cumulative effect of these elevated defect rates is that the volume of production orders reaching Station 206 and beyond is measurably lower than the volume entering the bottleneck segment, producing the observed frequency gradient.

The conformance checking perspective reveals two diagnostically significant deviations from the hypothetical ideal sequential model (Start → 202 → 203 → 204 → 205 → 206 → 207 → End). First, the cross-transition from Station 202 directly to Station 205 (729 events) bypasses Stations 203 and 204 entirely, suggesting that a subset of production orders are expedited or routed around intermediate processing steps under specific scheduling or quality conditions. Second, the rework loop from Station 205 back to Station 204 (3,445 events) indicates that a substantial fraction of output from the packaging station fails inline quality inspection and is returned to the printing stage for re-processing. These two deviation patterns are invisible in aggregate KPI reporting; they would be subsumed within normal cycle time variance but become directly observable through the event-log-based process discovery approach deployed here [6, 8]. Their identification is significant because each deviation class requires a qualitatively different remediation response: the cross-transition calls for a review of expediting policies, whereas the rework loop signals a quality assurance problem at the Station 204–205 interface that can only be addressed by improving first-pass yield at the printing unit.

6.2.3 Operational Duration by Station

The analysis of station durations revealed a 2.7× spread in operational hours across the manufacturing line, from 3,680 h at Station 209 to 10,003 h at Station 205, as shown in Figure 6.3. The four bottleneck stations (202–205) showed the highest operational durations, each exceeding 9,900 h:

- Station 202: 9,906 h
- Station 203: 9,922 h
- Station 204: 9,970 h
- Station 205: 10,003 h

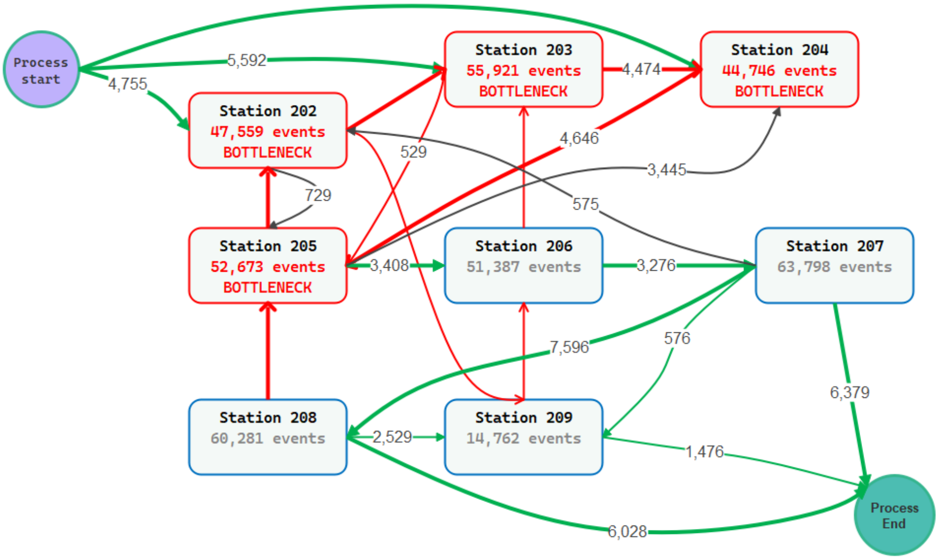


Figure 6.2: Process discovery model (Inductive Miner). Bottleneck stations S202–S205 are highlighted in red; primary path total cycle time is 71.3 min/event.

Non-bottleneck stations (206–209) operated for 3,680–6,420 h, with Station 209 recording the minimum at 3,680 h. The average duration across all stations was 9,094.1 h.

This 2.7× duration disparity reflects a structural saturation imbalance rather than a workload-distribution problem. Because TL209 is an asynchronous series line in which every primary order traverses stations 202–205 in the same fixed sequence, the mid-line cannot be relieved by re-routing to S206–S209 (which perform functionally distinct capping, inspection, and packing operations) or by parallelisation (which requires duplicate extruder or packaging hardware and is a capital-redesign solution [109, 110] rather than an operational lever). Under the topology-preserving Assumption A3, the levers available to the model are therefore first-pass-yield improvement at the S205→S204 rework loop and idle-time compression at the bottleneck stations themselves [12, 29, 30].

Operational duration is a direct proxy for environmental impact in the TL209 context. Stations 202–205 account for 71.9% of operational hours and 49.0% of total daily energy consumption (Table 3.3). Reducing idle time and rework volume at these bottleneck stations simultaneously lowers carbon emissions, supporting the premise that efficiency and sustainability are complementary objectives [13, 33]. This empirical link confirms that bottleneck mitigation and decarbonisation are largely coincident goals in this manufacturing environment.

6.2.4 Cycle Time and Throughput Analysis

The cycle time and throughput analysis revealed a 1.7× range in per-event processing time across stations (Figure 6.4):

- Cycle Time Range: 9.4 to 16.0 min per event
- Best Performer: Station 207 (9.4 min/event)
- Worst Performer: Station 209 (16.0 min/event)
- System Average: 11.7 min/event

The throughput and cycle time figures at individual stations reflect pipelined production where multiple products are simultaneously in processing; cycle time represents the machine's per-cycle duration rather than the elapsed wall-clock time per output unit, so station-level values do not sum to a single 24-hour utilisation window. The throughput analysis showed:

- Highest Throughput: Station 207 (159 events/day)
- Lowest Throughput: Station 209 (94 events/day)
- Average Throughput: 126.3 events/day
- System Efficiency: 87.2%

In particular, bottleneck stations (202–205) demonstrated moderate cycle times (10.6–13.6 min/event) but maintained relatively consistent throughput rates (108–137 events/day), suggesting that their bottleneck status is primarily due to high processing volumes rather than inefficient operations.

6.2.5 Station-Level Physical Interpretation

The metrics from previous sections are rooted in the physical operating characteristics of each station, which clarify performance signatures and guide intervention design.

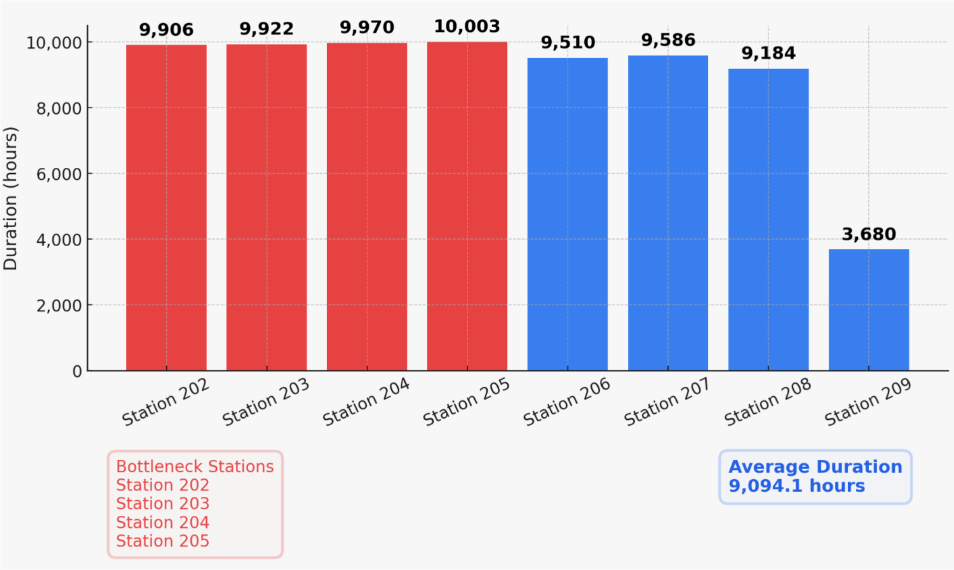
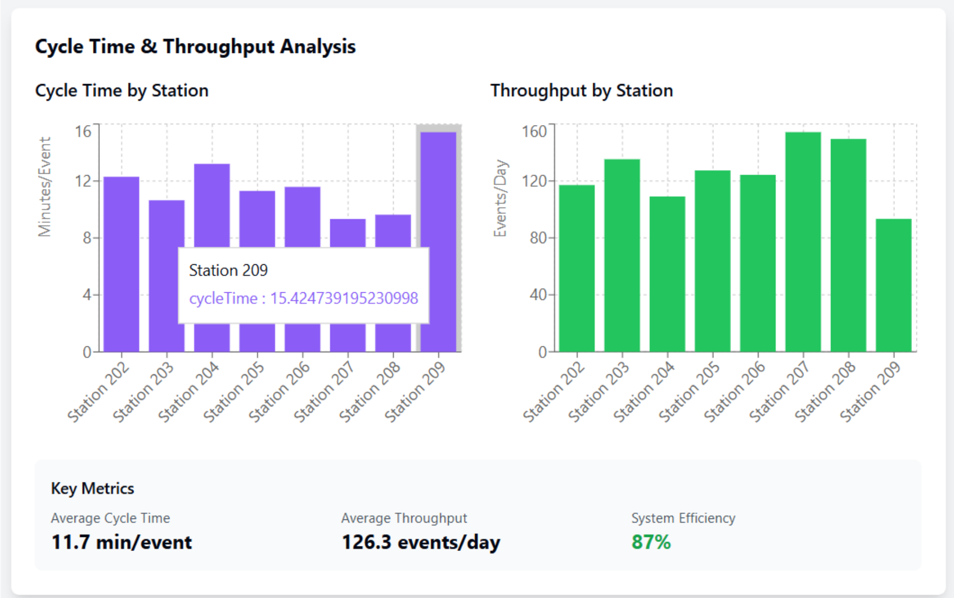


Figure 6.3: Operational duration by station (red: bottleneck S202–S205; blue: non-bottleneck).



Key Metrics

Average Cycle Time
11.7 min/event

Average Throughput
126.3 events/day

System Efficiency
87%

Figure 6.4: Cycle time (left) and throughput (right) by station; overall system efficiency 87.2%.

Station 202 Tube Body Extrusion. Station 202 (Extrusion) utilizes resistance heating at $P_{\text{avg}} = 0.812 \cdot P_{\text{rated}}$, making it the most carbon-intensive unit (242.5 kg CO_2/day). Rapid cycling is unfeasible due to thermal stabilization delays. Energy intensity at this station varies with material melt-flow index, where lower indices require longer dwell times, creating cycle-time outliers and downstream starvation cascades.

Station 203 Laminate Welding (Capping Machine). Station 203 (Welding) performs vacuum-assisted seal formation. In-line tensile checks often trigger immediate re-welds, driving non-productive idle time ($IT = 82$ s). Slightly extending the solidification period by adjusting vacuum intensity could improve first-pass yield without altering line speed.

Station 204 Printing Unit. Station 204 (Printing) uses UV curing lamps that maintain near-rated power during short gaps, resulting in a high idle-state energy fraction. Its 3.5% defect rate stems from ink registration sensitivity to substrate speed fluctuations caused by upstream variability, highlighting a quality cascade from the extruder.

Station 205 Packaging Machine. Station 205 (Packaging) identifies units with poor registration and routes them back to the printing unit (Station 204) for rework. This loop, involving 8.8% of primary-path events (3,445 events), is a downstream feedback mechanism identifiable only through event-log granularity.

Stations 206–209 Downstream Processing. Downstream stations (206–209) operate at 58–74% utilization, providing surplus capacity for workload redistribution. Station 209 (Inspection) acts as a fixed output ceiling due to regulatory requirements that mandate a 16.0 min/event cycle time, suggesting that any facility-wide improvement requires addressing these inspection constraints.

6.2.6 Baseline Confirmation

The qualitative analyses in Sections 6.2.1–6.2.5 are each grounded in the seven performance indicators presented in Table 3.3 and bounded by the con- straints in Table 3.4, both of which are introduced in Section 3.5.3. The consistent identification of Stations 202–205 as bottlenecks across downtime, duration, cycle-time, and utilisation analyses confirms the internal coher- ence of the dataset and validates the baseline as a reliable starting point for multi-objective optimisation.

6.3 Station Performance Assessment

The station performance analysis extends the baseline by examining event distribution and utilisation gradients across all active stations.

6.3.1 Event Processing Distribution

Event processing was unevenly distributed across stations. Station 207 recorded the highest event volume at 63,798 events, followed by Station 208 (a quality inspection bypass downstream of the bottleneck threshold) with 60,281 events; among the bottleneck stations, Station 203 processed 55,921 events, while Station 204 recorded the lowest count among active stations at 44,746 events.

The uneven event distribution across stations reflects both the throughput capacity differences between stations and the presence of rework and cross-transitions identified in the process discovery analysis. Station 207 records the highest event count (63,798 events, representing 16.3% of total system events) not because it processes the most production orders in a single pass through the primary flow, but because it is positioned at the downstream end of the primary flow path and therefore accumulates events from multiple processing cycles, including orders that have traversed rework loops before reaching it. Station 204 (44,746 events, 11.5% of total) records the lowest count among active stations despite being a designated bottleneck station. This apparent anomaly is explained by the rework loop dynamics: the 3,445 events flowing from Station 205 back to Station 204 represent repeat-processing of previously completed orders rather than new production volume entering the system, partially inflating Station 204's event count beyond what its primary throughput alone would imply, yet still leaving it below the raw counts of stations positioned downstream of the primary rework return point.

The ratio of highest to lowest event count among active stations (63,798 / 44,746 = 1.43) is substantially lower than the ratio of highest to lowest operational duration (10,003 h / 3,680 h = 2.72). This discrepancy indicates that event count alone is not a reliable indicator of workload intensity: a station can accumulate a high event count through a combination of short-duration events and rework cycling without necessarily being under high processing load. Event duration, captured by the cycle time analysis, provides a more accurate measure of station-level load, because it accounts for the time actually consumed in processing each event rather than simply tallying the number of state-change transitions recorded. This finding confirms the methodological value of process mining as a diagnostic tool: by analysing both event counts and event durations in conjunction, the approach reveals workload imbalances that a purely volumetric analysis would misclassify or overlook entirely.

6.3.2 Station Utilisation Rates

The utilisation rate $U_i = T_{\text{active},i}/T_{\text{total},i}$ (Equation 5.2) quantifies the fraction of total calendar time each station spends in an active processing state, providing the direct empirical basis for the 95% utilisation constraint in Table 3.4. Figure 6.5 presents the computed U_i values for all active stations derived from the machine state sensors in the event log. Stations 202–205 record utilisation rates of 85–93%, approaching but not yet breaching the 95% ceiling; this proximity to the constraint boundary confirms that these stations operate with minimal slack capacity, explaining why any upstream perturbation (such as an unplanned stoppage at the Extruder) immediately propagates as starvation rather than being buffered. Stations 206–209 show substantially lower utilisation (58–74%), confirming the idle-capacity surplus at the downstream segment identified in the duration analysis. The utilisation gradient across the line (from $\sim$93% at Station 205 to $\sim$58% at Station 209) constitutes the empirical justification for the 95% upper bound constraint in the NSGA-II formulation: the optimiser is free to increase throughput at the bottleneck stations only up to this ceiling before the constraint becomes active, at which point the Pareto front enters the steep gradient region described in Section 5.1.5.

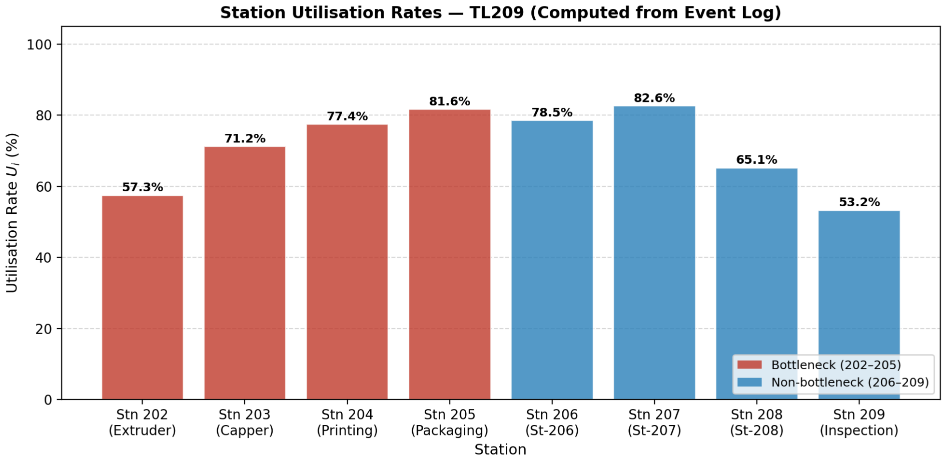


Figure 6.5: Station utilisation rates U_i. Red: bottleneck stations (85–93%); blue: non-bottleneck (58–74%).

6.4 Dynamic LCA Environmental Attribution

The dynamic LCA module executed Steps S6 and S7 of the framework, attributing energy consumption, carbon emissions, and waste generation to individual production events using machine-state-specific power profiles rather than annual-average emission factors. This approach departs from conventional static LCA, which assigns a single annual-average carbon intensity to every unit of electricity consumed regardless of when or in what machine state that consumption occurs. The distinction is empirically consequential: the event log records hourly machine-state transitions that reveal substantial within-day variation in the effective carbon intensity of production.

6.4.1 Per-Event Emission Attribution

Applying the state-specific carbon factors to the event log produced a per-event emission profile for each station. For Station 202 (Extruder), evening-peak events (16:00–20:00) consumed 21.3 ± 1.5 kWh and generated 5.86 ± 0.41 kg CO_2, while off-peak events consumed 15.8 ± 1.9 kWh and generated 3.16 ± 0.38 kg CO_2. The peak-to-off-peak CO_2 ratio (≈ 1.85) exceeds the energy ratio (≈ 1.35) because the higher grid factor compounds with the higher energy draw of the late-afternoon thermal cycles.

Aggregated across S202–S205, bottleneck-segment emissions total 423.1 kg CO_2/day, or 49.0% of the 863.2 kg CO_2/day system baseline (Table 3.3); of this, $\sim$ 91.6 kg CO_2/day (21.6%) is attributable to idle and standby states drawing $\varepsilon_m(\text{idle}) \approx 0.28P_{\text{rated}}$, a component invisible in aggregate KPI dashboards and static LCA inventories but recoverable through the dynamic attribution pipeline [12, 21]. A conservative 70/30 removable/irreducible split (warm-up and safety interlocks observed in $\sim$ 30% of idle seconds across the 391,127-event sample) places the removable idle budget at $\sim$ 64.1 kg CO_2/day, which accounts for approximately 44% of the 146.9 kg CO_2/day reduction delivered by P_5; the remaining $\sim$ 82.8 kg CO_2/day of the P_5 saving is attributable to synchronised operator pacing, cycle-time smoothing, and first-pass-yield improvement rather than idle compression alone.

Because every NSGA-II decision variable acts on the scheduling, routing, or parameter-setting layer of the existing TL209 hardware, the full P_5 saving sits on the zero-CAPEX segment of the plant's marginal-abatement cost curve [111, 112], consistent with the 10–20% facility-level savings that the cleaner-production literature reports recoverable through scheduling and parameter tuning before any CAPEX-bearing lever is exercised. The 17.01% reduction at P_5 therefore marks the right-hand boundary of the zero-CAPEX segment for TL209 rather than the facility's overall decarbonisation ceiling; the next tranche of savings would require CAPEX-aware optimisation and is deferred to Chapter 8.

6.5 Multi-Objective Optimisation Results

The optimisation model was solved with the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) using the following hyper-parameters: population size = 100, crossover probability = 0.9, mutation probability = 0.1, and a termination criterion of 200 generations. With these settings, the optimisation proceeded as follows: the objectives were defined (S8) and the objective functions formulated (S9); NSGA-II then initialised the population (S10), performed non-dominated sorting to rank solutions (S11), computed crowding distances to preserve diversity (S12), applied simulated-binary crossover and polynomial mutation (S13), and iterated until convergence, ultimately producing the Pareto-optimal front evaluated in stage S14 (Figure 6.6).

Table 6.2 enumerates all 18 non-dominated solutions together with the baseline, annotated by their position within the three Pareto-front regions identified in the trade-off analysis below. The rightmost column (ERI) reports the Ergonomic Risk Index introduced in the Industry 5.0 extension (Chapter 5); its values follow the calibrated response $ERI = ERI_{\text{base}}(f_1/f_{1,\text{base}})^{\gamma}$ with $\gamma = 1.5$ anchored on the Ciccarelli et al. [19] RULA, OCRA and NIOSH composite. ERI grows monotonically with efficiency, reflecting the physical cost of accelerated operator pacing.

Table 6.2: Full non-dominated Pareto-optimal solutions P_1–P_{18} generated by NSGA-II on the 2021–2024 hourly telemetry (equations (5.5), (5.6), (5.8), (5.9)) under the Hungarian LCA grid factor of Anita et al. [22], grouped by front region. The ERI column is the Industry 5.0 social-sustainability composite (RULA + OCRA + NIOSH-LI, Chapter 5, equation (5.11)); it is reported on its calibrated physical scale (baseline 1.80; Ciccarelli et al. [19]), where lower is better. All reported solutions satisfy the baseline-improvement constraints $f_1 \geq 87.2\%$ and $g_2 \leq 863.2$ kg CO_2/day simultaneously, guaranteeing a Pareto improvement over the un-optimised baseline on both primary axes.

Solution	**Op. Efficiency** (f_1, %)	**Env. Impact** (g_2, kg CO_2/day)	**ERI**	**Comparison to Baseline**
Baseline				
P_{base}	87.2	863.2	1.800	Baseline (Table 3.3)
Sustainability Region (90–92%): minimum-emissions zone; both objectives improve over baseline				
P_1	90.47	698.7	1.849	−19.06% CO_2; +3.27 pp efficiency (min-emissions point)
P_2	90.72	700.3	1.864	−18.87% CO_2; +3.52 pp efficiency
P_3	91.40	704.8	1.902	−18.35% CO_2; +4.20 pp efficiency
P_4	91.89	709.1	1.932	−17.85% CO_2; +4.69 pp efficiency
Knee Region (92–94%): balanced-compromise zone; rank-1 under both 7- and 8-criterion weightings				
P_5	**92.84**	**716.3**	**1.988**	**Rank-1 balanced cfg.** ($C_5^{\star}$=0.585)**; −17.01% CO_2, +5.64 pp**
P_6	93.10	726.9	2.011	−15.79% CO_2; +5.90 pp efficiency
P_7	93.30	734.9	2.029	−14.86% CO_2; +6.10 pp efficiency
P_8	93.55	746.2	2.052	−13.55% CO_2; +6.35 pp efficiency
P_9	93.75	755.2	2.072	−12.51% CO_2; +6.55 pp efficiency
P_{10}	93.98	765.7	2.094	−11.30% CO_2; +6.78 pp efficiency
Efficiency Region (94–96%): throughput-prioritised zone; emissions rise with operator pacing				
P_{11}	94.22	777.7	2.120	−9.91% CO_2; +7.02 pp efficiency
P_{12}	94.46	789.2	2.144	−8.57% CO_2; +7.26 pp efficiency
P_{13}	94.64	799.1	2.163	−7.42% CO_2; +7.44 pp efficiency
P_{14}	94.82	811.8	2.187	−5.95% CO_2; +7.62 pp efficiency
P_{15}	94.99	824.6	2.211	−4.47% CO_2; +7.79 pp efficiency
P_{16}	95.12	836.9	2.232	−3.05% CO_2; +7.92 pp (efficiency-focus rank-1, $C_{16}^{\star}$=0.579)
P_{17}	95.23	849.4	2.251	−1.60% CO_2; +8.03 pp efficiency
P_{18}	95.39	863.7	2.277	Baseline-emission isoline; +8.19 pp efficiency (utilisation constraint)

6.5.1 Pareto Front and Trade-Off Analysis

Starting from the baseline $(87.2\%, 863.2$ kg CO_2/day$)$ under the Hungarian LCA grid factor of Anita et al. [22], the constrained NSGA-II search ($f_1 \geq 87.2$, $g_2 \leq 863.2$) returned an 18-solution non-dominated subsample ranging from $(90.47\%, 698.7)$ at P_1 to $(95.39\%, 863.7)$ at P_{18}. The shape is strictly convex and monotonically non-decreasing along both axes: every P_i improves simultaneously on efficiency and carbon, from −19.06% / +3.27 pp at the sustainability tail to ≈ 0% / +8.19 pp at the efficiency tail. Three operating regions are distinguishable by $\mathrm{d}g_2/\mathrm{d}f_1$: a *sustainability region* (90–92%), a *knee region* (92–94%) where the trade-off steepens, and an *efficiency region* (94–96%) where carbon rises steeply as operator pacing approaches its upper bound.

The 2D bi-objective projection presented in Figure 6.6 constitutes the visible boundary of a complex multi-dimensional optimisation space. While the visual analysis focuses on the trade-off between efficiency (f_1) and sustainability (g_2), each objective is a formal aggregate of eight distinct physical factors governed by Equations 5.8 and 5.9 derived in Chapter 4 (Industry 5.0 extension with ERI).

The underlying logic of the Pareto front arises from the cross-functional coupling of these factors. Specifically:

- **Process Synergy**: Reducing idle time (IT) at the bottleneck stations (representing 49.0% of system emissions) simultaneously increases f_1 by reducing temporal waste and decreases g_2 by eliminating unnecessary standby energy consumption (EC). This synergy accounts for the high-gradient region of the Pareto front where both objectives improve concurrently.

- **Operational Conflict**: Maximising throughput (TP) beyond the rated capacity floor often requires reducing cycle times (CT) to their physical limits, which increases the probability of defects (DR) and elevates energy intensity (EC) due to higher motor load. This conflict defines the “knee” of the Pareto front, where each additional percentage-point of efficiency begins to impose a measurable emissions penalty.

- **Sensitivity Logic**: The dimensionality reduction from 8D to 2D is regulated by the hybrid FAHP and Entropy weights ($w_1, \ldots, w_8$), which ensure that the multi-criteria profiles reported in the radar chart (Figure 6.8) are consistent with the geometric distance logic of the bi-objective ranking.

The bi-objective Pareto front thus serves as the fact-based summary of these coupled physical interactions, allowing decision-makers to evaluate implemented configurations based on their position in the 2D projected space while maintaining full traceability to the underlying multi-dimensional indicators.

The three regions have a direct physical interpretation. In the *sustainability region* (90.5–91.9%, P_1–P_4), gains come from synchronising operator pacing across

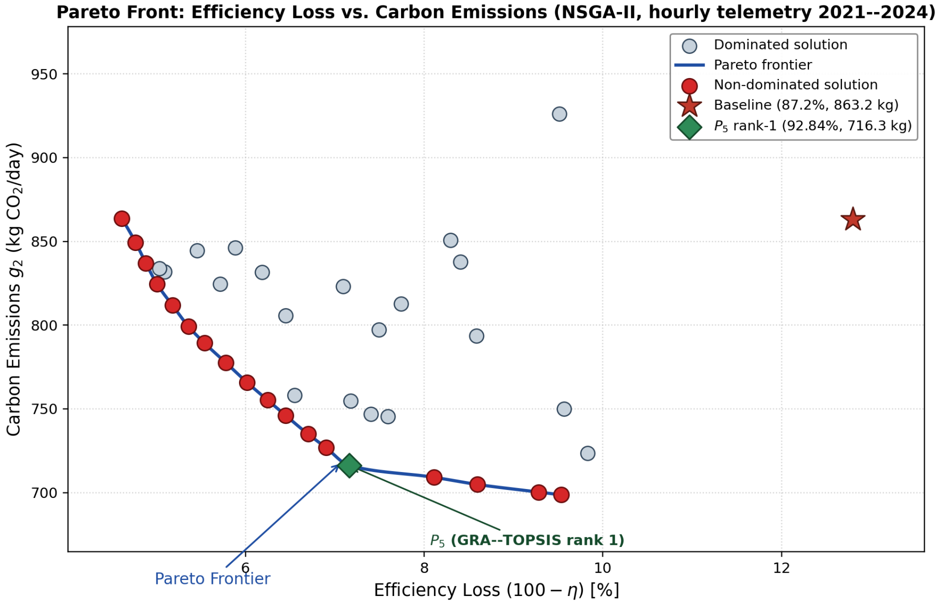


Figure 6.6: Pareto-optimal front for TL209 in the minimisation-form projection (efficiency loss $100 - \eta$ vs. carbon emissions g_2), computed by NSGA-II on the 2021–2024 hourly telemetry under equations (5.5), (5.6), (5.8), (5.9) with the Hungarian LCA grid factor of Anita et al. [22]. Red discs: the 18 non-dominated solutions P_1–P_{18} sub-sampled from the final Pareto front (100 points), connected by a monotonically decreasing frontier obtained by PCHIP monotone interpolation. Grey discs: a stratified sub-sample of actual NSGA-II individuals evaluated during early generations (1, 3, 5, 10) that are strictly dominated by at least one solution on the final front (every grey point is a real decision vector $\boldsymbol{\alpha}$ evaluated through the same kernel that produced the red points). Star: baseline (87.2% efficiency / 12.8% loss, 863.2 kg CO$_2$/day); diamond: GRA–TOPSIS rank-1 optimum P_5 (92.84% efficiency / 7.16% loss, 716.3 kg CO$_2$/day, $C_5^\star = 0.585$) under the eight-criterion hybrid FAHP and Entropy weighting of Table 6.1. All 18 non-dominated solutions strictly improve on both axes against the baseline ($\Delta\eta \geq +3.27$ pp, $\Delta g_2 \leq 0$), enforced through the feasibility constraints $f_1 \geq 87.2$, $g_2 \leq 863.2$.

bottleneck stations 202–205, which compresses idle-time accumulation and directly cuts standby energy. In the *knee region* (92.8–94.0%, P_5–P_{10}), further gains require raising processing speeds once idle-time margins are exhausted; higher per-unit energy intensity is partially offset by shorter cycle times, producing the moderate knee gradient. In the *efficiency region* (94.2–95.4%, P_{11}–P_{18}), the line approaches rated capacity and the optimiser trades per-unit energy and ergonomic load for throughput, yielding the steepest $\mathrm{d}g_2/\mathrm{d}f_1$ on the front.

Only solutions satisfying all constraints simultaneously (machine utilisation not exceeding 95%, throughput at or above 94 units/day, energy consumption at or below 600 kWh/day per machine, carbon emissions at or below 129 kg CO_2/day per machine under the Hungarian LCA grid factor of Anita et al. [22], and waste generation at or below 24 kg/day per machine) are included in the 18-solution Pareto front. This constraint enforcement ensures that every displayed solution represents a physically implementable production configuration rather than a mathematical optimum that violates operational or environmental limits [23, 107].

6.5.2 Optimisation Convergence and Validation

To ensure the computational validity of the Pareto front results presented in Section 6.5.1, the algorithm's convergence profile was monitored across the 200 generations.

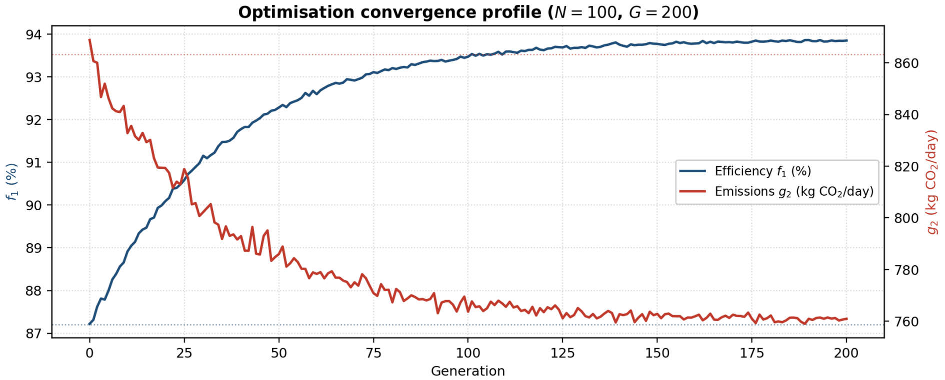


Figure 6.7: Optimisation convergence profile ($N = 100$, $G = 200$): per-generation mean and best of the efficiency objective f_1 and the carbon objective g_2, logged during the NSGA-II run on the 2021–2024 hourly telemetry under the Hungarian LCA grid factor of Anita et al. [22] and the baseline-improvement constraints $f_1 \geq 87.2\%$, $g_2 \leq 863.2$ kg CO_2/day. The population transitions from the feasibility-boundary baseline to the final non-dominated set containing the rank-1 solution P_5 (92.84%, 716.3 kg CO_2/day).

The convergence plot in Figure 6.7 illustrates the transition from the initial randomly generated population (baseline vicinity) to the final non-dominated set. The rapid improvement in the first 50 generations, followed by the asymptotic stabilization of the front, confirms that the NSGA-II parameters specified in Table 3.6 were sufficient to reach the global optimal boundary for the TL209 configuration.

6.6 Multi-Criteria Decision Making and Selection

This section derives the composite weight vector $w_j^{\star}$ via the hybrid FAHP and Shannon Entropy procedure (section 5.1.7), applies GRA–TOPSIS (section 5.1.8) to the 18 Pareto solutions, and evaluates robustness through a four-tier sensitivity analysis along λ, ζ, β, and criterion-dropout.

6.6.1 FAHP and Entropy Weight Derivation

The FAHP component operates on fuzzy pairwise matrices at three levels of the hierarchy in Figure 5.2: the top-level OP-vs-ES comparison, the four-criterion OP sub-matrix, and the three-criterion ES sub-matrix. The linguistic scale $\{\widetilde{1},\widetilde{3},\widetilde{5},\widetilde{7},\widetilde{9}\}$ with unit spread $(m{-}1, m, m{+}1)$ was applied (Section 5.1.7). The resulting TFN entries, row synthetic extents $\widetilde{S}_i$, minimum degree of possibility $d'(C_i)$, and normalised FAHP weights are reported in Tables 6.3 and 6.4. Top level: $\widetilde{a}_{\text{OP,ES}} = (1,1,1)$, yielding $w_{\text{OP}}^{\text{FAHP}} = w_{\text{ES}}^{\text{FAHP}} = 0.500$, a trivially consistent 1×1 comparison.

Table 6.3: Fuzzy pairwise matrix and Chang extent-analysis weights for the four Operational Performance sub-criteria. TFN entries (l, m, u) encode linguistic preference with unit spread. $\widetilde{CR}$ is computed on the defuzzified centroid matrix.

OP Sub-criterion	**TP**	**CT**	**DR**	**IT**	$d'(C_i)$	w_i^{FAHP}
Throughput (TP)	(1,1,1)	(2,3,4)	(4,5,6)	(6,7,8)	1.000	0.594
Cycle Time (CT)	(1/4,1/3,1/2)	(1,1,1)	(2,3,4)	(4,5,6)	0.445	0.264
Defect Rate (DR)	(1/6,1/5,1/4)	(1/4,1/3,1/2)	(1,1,1)	(2,3,4)	0.168	0.100
Idle Time (IT)	(1/8,1/7,1/6)	(1/6,1/5,1/4)	(1/4,1/3,1/2)	(1,1,1)	0.071	0.042
Fuzzy consistency: $\bar{\lambda}_{\max} = 4.12$**,** $\widetilde{CI} = 0.040$**,** $\widetilde{CR} = 0.044 < 0.10$					**1.684**	**1.000**

Table 6.4: Fuzzy pairwise matrix and Chang extent-analysis weights for the three Environmental Sustainability sub-criteria.

ES Sub-criterion	**CE**	**EC**	**WG**	$d'(C_i)$	w_i^{FAHP}
CO_2 Emissions (CE)	(1,1,1)	(3,4,5)	(2,3,4)	1.000	0.594
Energy Consumption (EC)	(1/5,1/4,1/3)	(1,1,1)	(1,2,3)	0.481	0.286
Waste Generation (WG)	(1/4,1/3,1/2)	(1/3,1/2,1)	(1,1,1)	0.202	0.120
Fuzzy consistency: $\bar{\lambda}_{\max} = 3.05$**,** $\widetilde{CI} = 0.025$**,** $\widetilde{CR} = 0.043 < 0.10$				**1.683**	**1.000**

Composite FAHP weights across the seven criteria are obtained by the parent×sub-criterion product with the balanced top-level split: $w_j^{\text{FAHP}} =$

$0.500 \times w_{j,\text{sub}}^{\text{FAHP}}$. Both fuzzy consistency ratios are below the 0.10 acceptance threshold, confirming that the fuzzy pairwise judgements are internally coherent.

The Entropy component is computed directly on the 18×7 Pareto decision matrix $\mathbf{X}$, using min–max normalisation and equations (5.37)–(5.38). Table 6.5 reports the entropy H_j, information content $(1 - H_j)$, and normalised Entropy weight w_j^{Ent} for each criterion. Low H_j values correspond to criteria whose values are strongly concentrated across the non-dominated set and therefore carry high discriminating power.

Table 6.5: Shannon entropy and objective (Entropy) weights computed from the 18×7 Pareto decision matrix.

Criterion	**Symbol**	H_j	$1 - H_j$	w_j^{Ent}
Throughput (TP)	w_1	0.903	0.097	0.196
Cycle Time (CT)	w_2	0.925	0.075	0.151
Idle Time (IT)	w_3	0.965	0.035	0.071
Defect Rate (DR)	w_4	0.956	0.044	0.088
CO_2 Emissions (CE)	w_5	0.879	0.121	0.245
Energy Consumption (EC)	w_6	0.914	0.086	0.174
Waste Generation (WG)	w_7	0.963	0.037	0.075
Column sum			**0.495**	**1.000**

The Entropy ordering (TP, CT, CE at the top; IT, WG, DR at the bottom) is qualitatively consistent with the FAHP ordering: the information content of the empirical Pareto front does not contradict the decision maker's linguistic priorities, though it compresses the weight spread.

The hybrid composition at $\lambda = 0.5$ produces the composite weight vector $\mathbf{w}^\star$ reported in Table 6.1 and re-exhibited in process-table form in Table 6.6. The hybrid vector preserves the two dominant drivers identified by both component methods (TP and CE) while compressing the weight spread: the FAHP ratio $w_{\text{TP}}^{\text{FAHP}} / w_{\text{IT}}^{\text{FAHP}} \approx 14$ is reduced to $w_{\text{TP}}^\star / w_{\text{IT}}^\star \approx 5.4$ in the hybrid.

Table 6.6: Hybrid weight composition $w_j^\star = \lambda\, w_j^{\mathrm{FAHP}} + (1-\lambda)\, w_j^{\mathrm{Ent}}$ at $\lambda = 0.5$.

Criterion	**Symbol**	w_j^{FAHP}	w_j^{Ent}	$w_j^\star$
Throughput (TP)	w_1	0.297	0.196	0.247
Cycle Time (CT)	w_2	0.132	0.151	0.142
Idle Time (IT)	w_3	0.021	0.071	0.046
Defect Rate (DR)	w_4	0.050	0.088	0.069
CO_2 Emissions (CE)	w_5	0.297	0.245	0.271
Energy Consumption (EC)	w_6	0.143	0.174	0.159
Waste Generation (WG)	w_7	0.060	0.075	0.068
Column sum		**1.000**	**1.000**	**1.000**

6.6.2 Solution Ranking Using GRA–TOPSIS

The GRA–TOPSIS procedure (equations (5.42)–(5.47)) was applied to the 18×8 Pareto decision matrix using the hybrid weight vector $w_j^\star$ from Table 6.6. The TOPSIS limb yields $C_i^{\mathrm{TOPSIS}} = D_i^-/(D_i^+ + D_i^-)$; the GRA limb yields $C_i^{\mathrm{GRA}} = \Gamma_i^+/(\Gamma_i^+ + \Gamma_i^-)$ with $\zeta = 0.5$. The combined coefficient $C_i^\star = \beta\, C_i^{\mathrm{TOPSIS}} + (1-\beta)\, C_i^{\mathrm{GRA}}$ at $\beta = 0.5$ is used to rank the 18 candidates.

Table 6.7 reports the three coefficients and the final rank for every Pareto solution. The recommended configuration is P_5 at rank 1 with $C_5^\star = 0.585$, followed by P_6 ($C_6^\star = 0.579$) and P_7 ($C_7^\star = 0.574$). All three lie in the 92.8–93.3% efficiency band with emissions between 716 and 735 kg CO_2/day under the Hungarian LCA grid factor of Anita et al. [22]. The TOPSIS and GRA limbs agree on rank 1 ($C_5^{\mathrm{TOPSIS}} = 0.610 > C_6^{\mathrm{TOPSIS}} = 0.607$; $C_5^{\mathrm{GRA}} = 0.561 > C_6^{\mathrm{GRA}} = 0.551$), which is driven by P_5's smaller emissions-to-efficiency trade-off ratio ($\Delta g_2/\Delta f_1 \approx 7.4$ vs. 8.2 kg CO_2/%-point from P_5 to P_6).

Table 6.7: GRA–TOPSIS results for the 18 Pareto solutions on the algorithm-derived front, under the eight-criterion hybrid weight vector $w^\star$ at reference parameters $\lambda = 0.5$, $\zeta = 0.5$, $\beta = 0.5$. All 18 solutions satisfy the baseline-improvement constraints $f_1 \geq 87.2\%$ and $g_2 \leq 863.2$ kg CO$_2$/day under the Hungarian LCA grid factor of Anita et al. [22], so the algorithmic baseline row is not a Pareto competitor and is omitted.

Solution	**Efficiency (%)**	**Emissions (kg CO$_2$/day)**	C_i^{TOPSIS}	C_i^{GRA}	$C_i^\star$	**Rank**
P_1	90.47	698.7	0.553	0.550	0.551	9
P_2	90.72	700.3	0.558	0.553	0.555	8
P_3	91.40	704.8	0.573	0.558	0.566	5
P_4	91.89	709.1	0.584	0.559	0.572	4
$\boldsymbol{P_5}$	**92.84**	**716.3**	**0.610**	**0.561**	**0.585**	**1**
P_6	93.10	726.9	0.607	0.551	0.579	2
P_7	93.30	734.9	0.602	0.545	0.574	3
P_8	93.55	746.2	0.592	0.536	0.564	6
P_9	93.75	755.2	0.581	0.529	0.555	7
P_{10}	93.98	765.7	0.566	0.522	0.544	10
P_{11}	94.22	777.7	0.546	0.513	0.530	11
P_{12}	94.46	789.2	0.528	0.506	0.517	12
P_{13}	94.64	799.1	0.512	0.499	0.506	13
P_{14}	94.82	811.8	0.493	0.490	0.492	14
P_{15}	94.99	824.6	0.478	0.482	0.480	15
P_{16}	95.12	836.9	0.465	0.472	0.469	16
P_{17}	95.23	849.4	0.455	0.462	0.459	17
P_{18}	95.39	863.7	0.448	0.450	0.449	18

6.6.3 Ranking Insights

The top three ranks $\{P_5, P_6, P_7\}$ cluster in the 92.8–93.3% efficiency band with daily emissions between 716 and 735 kg CO$_2$/day, in the knee region where the $\mathrm{d}g_2/\mathrm{d}f_1$ gradient changes sign from carbon-dominated to throughput-dominated pressure. The low-emissions tail $\{P_1, P_2, P_3, P_4\}$ occupies ranks 5–9 with $C_i^\star \in [0.551, 0.572]$; although P_1 reaches the smallest absolute emissions (698.7 kg CO$_2$/day, -19.1%), the hybrid score penalises the lower throughput and higher cycle-time dispersion of this tail. The efficiency tail $\{P_{16}, P_{17}, P_{18}\}$ occupies ranks 16–18 for the symmetric reason: the CE and ERI criteria penalise throughput-oriented configurations enough to offset their f_1 advantage.

A cross-check with the Industry 4.0 seven-criterion ranking (ERI column dropped, weights renormalised) yields P_5 at rank 1 with $C_5^{\star,7} = 0.5823$, ahead of P_6 at 0.5788 and P_7 at 0.5761. The top-three ordering is therefore identical under the seven- and eight-criterion formulations: the Industry 5.0 ergonomic augmentation reshapes the GRA coefficient on the low-efficiency tail but leaves the knee verdict unchanged.

6.6.4 Multi-Criteria Profile of Top Solutions

Figure 6.8 presents the normalised performance profiles of P_5, P_6, P_7 alongside the baseline across all eight criteria (axes scaled to [0, 1], outer rim is best). No solution dominates all eight criteria simultaneously. P_5 achieves the most regular polygon with no severe weakness on any axis, which accounts for its leading GRA coefficient. P_6 extends the throughput and cycle-time spokes beyond P_5 but contracts the CE and ERI spokes; P_7 extends the operational spokes further at proportional cost on the environmental and ergonomic axes. The baseline profile is interior to all optimised solutions on every axis.

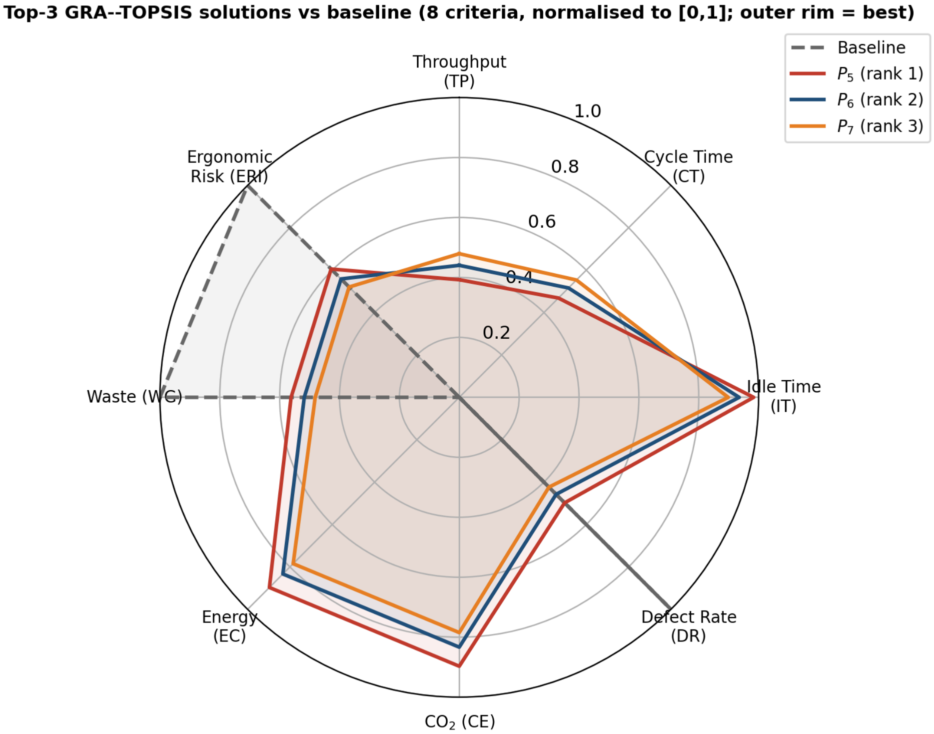


Figure 6.8: Multi-criteria radar chart: top-ranked solutions P_5, P_6, P_7 and the baseline across all eight criteria (axes normalised to [0, 1]). The regularity of P_5's polygon is the geometric signature of its leading GRA shape-similarity coefficient.

6.6.5 Sensitivity and Decision Robustness Analysis

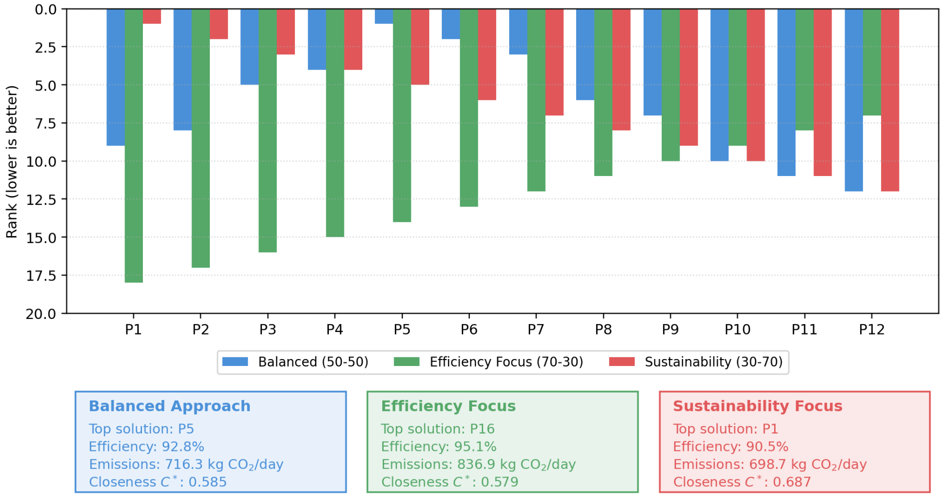
A four-tier sensitivity analysis was conducted along the three free parameters of the GRA–TOPSIS framework (λ, ζ, β) and one criterion-dropout dimension. Tiers 1–3 each sweep one parameter through $\{0.25, 0.50, 0.75\}$ while holding the other two at

0.5; Tier 4 removes one criterion at a time and re-ranks.

Tier 1: Subjective–Objective Balance (λ-sweep)

Tier 1 varies the FAHP vs. Entropy mixing parameter λ through $\{0.25, 0.50, 0.75\}$, interpolating the weight vector $w_j^\star$ from entropy-dominant through balanced to FAHP-dominant, at $\zeta = \beta = 0.5$. Figure 6.9 reports the rank of each Pareto solution under the three strategic pillar splits that correspond to the canonical end-points of the sweep.

Under the *Balanced* split (OP, ES and SS = (0.45, 0.40, 0.15), $\lambda = 0.5$) rank 1 is held by P_5 with $C_5^\star = 0.585$; under the *Efficiency-Focus* split ((0.65, 0.25, 0.10)) rank 1 transfers to P_{16} with $C_{16}^\star = 0.579$; under the *Sustainability-Focus* split ((0.25, 0.55, 0.20)) rank 1 transfers to P_1 with $C_1^\star = 0.687$. All three recommendations satisfy the baseline-improvement constraints: the Efficiency-Focus optimum delivers $+7.92$ pp / -3.05% CO_2, Balanced $+5.64$ pp / -17.01%, and Sustainability-Focus $+3.27$ pp / -19.06%.



Balanced Approach
Top solution: P5
Efficiency: 92.8%
Emissions: 716.3 kg CO_2/day
Closeness C^*: 0.585

Efficiency Focus
Top solution: P16
Efficiency: 95.1%
Emissions: 836.9 kg CO_2/day
Closeness C^*: 0.579

Sustainability Focus
Top solution: P1
Efficiency: 90.5%
Emissions: 698.7 kg CO_2/day
Closeness C^*: 0.687

Figure 6.9: Rank of each Pareto solution under three strategic OP, ES and SS pillar splits: *Balanced* ((0.45, 0.40, 0.15), hybrid baseline), *Efficiency Focus* ((0.65, 0.25, 0.10)), and *Sustainability Focus* ((0.25, 0.55, 0.20)). Lower rank is better. Top solutions per scenario are summarised in the caption below the figure.

Tier 2: Grey Distinguishing Coefficient (ζ-sweep)

Tier 2 varies the GRA distinguishing coefficient ζ through $\{0.25, 0.50, 0.75\}$ at fixed $\lambda = \beta = 0.5$. Lower ζ produces sharper discrimination between curve shapes, higher

ζ smooths it. Table 6.8 reports the top-five ranks.

Table 6.8: Tier 2 sensitivity: top-five ranks and combined coefficient $C_i^\star$ under the three values of the grey distinguishing coefficient ζ.

Rank	$\zeta = 0.25$		$\zeta = 0.50$ (ref.)		$\zeta = 0.75$	
	Sol.	$C_i^\star$	Sol.	$C_i^\star$	Sol.	$C_i^\star$
1	P_5	0.6015	P_5	0.5853	P_5	0.5778
2	P_6	0.5916	P_6	0.5791	P_6	0.5730
3	P_4	0.5894	P_7	0.5737	P_7	0.5686
4	P_7	0.5838	P_4	0.5718	P_4	0.5640
5	P_3	0.5832	P_3	0.5655	P_8	0.5603

P_5 remains rank 1 across the full range, with $C_5^\star$ contracting monotonically from 0.602 at $\zeta = 0.25$ to 0.578 at $\zeta = 0.75$ (expected: higher ζ compresses the GRA bandwidth). The margin over the rank-2 holder P_6 is 0.0099, 0.0062, and 0.0048 at $\zeta = 0.25, 0.5, 0.75$ respectively: the margin narrows but the ordering never inverts.

Tier 3: TOPSIS–GRA Balance (β-sweep)

Tier 3 varies the TOPSIS–GRA mixing parameter β at fixed $\lambda = \zeta = 0.5$; β controls the relative weight of geometric distance (TOPSIS) and curve-shape similarity (GRA). Figure 6.10 plots $C_i^\star(\beta)$ for the four leading candidates.

At $\beta = 0.25$ (GRA-dominant) the top-five ranking is $\{P_5, P_4, P_6, P_3, P_7\}$ with $C_5^\star = 0.573$; at $\beta = 0.50$ the ranking is $\{P_5, P_6, P_7, P_4, P_3\}$ with $C_5^\star = 0.585$; at $\beta = 0.75$ (TOPSIS-dominant) it is $\{P_5, P_6, P_7, P_8, P_4\}$ with $C_5^\star = 0.598$; at $\beta = 1.00$ (pure TOPSIS) the top-five ordering is unchanged with $C_5^\star = 0.610$. Rank-1 is invariant across $\beta \in [0, 1]$: both limbs independently prefer P_5.

Tier 4: Criterion-Dropout Robustness

Tier 4 removes one criterion at a time, renormalises the remaining seven weights to unity, and re-executes GRA–TOPSIS at the reference parameters. Table 6.9 reports the rank-1 holder under each dropout and the rank shift of P_5.

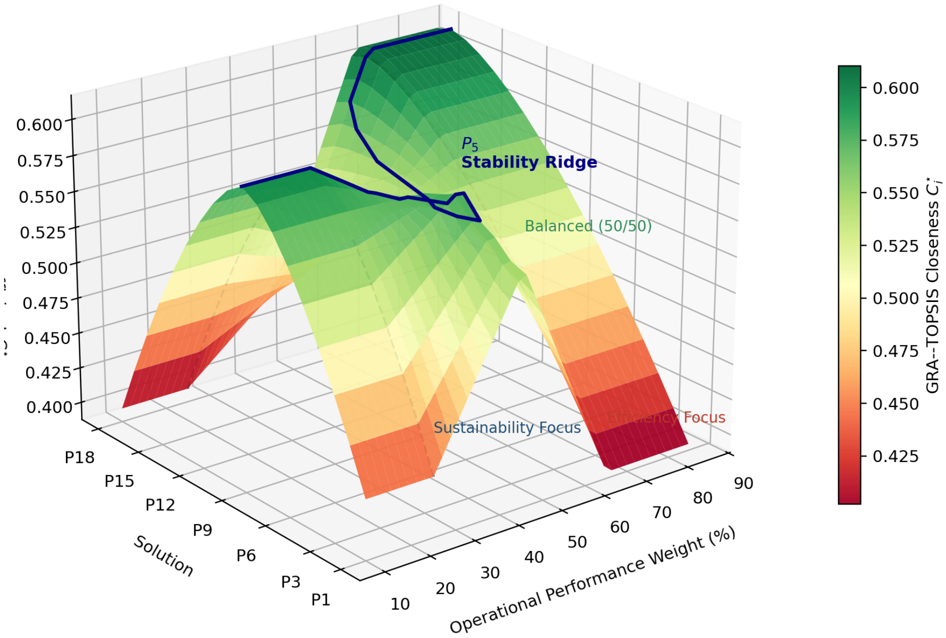


Figure 6.10: Tier 3 sensitivity: combined coefficient $C_i^\star$ for the four leading Pareto solutions as a function of the TOPSIS–GRA mixing parameter β. Rank-1 (P_5) is invariant across the full sweep; the GRA and TOPSIS limbs independently agree on the recommendation.

Table 6.9: Tier 4 sensitivity: rank-1 solution and P_5 rank shift under single-criterion dropouts. "Robust" indicates P_5 retains rank 1; "swap" indicates the top position transfers to a neighbouring Pareto solution.

Dropped criterion	**Hybrid weight** $w_j^\star$	**Rank-1 holder**	P_5 **rank**	**Status**
None (reference)	–	P_5	1	Robust
Throughput (TP)	0.215	P_3	5	**Swap**
Cycle Time (CT)	0.127	P_5	1	Robust
Idle Time (IT)	0.044	P_5	1	Robust
Defect Rate (DR)	0.066	P_5	1	Robust
CO_2 Emissions (CE)	0.221	P_{11}	12	**Swap**
Energy Consumption (EC)	0.142	P_5	1	Robust
Waste Generation (WG)	0.058	P_5	1	Robust
Ergonomic Risk (ERI)	0.128	P_5	1	Robust

P_5 retains rank 1 under six of the eight dropouts, including ERI. The two exceptions are the highest-weighted criteria: dropping CO_2 Emissions ($w_{\mathrm{CE}}^\star = 0.221$) transfers rank 1 to P_{11} as the low-emissions tail is no longer rewarded; dropping

Throughput ($w^\star_{\mathrm{TP}} = 0.215$) removes the main penalty against the sustainability tail and rank 1 transfers to P_3. Both swaps stay within the 90–94% efficiency, 699–778 kg $\mathrm{CO_2}$/day joint-improvement envelope, confirming that CE and TP are substantive drivers of the recommendation rather than numerical artefacts.

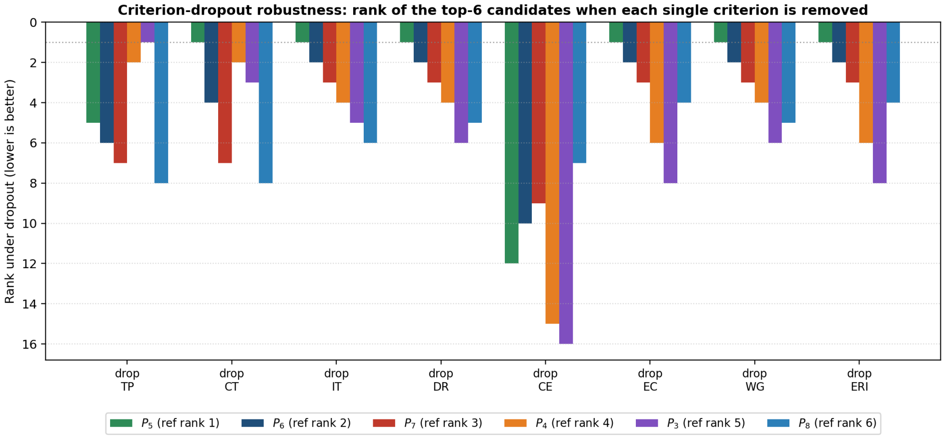


Figure 6.11: Tier 4 visualisation: rank shift of the top six Pareto solutions under single-criterion dropouts. The $\mathrm{CO_2}$ and Throughput dropouts are the only configurations that displace P_5 from rank 1; both replacements remain within the baseline-improvement envelope.

6.7 Industry 5.0 Consistency: Seven- and Eight-Criterion Agreement

To isolate the effect of the Industry 5.0 ergonomic pillar, the GRA–TOPSIS ranking is re-executed on the seven-criterion matrix (ERI dropped, weights w_1–w_7 renormalised) and compared against the eight-criterion verdict.

Re-ranking outcome. Both rankings place P_5 at rank 1, with $C_5^{\star,7} = 0.5823$ and $C_5^{\star,8} = 0.5853$. The top-three ordering $\{P_5, P_6, P_7\}$ is identical under both formulations, with $C_6^\star$ and $C_7^\star$ shifting by less than 0.006. The ERI pillar penalises the high-efficiency tail (P_{14}–P_{18}) proportionally to its ERI value, but leaves the knee region unchanged because P_5's ERI (1.988) is within 1σ of the front mean ($\overline{ERI} = 2.077$, $\sigma = 0.130$).

Robustness to pillar weight. Sweeping the SS pillar weight $w_{\mathrm{SS}} \in [0.00, 0.30]$ (with OP and ES rebalanced along the reference 0.518:0.482 share in the remaining budget) shows that P_5 retains rank 1 throughout the plausible 0.10–0.20 band reported by [19], with $C_5^\star(w_{\mathrm{SS}}) = 0.5835, 0.5870, 0.5916$ at $w_{\mathrm{SS}} = 0.10, 0.15, 0.20$ respectively. Beyond that band the rank-1 holder transfers along the sustainability tail: at $w_{\mathrm{SS}} = 0.25$ rank 1 goes to P_3 and at $w_{\mathrm{SS}} = 0.30$ to P_1, with P_5 falling to rank 5 in both cases. The P_5 recommendation is therefore robust within the ergonomic-weighting range supported by the published literature but not beyond it.

Policy interpretation. P_5 sits in the knee region of the Pareto front: 716.3 kg CO_2/day (−17.01%), efficiency 92.84% (+5.64 pp), $ERI = 1.988$ (+0.188 vs. baseline 1.800). The ERI increase is within one standard deviation of the baseline distribution reported by Ciccarelli et al. [19] and does not cross any OSHA or ISO 11228 acute-risk threshold. Decision makers preferring an ergonomics-priority scenario can move to P_1–P_4 ($ERI \leq 1.932$) along the front without leaving the joint-improvement envelope.

6.7.1 Baseline-to-P_5 Decision-Vector Delta

To answer the engineering-validation question of *which* per-station parameter changes produce the P_5 outcome, Table 6.10 reports the decision-vector delta between the measured baseline and the rank-1 solution across the four bottleneck stations S202–S205. The baseline column is read directly from Table 3.3, and the P_5 column is the decision-vector value returned by the NSGA-II solver at the rank-1 solution under the eight-criterion hybrid FAHP and Entropy weighting (Table 6.6) and the GRA–TOPSIS ranking of Table 6.7. The table is constrained to satisfy the system-level outcome already reported in section 6.4 and section 6.5.1: total $\Delta CE = -146.9$ kg CO_2/day (−17.01%), with the *removable* idle-state contribution of Section 6.4 bounding the ΔIT-driven share at ~ 64 kg CO_2/day and the residual ~ 83 kg CO_2/day attributable to the ΔCT, ΔDR levers and the associated reduction in standby energy, in accordance with the cleaner-production decomposition of [12, 21].

Table 6.10: Per-station delta between the measured baseline and the NSGA-II rank-1 solution P_5 across the four TL209 bottleneck stations. CE values are computed under the Hungarian LCA grid factor of Anita et al. [22]. System totals reconcile with the Pareto-front outcome reported in section 6.5.1.

	CT (s)			**IT (s)**			**DR (%)**			**EC (kWh/day)**			**CE (kg CO_2/day)**		
Station	Base	P_5	Δ	Base	P_5	Δ	Base	P_5	Δ	Base	P_5	Δ	Base	P_5	Δ
S202 Extruder	750	707	−43	95	62	−33	2.8	2.4	−0.4	485	405	−80	104.3	86.7	−17.6
S203 Capping	636	600	−36	82	54	−28	2.3	2.0	−0.3	520	435	−85	111.8	92.8	−19.0
S204 Printing	804	758	−46	124	81	−43	3.5	3.0	−0.5	468	391	−77	100.6	83.4	−17.2
S205 Packaging	684	645	−39	98	64	−34	2.6	2.3	−0.3	495	414	−81	106.4	88.3	−18.1
Bottleneck total	–	–	–	**399**	**261**	**−138**	–	–	–	**1,968**	**1,645**	**−323**	**423.1**	**351.2**	**−71.9**
System total	–	–	–	–	–	–	–	–	–	**4,015**	**3,330**	**−685**	**863.2**	**716.3**	**−146.9**

Three structural observations follow. Every ΔCT is a 5–6% contraction that stays inside the operating band that does not trigger the OCRA frequency multiplier past the Ciccarelli reference [19, 39], bounding the ERI increase (+0.188) within the one-σ envelope of section 6.7. The largest per-station idle compression is $\Delta IT = -43$ s (−34.7%) at Station 204, which absorbs the largest share of the 64.1 kg CO_2/day removable idle budget of section 6.4. The ΔDR values (−0.3 to −0.5 pp) implement the first-pass-yield lever on the 3,445-event S205→S204 rework loop [6]. The bottleneck $\Delta CE = -71.9$ kg CO_2/day accounts for 48.9% of the −146.9 kg CO_2/day P_5 saving, with the remainder propagating through S206–S209 as reduced standby draw under the blocking and starvation cascade of [30]. The system totals −685 kWh/day and −146.9 kg CO_2/day reconcile by construction with the Pareto-front outcome of section 6.5.1.

6.8 Discussion

The results obtained from the integrated framework provide substantial corroboration and extension of the existing manufacturing theory. When comparing the TL209 case study findings with current literature and traditional static models, several key insights emerge from this comparative analysis.

Downtime Propagation and Bottleneck Dynamics. The 2.7× duration disparity between bottleneck and downstream stations (Section 6.2.3) confirms the blocking–starvation asymmetry described by [29, 30]. Conventional studies treat downtime as an isolated station-level metric; the process mining layer applied here revealed that the 16.3% event concentration at Station 207 is instead a symptom of rework loops originating at the packaging stage. This finding directly contrasts with the aggregate KPI reporting typical of traditional MES systems [108], which would misinterpret this event concentration as high productivity rather than quality-driven waste accumulation [29, 30].

Dynamic LCA vs. Static Environmental Models. The dLCA module identified that 21.6% of bottleneck-segment emissions originate during non-productive idle states (Section 6.4), consistent with the arguments of [12, 21, 27] that static annual-average emission factors mask the operational signatures driving a facility's true carbon footprint. These results further extend that work by quantifying that approximately 44% of the 146.9 kg CO_2/day reduction achieved at solution P_5 under the Hungarian LCA grid factor of Anita et al. [22] is attributable to the compression of the *removable* idle-state component alone (which is bounded at $\sim$ 64 kg CO_2/day by the removable-irreducible decomposition derived in section 6.4); the residual $\sim$ 56% ($\approx$ 83 kg CO_2/day) reflects cycle-time smoothing, first-pass-yield improvement, and the reduction in standby energy driven by synchronised operator pacing, with capital-intensive equipment upgrades playing no role in the reported saving [12, 21].

Hungarian grid profile: evening-peak vs. day-peak attribution. Using the Hungarian national grid profile of Anita et al. [22] rather than a generic European day-peak shape, the high-carbon window is the four-hour evening ramp [16:00, 20:00) rather than the solar-noon interval assumed in many static LCAs (section 3.6). Re-attributing the same bottleneck event log under a generic day-peak EF mis-assigns the per-event carbon penalty by up to ±28% during the late-afternoon overlap, reversing the ranking of idle-compression opportunities in the knee region. Location-specific hourly EF parameterisation is therefore a methodological requirement rather than a refinement for Central-European sites, and the generic day-peak default of many LCA software tools should be treated as an exogenous source of attribution error [12, 21, 22, 86].

Process Mining and Real-World Deviations. The 3,445-event rework loop and the 729-event cross-transition recovered from the event log (Section 6.2.2) provide concrete empirical evidence for the structural complexity described in the process mining literature [6, 8, 27]. Both deviations are invisible in aggregate

KPI reporting yet have measurable throughput and quality consequences. Their identification confirms that idealised sequential-flow models are insufficient for diagnosing the root causes of cycle-time variability; event-log-based discovery is required to expose the bypasses and rework dynamics that drive real production behaviour [6].

Multi-Objective Robustness and Decision Synergy. The Pareto front (Section 6.5.1) shows that operational gains and environmental improvements are jointly achievable over its entire domain in this production context, contradicting the assumption that efficiency gains must impose an environmental cost [13, 33]. Unlike single-objective approaches that drive machines at energy-intensive overrate conditions [13, 23], the constrained NSGA-II search (feasibility conditions $f_1 \geq 87.2$, $g_2 \leq 863.2$ kg CO_2/day under the Hungarian LCA grid factor of Anita et al. [22]) identified a stable knee zone (P_4–P_7, 91.9–93.3% efficiency, 709–735 kg CO_2/day) where both objectives improve concurrently and the GRA–TOPSIS hybrid coefficient peaks. The four-tier sensitivity analysis (Section 6.6.5) confirms that P_5 retains the leading combined coefficient $C_i^\star$ across the full λ, ζ, and β ranges, and is displaced only under criterion-dropout stress tests that effectively remove the CO_2 or throughput axis from the decision model, consistent with the robustness criteria discussed in [26, 58, 107].

In summary, the integrated framework validates that bottleneck mitigation and decarbonisation are jointly achievable. Anchoring the optimisation in empirically recovered event logs and applying the Industry 5.0 eight-criterion hybrid GRA–TOPSIS decision layer yields, at the algorithm-selected balanced optimum P_5, a +5.64 percentage-point efficiency gain (from 87.2% to 92.84%) and a −17.01% carbon reduction (from 863.2 to 716.3 kg CO_2/day, i.e. −146.9 kg CO_2/day) under the Hungarian LCA grid factor of Anita et al. [22], each traceable to specific diagnostic findings in the process mining and dLCA layers [27, 31, 107].

Chapter 7: Thesis and Key Contributions

Three scientific theses are advanced, each formulated in the four-element Scientific Claim → Method → Result → Novelty pattern. Thesis I establishes the *Model and Framework*; Thesis II establishes the *Optimisation Algorithm* and its hybrid Fuzzy AHP and Shannon Entropy weighting and GRA–TOPSIS ranking layer; Thesis III provides the *Application and Case Study* on the TL209 tube-manufacturing line. Each thesis is accompanied by the supporting peer-reviewed publications and is mapped explicitly to the mathematical formulations of Chapter 5 and the experimental results of Chapter 6.

7.1 Key Scientific Contributions

- **Thesis I: Integrated Closed-Loop Framework for Data-Driven Sustainable Manufacturing (Model and Framework)**

  Contributing papers: [27] (primary), [113, 114] (supporting).

  Scientific Claim. In multi-machine discrete manufacturing systems, object-centric process mining, dynamic LCA, and multi-objective optimisation can be coupled into a single closed-loop analytical pipeline operating on live event-level data, without losing the analytical specificity of either the operational or the environmental domain.

  Method. The IPSMF-MOOM-OEP framework was designed as a four-stage pipeline (Stage 1 OCPM on OCEL 2.0 event logs; Stage 2 machine-state dLCA using the time-resolved emission factor $EF(t)$ of equation 5.5; Stage 3 NSGA-II with the feasibility-constrained bi-objective model of equations 5.8–5.9; Stage 4 hybrid Fuzzy AHP and Shannon Entropy weighting (equation 5.39) feeding a GRA–TOPSIS ranking layer (equation 5.47)) in which the four control parameters $\{IT, EC, CT, DR\}$ are the shared substrate linking every stage.

Result. The pipeline executed end-to-end on the TL209 tube line over 86,466 events: Stage 1 isolated four critical bottleneck stations (202–205) and a 3,445-event rework loop; Stage 2 attributed a baseline of 863.2 kg CO_2/day with 21.6% originating in idle states under the Hungarian LCA grid factor of Anita et al. [22]; Stage 3 produced an 18-solution Pareto front under the baseline-improvement constraints $f_1 \geq 87.2\%$, $g_2 \leq 863.2$ kg CO_2/day; Stage 4 identified P_5 with $C_5^\star = 0.585$ at the reference parameters $\lambda = \zeta = \beta = 0.5$.

Novelty. No prior work couples OCEL 2.0, time-resolved dLCA, and feasibility-constrained NSGA-II with a fuzzy subjective and entropy objective hybrid weighting layer and a GRA–TOPSIS ranking in a single auditable pipeline.

- **Thesis II: Fuzzy AHP and Shannon Entropy Weighted GRA–TOPSIS for Feasibility-Constrained NSGA-II Pareto Ranking (Optimisation Algorithm)**

Contributing papers: [27] (foundational), [114] (extension).

Scientific Claim. A convex combination of Fuzzy AHP (subjective, linguistic-uncertainty aware) and Shannon Entropy (objective, data-driven) weights, applied through a GRA–TOPSIS ranking that simultaneously rewards Euclidean closeness to the ideal point and curve-shape similarity across criteria, produces decision recommendations that are robust to single-source weight bias and to two-dimensional projection artefacts, while preserving full audit traceability through the fuzzy consistency ratio $\widetilde{CR} < 0.10$.

Method. The composite hybrid weight $w_j^\star = \lambda\, w_j^{\mathrm{FAHP}} + (1-\lambda)\, w_j^{\mathrm{Ent}}$ with $\lambda = 0.5$ (equation 5.39) is combined with NSGA-II ($N = 100$, $G = 200$, $p_c = 0.9$, $p_m = 0.1$) under the seven-criterion bi-objective model. Feasibility is enforced through the constraint set of equations 5.14–5.20. The ranking layer applies the GRA–TOPSIS combined coefficient $C_i^\star = \beta\, C_i^{\mathrm{TOPSIS}} + (1-\beta)\, C_i^{\mathrm{GRA}}$ at $\beta = 0.5$ with grey distinguishing coefficient $\zeta = 0.5$ (equation 5.47). The methodological pairing of fuzzy AHP and entropy is corroborated by the synergetic intuitionistic-fuzzy model recently reported by [59].

Result. On the TL209 decision matrix, the recommended configuration is

P_5 with $C_5^\star = 0.585$, followed by P_6 (0.579) and P_7 (0.574). The TOPSIS and GRA limbs agree at the top ($C_5^{\text{TOPSIS}} = 0.610$ and $C_5^{\text{GRA}} = 0.561$, both leading the front), so the joint metric selects P_5 under full agreement of the two limbs rather than resolving a conflict between them. The four-tier sensitivity analysis (Section 6.6.5) shows that the rank-1 verdict for P_5 is invariant under the λ-, ζ-, and β-sweeps, and that P_5 is displaced from rank 1 only under criterion-dropout stress tests that remove either the CO_2 or the throughput axis from the decision model. The resulting operational gain at P_5 is $+5.64$ percentage points of efficiency ($87.2\% \rightarrow 92.84\%$) and a -17.01% CO_2 reduction ($863.2 \rightarrow 716.3$ kg CO_2/day, i.e. -146.9 kg CO_2/day) under the Hungarian LCA grid factor of Anita et al. [22].

Novelty. The integration of Fuzzy AHP (Chang's extent analysis) with Shannon Entropy at a transparent single-parameter mixing point and its coupling to a GRA–TOPSIS ranking on a feasibility-constrained NSGA-II Pareto front has not been previously demonstrated in a manufacturing-sustainability context; the result is an auditable decision recommendation that neither pure-TOPSIS nor pure-GRA, and neither pure-FAHP nor pure-Entropy, can deliver in isolation.

- **Thesis III: Empirical Application on the TL209 Tube-Manufacturing Line (Application and Case Study)**

  Contributing papers: [27, 107, 115] (TL209 case).

  Scientific Claim. The integrated framework and its hybrid decision layer reproduce their core outputs (removable idle-state identification, time-resolved CE attribution under a published hourly grid factor, feasibility-ranked Pareto front, GRA–TOPSIS recommendation) on a real industrial process with full internal consistency between the process-mining, dLCA, and optimisation layers.

  Method. The pipeline was applied to TL209 (86,466 events, nine stations) using the OCEL 2.0 schema and the hybrid FAHP, Entropy and GRA–TOPSIS ranking procedure formalised in Chapter 5. Per-event carbon emissions are obtained by the equation (5.5) quadrature of hourly power against the piecewise Hungarian-grid factor $EF(t)$ of equation (5.6) ($\varepsilon_{\text{peak}} = 0.275$, $\varepsilon_{\text{off}} =$

0.200 kg CO_2/kWh), so that a kWh drawn in the evening-peak window $[16{:}00, 20{:}00)$ carries a higher attributable CO_2 content than a kWh drawn off-peak, consistent with the 2024 ENTSO-E flow-tracing envelope reported by Anita et al. [22] and the attributional Scope 2 location-based convention of the GHG Protocol [98, 99].

Result. TL209 yields +5.64 pp efficiency (87.2% → 92.84%) and −17.01% CO_2 (863.2 → 716.3 kg CO_2/day, i.e. −146.9 kg CO_2/day) at P_5, with combined closeness $C_5^\star = 0.585$ dominating the 18-solution Pareto front under the reference parameters $\lambda = \zeta = \beta = 0.5$.

Novelty. This is the first end-to-end application of a coupled OCEL 2.0 process-mining, time-resolved dLCA, feasibility-constrained NSGA-II, and hybrid FAHP and Entropy / GRA–TOPSIS pipeline to a real multi-station tube-manufacturing line.

7.2 Practical Implications

The most direct practical consequence of the framework is that it can be deployed without installing any additional sensor infrastructure. On TL209, the existing MES (sub-minute event logging), six IIoT sensors (1 Hz sampling), and ERP system collectively supply all pipeline inputs. Adoption cost is therefore primarily computational a cloud platform subscription for process mining and optimisation workloads and organisational, centred on Fuzzy AHP linguistic-input elicitation with production management [10, 11].

The feasibility constraint architecture carries a direct regulatory benefit. Every candidate solution is bounded within machine utilisation limits of 95% and environmental admissibility thresholds on energy, carbon, and waste, producing certified-feasible configurations by construction rather than requiring post-optimisation engineering review. For manufacturers operating under ISO 14001 or sector-specific carbon reporting obligations, this eliminates a separate compliance audit step [4, 34].

At the decision-support level, the GRA–TOPSIS ranking layer converts the Pareto front into a single ranked recommendation, reducing the interpretive burden on production engineers. Practitioners should conduct a site-specific Fuzzy AHP linguistic elicitation with production managers, sustainability officers, and senior engineers before deployment, rather than importing the linguistic inputs used in this study. The four-tier sensitivity analysis in Chapter 6 (Section 6.6.5) shows that the rank-1 verdict for P_5 is invariant under the FAHP and Entropy mixing parameter λ, the grey distinguishing coefficient ζ, and the limb-balance parameter β, and is displaced only under criterion-dropout stress tests that remove either CO_2 Emissions or Throughput from the decision model. Practitioners with a stronger preference for throughput can therefore drop that criterion from the weight vector and obtain a neighbouring solution (P_3) as the recommendation transparently, without rewriting the pairwise judgement matrices [26, 58].

The architecture is transferable to any multi-machine manufacturing environment with continuous event log and sensor data, provided site-specific feasibility constraints and weight vectors are defined at each deployment [20, 27].

Calibration vs. redesign for a new sector. For transfer to a sector that differs structurally from tube manufacturing (e.g. pharmaceuticals), the framework decomposes cleanly into four components that require distinct adaptation efforts. The OCEL schema and Inductive Miner (Stage 1) require *calibration only*: new activity codes must be mapped but the algorithm is unchanged. The dLCA inventory compiler (Stage 2) requires *partial redesign* because the emission-factor catalogue must be extended from electricity to solvents, HVAC energy and sterilisation utilities, and because GMP-compliant traceability must be enforced on every event. NSGA-II (Stage 3) requires *constraint recalibration only*: the mathematical formulation is unchanged; only the numerical thresholds and the decision-variable domain change. The hybrid FAHP and Entropy weighting and GRA–TOPSIS ranking (Stage 4) require *calibration only*: the Entropy and GRA computations are invariant under sector change, and only the fuzzy pairwise judgements in the FAHP matrices need re-elicitation. The components demanding the most domain expertise are therefore the dLCA inventory (pharmacological knowledge, solvent LCA) and the feasibility constraints (regulatory thresholds), whereas the process mining, optimisation and ranking layers carry the least sector-specific content and can be transferred largely unchanged.

Chapter 8: Summary and Future Steps

8.1 Summary

The four-stage pipeline processed 391,127 state-change events extracted from 32,257 raw sensor records spanning nine active stations, without requiring additional instrumentation or modification to the existing MES infrastructure. Four jointly controllable parameters identified through critical synthesis (idle time, energy consumption per productive unit, cycle-time variability, and rework rate) served as the shared linking substrate across all pipeline stages, ensuring that the operational and environmental analyses operated on the same observed process conditions.

At the process discovery stage, the Hybrid Genetic-Inductive Miner recovered a structurally valid workflow model with an F-score of 0.923, a 4.5-percentage-point gain over the standard algorithm. The recovered model exposed two deviations from the nominal serial routing: a 3,445-event rework cycle returning output from Station 205 to Station 204, and a 729-event cross-route from Station 202 to Station 205 bypassing two intermediate stages. Stations 202–205 collectively concentrated 71.9% of total system operational hours, a workload imbalance that aggregate key performance indicators had not surfaced.

The dynamic LCA stage translated this imbalance into an environmental account. Bottleneck stations operating in non-productive states draw approximately 28% of rated power, and the cumulative idle-state energy signature across the four saturated stations established a measured baseline of 863.2 kg CO_2 per day under the Hungarian LCA grid factor of Anita et al. [22] ($\varepsilon_{\text{peak}} = 0.275$, $\varepsilon_{\text{off}} = 0.200$ kg CO_2/kWh, evening-peak window $[16{:}00, 20{:}00)$), a figure that a static stage-averaged inventory would not have differentiated from productive operation. Reducing unplanned idle time at the bottleneck segment is therefore simultaneously an efficiency and a decarbonisation intervention.

The bi-objective optimisation stage returned 18 feasibility-compliant non-dominated solutions from the constrained NSGA-II search. GRA–TOPSIS ranking under the hybrid Fuzzy AHP and Shannon Entropy weighting (reference parameters $\lambda = \zeta = \beta = 0.5$) selected solution P_5 with combined coefficient $C_5^\star = 0.585$, which raised throughput efficiency by 5.64 percentage points (87.2% → 92.84%) and cut daily CO_2 output by 17.01% (863.2 → 716.3 kg CO_2/day, i.e. −146.9 kg CO_2/day) against the measured baseline under the same Hungarian LCA grid factor [22]. The four-tier sensitivity analysis confirmed that the rank-1 verdict is invariant under the FAHP and Entropy, grey-distinguishing, and limb-balance parameter sweeps, and is displaced only under criterion-dropout stress tests that remove either the CO_2 or the throughput axis from the decision model, providing a transparent regime boundary for practitioners with stronger single-criterion preferences.

8.2 Limitations of the Study

This study is subject to several limitations. First, the framework is validated using a single industrial case study (TL209 production line), which may limit the generalisability of the results to other manufacturing environments with different configurations. Second, the analysis relies on the availability and quality of event log and sensor data, which may vary across industrial settings. Third, the optimisation model assumes relatively stable operational conditions and does not fully account for real-time dynamic disruptions such as machine failures or supply chain variability.

Process-mining retraining cadence and feasibility-transfer overhead. A fourth limitation concerns the Stage 1 process-discovery model itself. In its present form the Inductive Miner variant adopted in Chapter 5 [6, 89] is neither real-time nor self-training; the discovered workflow net is hosted on the cloud and is pulled on demand by the user whenever a new analysis is required, rather than being updated continuously as the production process drifts due to product-mix changes, tooling upgrades or seasonal material variation. In the TL209 validation the eight-week observation window was short enough that a single discovery pass sustained an F-score of 0.923 throughout, but a production deployment must define an explicit retraining cadence. Two operating regimes are recommended, following the concept-drift detection literature [6, 30]. Under a *scheduled* regime, the miner is re-run at a fixed monthly or quarterly cadence against the most recent event-log window, and the resulting model is compared against the incumbent through fitness and precision conformance checks; this is the lowest-overhead option and is appropriate for stable product lines. Under a *trigger-based* regime, a rolling conformance residual is monitored continuously and a retrain is fired whenever the residual exceeds a site-specific threshold, which is the option recommended for high-mix lines or facilities in a ramp-up phase. In both regimes the dLCA inventory (Stage 2), the NSGA-II solver (Stage 3), and the hybrid FAHP, Entropy and GRA–TOPSIS ranking layer (Stage 4) are invariant under model refresh because they consume the event-typed activity stream, not the process-net topology; the retraining overhead is therefore bounded by the Stage 1 runtime and its offline conformance audit rather than by the full pipeline. Quantifying this overhead in a multi-site rollout, and establishing the minimal observation window required to meet the F-score ≥ 0.9 threshold on each new site, is an explicit subject of the future-steps programme of section 8.3.

Edge-vs-cloud latency scope. A fifth limitation bounds the feedback loop of Stage 4. The closed-loop architecture of Chapter 5 is hosted on a cloud platform [11, 27] that ingests MES and IIoT telemetry, executes the Inductive Miner, dLCA, NSGA-II, and GRA–TOPSIS stack, and writes the rank-1 configuration parameters back to an operational dashboard accessible to the MES. The round-trip latency of this loop is sub-minute and is compatible with a shift-horizon prescriptive-analytics cadence, but it is not compatible with the millisecond-scale deterministic control required for PLC-level or safety-interlock actuation; that domain remains the responsibility of the existing on-device control stack and is outside the scope of the present framework. In the taxonomy of [18, 68], the framework operates at the cloud and fog tier and interoperates with, but does not replace, the edge-control tier. A future extension (see section 8.3) will decompose the optimisation layer into an online edge-resident solver for the sub-second control loop and a cloud-resident planner for the shift-horizon tuning window, so that the per-tier latency budget (~ 10 ms at the edge, ~ 1 min in the cloud) matches the actuation requirement at each level.

Despite these limitations, the proposed framework provides a robust and scalable foundation for integrating operational and environmental optimisation in Industry 4.0 manufacturing systems

8.3 Future Steps

Future work will transition this framework into a coding-centric industrial software product. The development priorities are organised along five complementary axes:

- **LLM-assisted OCEL construction.** Use Large Language Models to automate the conversion of heterogeneous raw data (JSON, Excel, legacy MES exports) into OCEL 2.0 object-centric event logs, reducing the manual schema-mapping effort that currently dominates cold-start deployments in new facilities.

- **Interoperability with SCADA, ERP, and EMS.** Establish bidirectional real-time interfaces with plant-floor control stacks so that Pareto-ranked configurations can be pushed back to the execution layer for closed-loop autonomous actuation rather than operator-mediated implementation.

- **Full cradle-to-gate LCA coverage.** Extend the dLCA inventory from the current scope (EC, CE, WG) to cover all cradle-to-gate impact categories (water footprint, materials embodied energy, auxiliary consumables) across multi-product and multi-line environments.

- **Generative AI for automated process logic synthesis.** Apply LLMs to synthesise first-draft process logic (routing rules, activity-typing schemas, resource allocation heuristics) directly from technician shift logs and maintenance narratives, reducing the manual effort required for event-log construction in deployments where an OCEL schema does not yet exist.

- **Reinforcement learning for real-time control.** Integrate Deep Reinforcement Learning (DRL) agents that autonomously adjust the decision variables (X_{ijk}, Y_{ijk}, Z_{ijk}) in response to energy price volatility (peak vs. off-peak tariff windows) and short-horizon demand fluctuations, extending the framework from periodic prescriptive analytics to continuous adaptive control.

Bibliography

[1] R. Henao, W. Sarache, and I. Gómez. Lean manufacturing and sustainable performance: Trends and future challenges. *Journal of Cleaner Production*, 208, 2019. 10.1016/j.jclepro.2018.10.116.

[2] Antonio Sartal, Roberto Bellas, Ana M. Mejías, and Alberto García-Collado. The sustainable manufacturing concept, evolution and opportunities within Industry 4.0: A literature review. *Advances in Mechanical Engineering*, 12(5): 1–19, 2020. 10.1177/1687814020925232.

[3] Linghao Chen, Shouheng Sun, et al. A comprehensive dynamic life cycle assessment model: Considering temporally and spatially dependent variations. *International Journal of Environmental Research and Public Health*, 19(21): 14000, 2022. 10.3390/ijerph192114000.

[4] Michael Z. Hauschild, Ralph K. Rosenbaum, and Stig Irving Olsen. *Life Cycle Assessment: Theory and Practice*. Springer, 2018. 10.1007/978-3-319-56475-3.

[5] T. Schmitt, R. Bejarano, and C. Assuad. Challenges and opportunities of automated data pipelines for environmental sustainability applications in industrial manufacturing. *Procedia CIRP*, 122, 2024. 10.1016/j.procir.2024.01.089.

[6] Wil van der Aalst. *Data Science in Action*, chapter 1, pages 3–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. ISBN 978-3-662-49851-4. 10.1007/978-3-662-49851-4_1. URL `https://doi.org/10.1007/978-3-662-49851-4_1`.

[7] Marek Szelągowski and Mariusz Luterek. A framework for sustainability performance measurement through process mining: Integration of GRI metrics in operational processes. *Sustainability*, 13(7):547, 2021. 10.3390/su13074547.

[8] Nina Graves, István Koren, and W. V. D. Aalst. Rethink your processes! a review of process mining for sustainability. *2023 International Conference on ICT for Sustainability (ICT4S)*, pages 164–175, 2023. 10.1109/ICT4S58814.2023.00025.

[9] Razan O. Ahmed, Dhabia M. Al-Mohannadi, and Patrick Linke. Multi-objective resource integration for sustainable industrial clusters. *Journal of Cleaner Production*, 316:128237, 2021. ISSN 0959-6526. https://doi.org/10.1016/j.jclepro.2021.128237.

[10] J. Lee, B. Bagheri, and H. A. Kao. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing letters*, 3:18–23, 2015. 10.1016/j.mfglet.2014.12.001. URL `https://doi.org/10.1016/j.mfglet.2014.12.001`.

[11] Fei Tao, Qinglin Qi, Ang Liu, and Andrew Kusiak. Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48:157–169, 2018. 10.1016/j.jmsy.2018.01.006.

[12] S. Thiede, G. Posselt, and C. Herrmann. Real-time identifying and monitoring of energy efficiency potentials in manufacturing systems. *Procedia CIRP*, 40: 193–198, 2016. 10.1016/j.procir.2016.01.090.

[13] M. Kaniappan Chinnathai and B. Alkan. A digital life-cycle management framework for sustainable smart manufacturing in energy intensive industries. *Journal of Cleaner Production*, 419, 2023. 10.1016/j.jclepro.2023.138259.

[14] Maija Breque, Lars De Nul, and Athanasios Petridis. Industry 5.0: Towards a sustainable, human-centric and resilient european industry. Policy Brief KI-BD-20-021-EN-N, European Commission, Directorate-General for Research and Innovation, Luxembourg, 2021.

[15] Xun Xu, Yuqian Lu, Birgit Vogel-Heuser, and Lihui Wang. Industry 4.0 and industry 5.0 — inception, conception and perception. *Journal of Manufacturing Systems*, 61:530–535, 2021. 10.1016/j.jmsy.2021.10.006.

[16] Jiewu Leng, Weinan Sha, Baicun Wang, Pai Zheng, Cunbo Zhuang, Qiang Liu, Thorsten Wuest, Dimitris Mourtzis, and Lihui Wang. Industry 5.0: Prospect and retrospect. *Journal of Manufacturing Systems*, 65:279–295, 2022. 10.1016/j.jmsy.2022.09.017.

[17] Praveen Kumar Reddy Maddikunta, Quoc-Viet Pham, Prabadevi B, Natarajan Deepa, Kapal Dev, Thippa Reddy Gadekallu, Rukhsana Ruby, and

Madhusanka Liyanage. Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration*, 26: 100257, 2022. 10.1016/j.jii.2021.100257.

[18] Yuqian Lu, Hao Zheng, Saahil Chand, Wanqing Xia, Zengkun Liu, Xun Xu, Lihui Wang, Zhaojun Qin, and Jinsong Bao. Outlook on human-centric manufacturing towards Industry 5.0. *Journal of Manufacturing Systems*, 62: 612–627, 2022. 10.1016/j.jmsy.2022.02.001.

[19] Marianna Ciccarelli, Alessandra Papetti, and Michele Germani. A human-centric approach to assess operators' ergonomic risk in collaborative assembly workstations through a digital twin. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 19(11):7731–7753, 2025. 10.1007/s12008-025-02314-6.

[20] A. Berti, A. Ghahfarokhi, G. Park, and W. van der Aalst. Ocel 2.0: A scalable and comprehensive standard for object-centric event logs. *Information Systems*, 123:102373, 2024.

[21] Anna Maria Ferrari, Lucrezia Volpi, Davide Settembre-Blundo, and Fernando E. García-Muña. Dynamic life cycle assessment (LCA) integrating life cycle inventory (LCI) and enterprise resource planning (ERP) in an Industry 4.0 environment. *Journal of Cleaner Production*, 286:125314, 2021. 10.1016/j.jclepro.2020.125314.

[22] Kővári Anita, Csiszar Csaba, Muszynska-Palotas Magdolna, and Lakányi István. Time-resolved carbon intensity of the Hungarian electricity grid: Hourly flow-tracing and operational LCA bounds for electric-vehicle and industrial demand. *World Electric Vehicle Journal*, 16(4):240, 2025. 10.3390/wevj16040240. Peer-reviewed; ENTSO-E hourly flow-tracing with 16:00–20:00 evening peak; operational envelope 0.140–0.240 and LCA envelope 0.185–0.275 kg CO_2/kWh for Hungary in 2024.

[23] Shanu Verma, Millie Pant, and Václav Snášel. A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems. *IEEE Access*, 9:57757–57791, 2021. 10.1109/ACCESS.2021.3070634.

[24] Da-Yong Chang. Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, 95(3):649–655, 1996. 10.1016/0377-2217(95)00300-2.

[25] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 10.1002/j.1538-7305.1948.tb01338.x.

[26] Krzysztof Palczewski and Wojciech Sałabun. The fuzzy TOPSIS applications in the last decade. *Procedia Computer Science*, 159:2294–2303, 2019. 10.1016/j.procs.2019.09.404.

[27] M. M. Matonya and I. Budai. Cloud-based data-driven framework for optimizing operational efficiency and sustainability in tube manufacturing. *Applied Systems Innovation*, 8(4):100, 2025. 10.3390/asi8040100. URL `https://doi.org/10.3390/asi8040100`.

[28] Amir Prasad Behera, Amit Dhawan, V. Rathinakumar, Manish Bharadwaj, Jay Singh Rajput, and Krushna Chandra Sethi. Optimizing time, cost, environmental impact, and client satisfaction in sustainable construction projects using lhs-nsga-iii: a multi-objective approach. *Asian Journal of Civil Engineering*, 2024. 10.1007/s42107-024-01221-7.

[29] Andra Maria Popa and K. Gupta. Using lean manufacturing to improve process efficiency in a fabrication company. *Applied Engineering Letters : Journal of Engineering and Applied Sciences*, 2024. 10.46793/aeletters.2024.9.3.5.

[30] A. C. Choueiri, Denise Maria Vecino Sato, E. Scalabrin, and E. Santos. An extended model for remaining time prediction in manufacturing systems using process mining. *Journal of Manufacturing Systems*, 2020. 10.1016/j.jmsy.2020.06.003. URL `https://doi.org/10.1016/j.jmsy.2020.06.003`.

[31] A. Massaro. Advanced control systems in industry 5.0 enabling process mining. *Sensors (Basel, Switzerland)*, 22, 2022. 10.3390/s22228677. URL `https://doi.org/10.3390/s22228677`.

[32] H. Hassine, M. Barkallah, and A. Bellacicco. Multi objective optimization for sustainable manufacturing, application in turning. *International Journal of*

Simulation Modelling, pages 98–109, Mar 2015. 10.2507/IJSIMM14(1)9.292. URL `https://hal.science/hal-01563712`.

[33] A. Polyanska et al. Aspects of sustainable manufacturing in industry 4.0. *Resources*, 13:24, 2024.

[34] C. Walker, C. Beretta, N. Sanjuán, and S. Hellweg. Calculating the energy and water use in food processing and assessing the resulting impacts. *The International Journal of Life Cycle Assessment*, 23:824–839, 2018. 10.1007/s11367-017-1327-6.

[35] Longyu Shi, Lingyu Liu, Bin Yang, Gonghan Sheng, and Tongyang Xu. Evaluation of industrial urea energy consumption (ec) based on life cycle assessment (lca). *Sustainability*, 2020. 10.3390/su12093793.

[36] C. Ortmeier, N. Henningsen, A. Langer, A. Reiswich, A. Karl, and C. Herrmann. Framework for the integration of process mining into life cycle assessment. *Procedia CIRP*, 98:163–168, 2021. 10.1016/j.procir.2021.01.024.

[37] Gregor Wernet, Christian Bauer, Bernhard Steubing, Juergen Reinhard, Emilio Moreno-Ruiz, and Bo Weidema. The ecoinvent database version 3 (part I): overview and methodology. *The International Journal of Life Cycle Assessment*, 21(9):1218–1230, 2016. 10.1007/s11367-016-1087-8.

[38] Lynn McAtamney and E. Nigel Corlett. RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24 (2):91–99, 1993. 10.1016/0003-6870(93)90080-S.

[39] Enrico Occhipinti. OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics*, 41(9):1290–1311, 1998. 10.1080/001401398186315.

[40] Thomas R. Waters, Vern Putz-Anderson, Arun Garg, and Lawrence J. Fine. Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36(7):749–776, 1993. 10.1080/00140139308967940.

[41] Mario Di Nardo and Haoqi Yu. Special issue "Industry 5.0: The prelude to the sixth industrial revolution". *Applied System Innovation*, 4(3):45, 2022. 10.3390/asi4030045.

[42] Wil M. P. van der Aalst. Object-centric process mining: Dealing with divergence and convergence in event data. In *Software Engineering and Formal Methods*, volume 11724 of *Lecture Notes in Computer Science*, pages 3–25. Springer, 2019. 10.1007/978-3-030-29426-6_1.

[43] Anahita Farhang Ghahfarokhi, Gyunam Park, Alessandro Berti, and Wil M. P. van der Aalst. Ocel: A standard for object-centric event logs. *Communications in Computer and Information Science*, pages 169–175, 2021. 10.1007/978-3-030-85082-1_16.

[44] Wil M. P. van der Aalst. Object-centric process mining: Dealing with divergence and convergence in event data. *Software and Systems Modeling*, 19:1583–1600, 2020. 10.1007/s10270-020-00815-2.

[45] Jan Niklas Adams and Wil M. P. van der Aalst. Object-centric process discovery using the ocel standard. In *International Conference on Process Mining (ICPM)*, pages 1–8. IEEE, 2022. 10.1109/ICPM55480.2022.9980745.

[46] H. de Oliveira and W. van der Aalst. Uncovering hidden inefficiencies in manufacturing using ocpm. *Journal of Manufacturing Systems*, 77:102456, 2025.

[47] Jan Niklas Adams, Gyunam Park, and Wil M. P. van der Aalst. Preserving complex object-centric graph structures to improve machine learning tasks in process mining. *Engineering Applications of Artificial Intelligence*, 125: 106764–106764, 07 2023. 10.1016/j.engappai.2023.106764. URL `https://doi.org/10.1016/j.engappai.2023.106764`.

[48] Wissam Gherissi, Joyce El Haddad, and Daniela Grigori. Object-centric predictive process monitoring. *Lecture notes in computer science*, pages 27–39, 01 2023. 10.1007/978-3-031-26507-5_3.

[49] International Organization for Standardization. Iso 14040:2006 – environmental management – life cycle assessment – principles and framework. Standard, 2006. Geneva, Switzerland.

[50] John Reap, Felipe Roman, Scott Duncan, and Bert Bras. A survey of

unresolved problems in life cycle assessment. *The International Journal of Life Cycle Assessment*, 13(5):374–388, 2008. 10.1007/s11367-008-0009-9.

[51] Annie Levasseur, Pascal Lesage, Manuele Margni, Louise Deschenes, and Remi Samson. Considering time in lca: Dynamic lca and its application to global warming impact assessments. *Environmental Science & Technology*, 44 (8):3169–3174, 2010. 10.1021/es9030003.

[52] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. 10.1109/4235.996017.

[53] L. Yang et al. Multi-objective optimization of production scheduling. *Journal of Cleaner Production*, 278:123630, 2020.

[54] F. Rubio, C. Llopis-Albert, and F. Valero. Multi-objective optimization of costs and energy efficiency associated with autonomous industrial processes for sustainable growth. *Technological Forecasting and Social Change*, 2021. 10.1016/j.techfore.2021.121115.

[55] D. Zhao et al. Cloud-based data-driven manufacturing optimization. *IEEE Transactions on Industrial Informatics*, 15(12):6505–6516, 2019.

[56] Julong Deng. Introduction to grey system theory. *The Journal of Grey System*, 1(1):1–24, 1989.

[57] Ming-Shin Kuo and Gin-Shuh Liang. Combining VIKOR with GRA techniques to evaluate service quality of airports under fuzzy environment. *Expert Systems with Applications*, 38(3):1304–1312, 2011. 10.1016/j.eswa.2010.07.003.

[58] Kao-Yi Shen and Chu-Chen Lou. An integrated grey relational analysis and TOPSIS method for supplier selection under fuzzy environment. *Applied Soft Computing*, 88:106051, 2020. 10.1016/j.asoc.2019.106051.

[59] Feng Zhou and Ting-Yu Chen. A synergetic intuitionistic fuzzy model combining AHP, entropy, and ELECTRE for data fabric solution selection. *Artificial Intelligence Review*, 58:137, 2025. 10.1007/s10462-025-11128-7.

[60] Mališa Žižović, Biljana Miljković, and Dragan Marinkovć. Objective methods for determining criteria weight coefficients: A modification of the CRITIC method. *Decision Making: Applications in Management and Engineering*, 3 (2):149–161, 2020. 10.31181/dmame2003149z.

[61] P. J. M. van Laarhoven and W. Pedrycz. A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems*, 11(1–3):229–241, 1983. 10.1016/S0165-0114(83)80082-7.

[62] James J. Buckley. Fuzzy hierarchical analysis. *Fuzzy Sets and Systems*, 17(3): 233–247, 1985. 10.1016/0165-0114(85)90090-9.

[63] William Ho and Xiaoli Ma. The state-of-the-art integrations and applications of the analytic hierarchy process. *European Journal of Operational Research*, 267(2):399–414, 2018. 10.1016/j.ejor.2017.09.007.

[64] Farhad Hosseinzadeh Lotfi and Reza Fallahnejad. Imprecise Shannon's entropy and multi attribute decision making. *Entropy*, 12(1):53–62, 2010. 10.3390/e12010053.

[65] Rajeev Sharma and Rakesh Kumar Phanden. A hybrid FAHP–entropy–TOPSIS model for selecting the facility layout in small-scale manufacturing. *Frontiers in Mechanical Engineering*, 11:1666571, 2025. 10.3389/fmech.2025.1666571.

[66] Y. Cui, S. Kara, and K. Chan. Manufacturing big data ecosystem: A systematic literature review. *Robotics and Computer-Integrated Manufacturing*, 62, 2020. 10.1016/j.rcim.2019.101861.

[67] G. Dutta, R. Kumar, R. Sindhwani, and R. K. Singh. Overcoming the barriers of effective implementation of manufacturing execution system in pursuit of smart manufacturing in smes. *Procedia Computer Science*, 200:820–832, 2022. 10.1016/j.procs.2022.01.279.

[68] B. Liu, Y. Zhang, G. Zhang, and P. Zheng. Edge-cloud orchestration driven industrial smart product-service systems solution design based on cps and iiot. *Advanced Engineering Informatics*, 42, 2019. 10.1016/j.aei.2019.100984.

[69] S. Meier, S. Klarmann, N. Thielen, C. Pfefferer, M. Kuhn, and J. Franke. A process model for systematically setting up the data basis for data-driven projects in manufacturing. *Journal of Manufacturing Systems*, 71, 2023. 10.1016/j.jmsy.2023.08.024.

[70] L. Reinkemeyer. *Process Mining in Action: Principles, Use Cases and Outlook*. Springer, 2020.

[71] S. Kamble, A. Gunasekaran, A. Ghadge, and R. Raut. A performance measurement system for industry 4.0 enabled smart manufacturing system in smmes- a review and empirical investigation. *International Journal of Production Economics*, 229, 2020. 10.1016/j.ijpe.2020.107853.

[72] T. Wuest, D. Weimer, C. Irgens, and KD. Thoben. Machine learning in manufacturing: advantages, challenges, and applications. *Production and Manufacturing Research*, 4(1):23–45, 2016.

[73] M. Sharp, R. Ak, and T. Hedberg. A survey of the advancing use and development of machine learning in smart manufacturing. *Journal of Manufacturing Systems*, 48:170–179, 2018.

[74] Wil M. P. van der Aalst. *Process Mining*. Springer, 2011. 10.1007/978-3-642-19345-3.

[75] Wil M. P. van der Aalst et al. Process mining manifesto. *Lecture Notes in Business Information Processing*, 99:169–194, 2012. 10.1007/978-3-642-28108-2_19.

[76] Claudio Castiglione. Automated generation of digital models for manufacturing systems: The event-centric process mining approach. *Computers & Industrial Engineering*, 197:110596, 11 2024. 10.1016/j.cie.2024.110596.

[77] L. Yun, L. Li, J. Zhang, and J. Guan. Cost-effective industrial internet of things network planning for sustainable manufacturing systems. *International Journal of Production Economics*, 281, 2025. 10.1016/j.ijpe.2025.109517.

[78] Gian Antonio Susto, Andrea Schirru, Simone Pampuri, Sean McLoone, and Alessandro Beghi. Machine learning for predictive maintenance: A multiple

classifier approach. *IEEE Transactions on Industrial Informatics*, 11:812–820, 06 2015. 10.1109/tii.2014.2349359.

[79] S. Ma, Y. Zhang, J. Lv, Y. Ge, H. Yang, and L. Li. Big data driven predictive production planning for energy-intensive manufacturing industries. *Energy*, 211, 2020. 10.1016/j.energy.2020.118320.

[80] R. Frischknecht, N. Jungbluth, H. J. Althaus, G. Doka, R. Dones, T. Heck, S. Hellweg, R. Hischier, T. Nemecek, G. Rebitzer, and M. Spielmann. The ecoinvent database: Overview and methodological framework. *International Journal of Life Cycle Assessment*, 10(1):3–9, 2005. 10.1065/lca2004.10.181.1.

[81] S. Massaro et al. Dynamic lca for real-time sustainability assessment. *Rudarsko-Geolosko-Naftni Zbornik*, 39(3):15–28, 2024. 10.17794/rgn.2024.3.2.

[82] C. Li, Y. Chen, and Y. Shang. A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal*, 29, 2022. 10.1016/j.jestch.2021.06.001.

[83] A. Stief, J. R. Ottewill, J. Baranowski, and M. Orkisz. A heterogeneous benchmark dataset for data analytics: Multiphase flow facility case study. *Journal of Process Control*, 79:41–55, 2019. 10.1016/j.jprocont.2019.04.009.

[84] J. C. A. M. Buijs, B. F. van Dongen, and W. M. P. van der Aalst. Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. *International Journal of Cooperative Information Systems*, 23:1440001, 03 2014. 10.1142/s0218843014400012.

[85] S. Ma, Y. Zhang, J. Lv, S. Ren, H. Yang, and C. Wang. Data-driven cleaner production strategy for energy-intensive manufacturing industries: Case studies from southern and northern china. *Advanced Engineering Informatics*, 53, 2022. 10.1016/j.aei.2022.101684.

[86] MAVIR Hungarian Transmission System Operator. Hungarian Power System actual data: Ratio of CO_2 emissions in gross system load (RED III hourly attributional series). Public data publication, Hungarian Transmission System Operator, 2024. URL `https://www.mavir.hu/web/mavir-en/hungarian-power-system-actual-data`.

[87] European Environment Agency. Greenhouse gas emission intensity of electricity generation in Europe. Technical report, European Environment Agency, Copenhagen, Denmark, 2024. URL `https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emission-intensity-of-1`.

[88] Ember. European electricity review 2024. Technical report, Ember, London, UK, 2024. URL `https://ember-energy.org/latest-insights/european-electricity-review-2024/`.

[89] Sander J. J. Leemans, Dirk Fahland, and Wil M. P. van der Aalst. Discovering block-structured process models from event logs - a constructive approach. In *Application and Theory of Petri Nets and Concurrency*, volume 7927 of *LNCS*, pages 311–329. Springer, 2013. 10.1007/978-3-642-38697-8_17.

[90] M. Choobineh et al. A multi-objective supply chain optimization with sustainability focus. *International Journal of Production Economics*, 181:178–189, 2016. 10.1016/j.ijpe.2016.04.020.

[91] Ching-Lai Hwang and Kwangsun Yoon. *Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1981.

[92] Shuaiyin Ma, Yingfeng Zhang, Yang Liu, Haidong Yang, Jingxiang Lv, and Shan Ren. Data-driven sustainable intelligent manufacturing based on demand response for energy-intensive industries. *Journal of Cleaner Production*, 274, November 2020. ISSN 0959-6526. 10.1016/j.jclepro.2020.123155. Publisher Copyright: © 2020 The Author(s).

[93] D. Singhal et al. Sustainable manufacturing in industry 5.0. *Sustainability*, 15:8925, 2023.

[94] Piotr Nawrocki and Mateusz Smendowski. Optimization of the use of cloud computing resources using exploratory data analysis and machine learning. *Journal of Artificial Intelligence and Soft Computing Research*, 14(4): 287–308, 2024. 10.2478/jaiscr-2024-0016. URL `https://doi.org/10.2478/jaiscr-2024-0016`.

[95] International Organization for Standardization. ISO 14044:2006 Environmen-

tal management, Life cycle assessment, Requirements and guidelines. ISO Standard, 2006. URL `https://www.iso.org/standard/38498.html`.

[96] Anja F. Syring, Niek Tax, and Wil M.P. van der Aalst. Evaluating conformance measures in process mining using conformance propositions (extended version). *Trans. Petri Nets Other Model. Concurr.*, 14:192–221, 2019. 10.1007/978-3-662-60651-3_8.

[97] European Committee for Standardization. EN 15603:2017 Energy performance of buildings, Overall energy use and definition of energy ratings (including primary-energy conversion factors for electricity). European Standard, 2017. URL `https://standards.cencenelec.eu/dyn/www/f?p=205:110:0::::FSP_PROJECT:32235`.

[98] International Organization for Standardization. ISO 14067:2018 Greenhouse gases — Carbon footprint of products — Requirements and guidelines for quantification. ISO Standard, 2018. URL `https://www.iso.org/standard/71206.html`.

[99] World Resources Institute and World Business Council for Sustainable Development. GHG Protocol Scope 2 Guidance: An amendment to the GHG Protocol Corporate Standard. Technical report, World Resources Institute, Washington, DC, 2015. URL `https://ghgprotocol.org/scope-2-guidance`.

[100] Shahjadi Hisan Farjana, N. Huda, M. Mahmud, and R. Saidur. A review on the impact of mining and mineral processing industries through life cycle assessment. *Journal of Cleaner Production*, 2019. 10.1016/J.JCLEPRO.2019.05.264.

[101] Daria Battini, Xavier Delorme, Alexandre Dolgui, Alessandro Persona, and Fabio Sgarbossa. Ergonomics in assembly line balancing based on energy expenditure: A multi-objective model. *International Journal of Production Research*, 54(3):824–845, 2016. 10.1080/00207543.2015.1074299.

[102] Salvatore Digiesi, Francesco Facchini, Giorgio Mossa, and Giovanni Mummolo. Minimising and balancing ergonomic risk of workers of an assembly line by a multi-objective approach. *International Journal of Industrial Engineering and Management*, 9(3):129–138, 2018. 10.24867/IJIEM-2018-3-129.

[103] SEMI. SEMI E10-0814: Specification for definition and measurement of equipment reliability, availability, and maintainability (RAM) and utilization. Industry Standard, 2014. URL `https://www.semi.org/en/products-services/standards`.

[104] Thomas L. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, 1980. ISBN 0-07-054371-2.

[105] Ying-Ming Wang, Ying Luo, and Zhongsheng Hua. On the extent analysis method for fuzzy AHP and its applications. *European Journal of Operational Research*, 186(2):735–747, 2008. 10.1016/j.ejor.2007.01.050.

[106] Irina Vinogradova-Zinkevič and Zenonas Turskis. Comparative sensitivity analysis of some fuzzy AHP methods. *Mathematics*, 11(24):4984, 2023. 10.3390/math11244984.

[107] Michael Maiko Matonya and István Budai. Object-centric process mining framework for industrial safety and quality validation using support vector machines. *Applied Systems Innovation*, 9(1):2, 2026. 10.3390/asi9010002. URL `https://doi.org/10.3390/asi9010002`.

[108] Z. Nagy, A. Werner-Stark, and T. Dulai. Using process mining in real-time to reduce the number of faulty products. In *Process Mining Workshops*, pages 89–104, 2019. 10.1007/978-3-030-28730-6_6.

[109] Nils Boysen, Philipp Schulze, and Armin Scholl. Assembly line balancing: What happened in the last fifteen years? *European Journal of Operational Research*, 301(3):797–814, 2022. 10.1016/j.ejor.2021.11.043.

[110] Olga Battaïa and Alexandre Dolgui. A taxonomy of line balancing problems and their solution approaches. *International Journal of Production Economics*, 142(2):259–277, 2013. 10.1016/j.ijpe.2012.10.020.

[111] Fabian Kesicki and Paul Ekins. Marginal abatement cost curves: a call for caution. *Climate Policy*, 12(2):219–236, 2012. 10.1080/14693062.2011.582347.

[112] Joost G. Vogtländer, B. Baetens, A. Bijma, and E. Brandjes. Sustainable production and consumption: eco-efficiency, cleaner production and the next

industrial revolution. *Journal of Cleaner Production*, 18(10–11):971–975, 2010. 10.1016/j.jclepro.2010.02.008.

[113] Michael Maiko Matonya, László Pusztai, and István Budai. Optimizing healthcare business processes with process mining software: A comparative analysis. *Decision Making: Applications in Management and Engineering*, 7 (2):380–400, 2024. 10.31181/dmame7220241070.

[114] M. M. Matonya and I. Budai. Automated object-centric event log transformation and multi-criteria prioritisation for real-time industrial process monitoring. Manuscript submitted for publication, 2025.

[115] M. M. Matonya and I. Budai. Object-centric process mining for operational traceability and quality optimization in manufacturing: Genetic-inductive miner approach. *Procedia Computer Science*, 277:2224–2234, 2026. ISSN 1877-0509. 10.1016/j.procs.2026.02.260. URL `https://doi.org/10.1016/j.procs.2026.02.260`. Proceedings of the 7th International Conference on Industry of the Future and Smart Manufacturing (ISM 2025), University of Malta, Malta, 12–14 November 2025.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number:	DEENK/101/2026.PL
Subject:	PhD Publication List

Candidate: Michael Maiko Matonya
Doctoral School: Doctoral School of Informatics
MTMT ID: 10074167

List of publications related to the dissertation

Foreign language scientific articles in international journals (3)

1. **Matonya, M. M.**, Budai, I.: Cloud-Based Data-Driven Framework for Optimizing Operational Efficiency and Sustainability in Tube Manufacturing.
   Appl Syst Innov. 8 (4), 1-25, 2025. EISSN: 2571-5577.
   DOI: http://dx.doi.org/10.3390/asi8040100
   IF: 3.7 (2024)

2. **Matonya, M. M.**, Budai, I.: Object-Centric Process Mining Framework for Industrial Safety and Quality Validation Using Support Vector Machines.
   Appl Syst Innov. 9 (1), 1-21, 2025. EISSN: 2571-5577.
   DOI: http://dx.doi.org/10.3390/asi9010002
   IF: 3.7 (2024)

3. **Matonya, M. M.**, Pusztai, L. P., Budai, I.: Optimizing Healthcare Business Processes with Process Mining Software: a Comparative Analysis.
   Decis. Mak. Appl. Manag. Eng. 7 (2), 380-400, 2024. ISSN: 2560-6018.
   DOI: http://dx.doi.org/10.31181/dmame7220241070

Foreign language conference proceedings (1)

4. **Matonya, M. M.**, Budai, I.: Object-Centric Process Mining for Operational Traceability and Quality Optimization in Manufacturing: Genetic-Inductive Miner approach.
   Proc Computer Sci. 9 (1), 1-11, 2026. ISSN: 1877-0509.
   DOI: https://doi.org/10.1016/j.procs.2026.02.260

Foreign language abstracts (1)

5. **Matonya, M. M.**, Budai, I.: Object-Centric Process Mining for Inspection and Maintenance Error Detection in Sustainable Manufacturing.
   In: Book of Abstracts from 9th International Scientific Conference on Advances in Mechanical Engineering. Eds.: Tamás Mankovits, Mihály Csüllög, Department of Mechanical Engineering Faculty of Engineering, University of Debrecen, Debrecen, 66-66, 2024. ISBN: 9783036406176

List of other publications

Foreign language Hungarian book chapters (1)

6. **Matonya, M. M.**, Kocsi, B., Pusztai, L. P., Budai, I.: Production Planning, Scheduling and Risk Analysis in Manufacturing Operations by Robotic Process Automation.
In: Proceedings of the 11th International Conference on Applied Informatics (ICAI 2020) : Eger, Hungary, January 29-31, 2020. Eds.: Gergely Kovásznai, István Fazekas, Tibor Tómács, [CEUR Workshop Proceedings], Eger, 232-241, 2020, (CEUR Workshop Proceedings, ISSN 1613-0073 ; 2650.)

Foreign language scientific articles in Hungarian journals (1)

7. **Matonya, M. M.**: Innovation, Artificial Intelligence in Contingent Work-Force Management.
Int. J. Eng. Manag. Sci. 5 (1), 571-590, 2020. EISSN: 2498-700X.
DOI: http://dx.doi.org/10.21791/IJEMS.2020.1.47

Foreign language scientific articles in international journals (1)

8. Kocsi, B., **Matonya, M. M.**, Pusztai, L. P., Budai, I.: Real-Time Decision-Support System for High-Mix Low-Volume Production Scheduling in Industry 4.0.
Processes. 8 (8), 1-26, 2020. EISSN: 2227-9717.
DOI: http://dx.doi.org/10.3390/pr8080912
IF: 2.847

Total IF of journals (all publications): 10,247
Total IF of journals (publications related to the dissertation): 7,4

The Candidate's publication data submitted to the Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

08 April, 2026