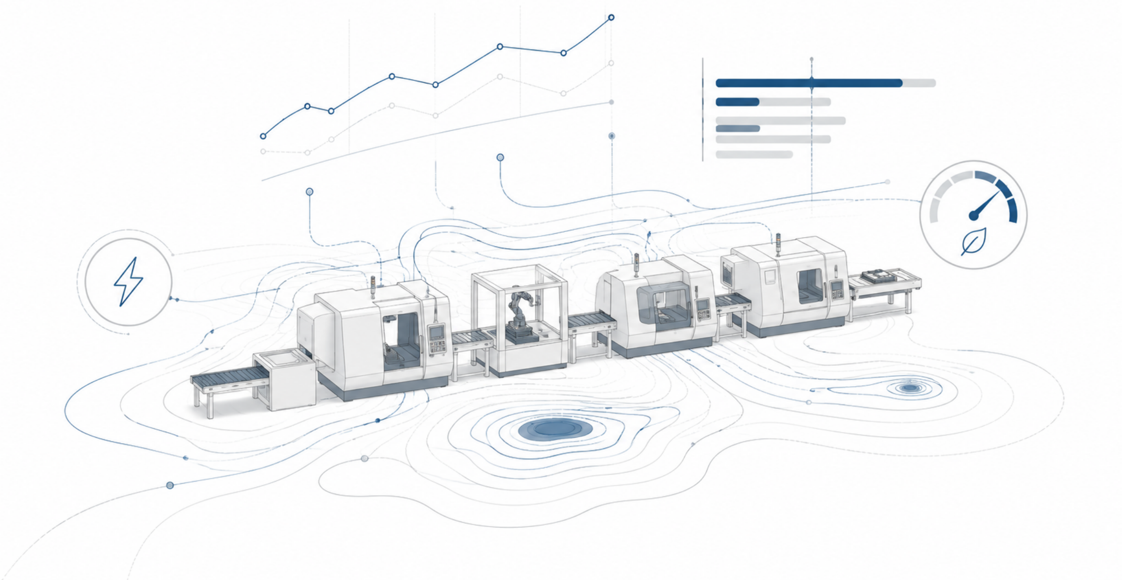

A Benchmarking-Based Framework for Energy Performance Variation Screening in Manufacturing Systems

A Case Study Using Local Benchmarking and Root Cause Screening

Master's thesis in Production Engineering

Jiana He
Zhuang Li

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2026
www.chalmers.se

MASTER'S THESIS 2026

A Benchmarking-Based Framework for Energy Performance Variation Screening in Manufacturing Systems

A Case Study Using Local Benchmarking and Root Cause Screening

Zhuang Li
Jiana He

Department of Industrial and Materials Science
Division of Production Systems
Chalmers University of Technology
Gothenburg, Sweden 2026

A Benchmarking-Based Framework for Energy Performance Variation Screening in Manufacturing Systems
A Case Study Using Local Benchmarking and Root Cause Screening
Zhuang Li and Jiana He

Supervisor: Qi Fang, Department of Industrial and Materials Science
Examiner: Mélanie Despeisse, Department of Industrial and Materials Science

Master's Thesis 2026
Department of Industrial and Materials Science
Division of Production Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Illustration related to energy-performance variation screening in manufacturing systems.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Abstract

Energy data are increasingly available in manufacturing systems, but valid interpretation depends on production context that is often incomplete. This thesis develops a benchmarking-based framework for screening and classifying within-resource energy-performance variation under limited data conditions. The framework is applied to an industrial case using three datasets provided by Volvo Trucks: 30-minute device-level energy records, stop logs, and stop-detail records.

The method integrates the datasets at a common resource–time-window level, constructs comparable active-like windows, establishes a local benchmark from historically lower-energy observations, and screens high benchmark-gap windows using the available stop information. Energy use per 30-minute window is treated as a screening variable rather than an output-normalised efficiency indicator. The application identifies distinct variation profiles across resources: some resources show larger relative gaps, whereas others show greater accumulated recorded energy above the local benchmark. Recorded stop events provide candidate explanations for only part of the flagged windows, demonstrating that richer data on output, product type, cycle time, shift, and machine state are required for causal diagnosis. The framework therefore supports transparent screening and prioritisation, but it does not by itself establish energy inefficiency, quantify savings potential, or confirm root causes.

Keywords: Energy performance variation, Local benchmarking, Manufacturing systems, Energy performance indicators, Industrial case study, Root-cause screening

Acknowledgements

We would like to express our sincere gratitude to Chalmers University of Technology for providing an inspiring academic environment and valuable learning opportunities throughout our master's studies. The knowledge, experiences, and challenges we encountered during our time at Chalmers have played an important role in both our academic and personal development.

We would also like to extend our heartfelt thanks to our supervisors and teachers for their guidance, constructive feedback, and encouragement throughout this thesis project. Their expertise and support have been invaluable in helping us complete this work.

Finally, we are deeply grateful to our families, friends, and fellow students for their continuous support, patience, and encouragement during this journey. Their understanding and motivation have helped us overcome challenges and stay focused on our goals.

This thesis marks the end of an important chapter in our academic lives, and we sincerely appreciate everyone who has contributed to making it possible.

Jiana He and Zhuang Li, Gothenburg, 06, 2026

List of Acronyms

Below is the list of acronyms used throughout this thesis, listed in alphabetical order.

EMS	Energy Management System
EnB	Energy Baseline
EnMS	Energy Management System
EnPI	Energy Performance Indicator
IIoT	Industrial Internet of Things
ISO	International Organization for Standardization
KPI	Key Performance Indicator
MES	Manufacturing Execution System
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SCADA	Supervisory Control and Data Acquisition
SEC	Specific Energy Consumption

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables used throughout this thesis.

Indices

r	Index for a resource, machine, process, or energy location
t	Index for a 30-minute analysis time window
i	Index for an observation within a comparable group
j	Index for a stop-related or operational variable

Sets

$\mathcal{R}$	Set of selected resources, machines, or energy locations
$\mathcal{T}$	Set of analysis time windows
$\mathcal{D}$	Integrated resource–time-window dataset used for the analysis
$\mathcal{C}_r$	Set of comparable active-like windows for resource r
$\mathcal{G}_r$	Good-performance reference group used to define the local benchmark for resource r
$\mathcal{H}_r$	Set of high benchmark-gap windows for resource r
$\mathcal{A}$	Set of problematic or high benchmark-gap windows used in comparison analysis
$\mathcal{N}$	Set of normal or reference windows used in comparison analysis

Parameters

Δt	Length of the aggregation time window; 30 minutes in this study

q_{good}	Share of comparable active-like windows used to define the good-performance reference group; $q_{\text{good}} = 0.20$ in this study
q_{gap}	Percentile threshold used to identify high benchmark-gap windows; $q_{\text{gap}} = 0.90$ in this study

Variables

$E_{r,t}$	Energy consumption of resource r in time window t
$Y_{r,t}$	Screening variable used in the analysis; in this study, $Y_{r,t} = E_{r,t}$
$Q_{r,t}$	Production output of resource r in time window t, when available
$SEC_{r,t}$	Specific Energy Consumption of resource r in time window t, when output data are available
$EnPI_{r,t}$	Energy Performance Indicator of resource r in time window t
$T_{\text{state},r,t}$	Duration of a selected machine state for resource r in time window t
μ_r^{low}	Low-load energy centre estimated for resource r
μ_r^{active}	Active-like energy centre estimated for resource r
τ_r^{active}	Active-like energy threshold used for comparable-window selection for resource r
B_r	Local benchmark value for resource r
$\Delta B_{r,t}$	Absolute benchmark gap between the observed value and the local benchmark for resource r in time window t
$g_{r,t}$	Relative benchmark gap for resource r in time window t
$Q_{90,r}$	90th percentile of positive relative benchmark gaps for resource r
N_r^H	Number of high benchmark-gap windows for resource r
$\bar{g}_r$	Mean relative benchmark gap in high benchmark-gap windows for resource r
E_r^{excess}	Total recorded energy above the local benchmark for resource r
$S_{r,t}$	Stop occurrence indicator for resource r in time window t
$X_{j,r,t}$	Value of the jth stop-related or operational variable for resource r in time window t
$X_{\text{fault},r,t}$	Fault-related stop duration for resource r in time window t
$X_{\text{wait},r,t}$	Waiting-related stop duration for resource r in time window t
$X_{\text{stop},r,t}$	Total stop-related duration for resource r in time window t
$X_{\text{uncoded},r,t}$	Uncoded stop duration for resource r in time window t

$\bar{X}_{j,\mathcal{A}}$ — Mean value of variable X_j in problematic or high benchmark-gap windows

$\bar{X}_{j,\mathcal{N}}$ — Mean value of variable X_j in normal or reference windows

ΔX_j — Difference in a selected stop-related or operational variable between problematic and reference windows

ΔX_{fault} — Difference in mean fault-related duration between problematic and reference windows

ΔX_{wait} — Difference in mean waiting-related duration between problematic and reference windows

ΔX_{stop} — Difference in mean total stop-related duration between problematic and reference windows

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

The manufacturing industry is one of the largest global consumers of energy and contributors to greenhouse gas emissions, and its operational performance has a direct impact on resource efficiency and value creation processes in manufacturing systems[1]. In modern manufacturing companies, energy management is not only a requirement for environmental compliance, but also an essential part of production and operations management. It affects cost control, capacity planning, equipment maintenance, and continuous improvement. With the widespread adoption of ISO 50001 energy management systems and ISO 50006 guidelines on energy performance indicators (EnPIs) and energy baselines (EnBs), the use of existing digital plant-floor data to evaluate energy performance in manufacturing systems has become an important research topic in manufacturing engineering[2, 3] .

In recent years, the development of Industrial Internet of Things and smart monitoring technologies has significantly improved the availability of energy and production data in manufacturing enterprises, providing a richer data foundation for energy performance evaluation[10]. However, transforming large amounts of data into actionable decision-support information is still a key challenge in energy management practice. Most existing energy-performance assessment methods operate at factory level and use indicators such as energy per unit of output or overall energy intensity. While these indicators support comparison, they cannot by themselves reveal differences among processes, equipment, and operating states, which limits their usefulness for detailed shop-floor analysis[12, 19]. Production routes and equipment states can materially affect energy use, and a single aggregate indicator may therefore conceal the source of observed variation[20]. Process-level assessment requires both a clearly defined indicator and sufficient production context to establish meaningful comparison conditions.

Manufacturing equipment also operates in states such as processing, waiting, warm-up, idling, and stopping, each with a different energy profile. The same recorded energy use can therefore correspond to different production loads and useful outputs. Energy data alone are insufficient for interpreting equipment energy performance; they must be considered together with production context such as machine state, stop events, cycle time, product type, and output[11, 21]. Model-based energy baselines can account for changing operating conditions when the relevant explanatory variables are available and reliable[3, 10]. When important variables are missing, however, a predicted deviation may reflect omitted production context rather than deteriorating performance. Under such data constraints, a transparent internal benchmark can provide a more defensible screening reference, provided that comparable windows and interpretation limits are stated explicitly.

Overall, the literature explains the components of an energy-performance assessment, proposes baseline and benchmarking methods, and identifies the production variables required for robust interpretation. Less explicit guidance is available on how these elements can be translated into as an auditable shop-floor screening procedure when energy and stop data are available but aligned output and machine-state data are incomplete. The research gap addressed in this thesis is therefore procedural rather than predictive: how to preserve meaningful within-resource comparison and operational follow-up while limiting the claims to what the available data can support.

Against this background, this study develops a process-level framework for screening energy-performance variation in manufacturing systems by integrating energy-monitoring data with digital stop-event records. The framework focuses on identifying and classifying variation in recorded energy use and screen potential explanatory factors. It does not claim to provide a complete assessment of energy efficiency.

1.2 Problem Statement

The industrial case provides three operational datasets supplied by Volvo Trucks: 30-minute device-level energy records, stop logs, and stop-detail records. These sources can be aligned by resource and time window, but they do not provide validated output quantities, product mix, cycle time, or complete machine-state labels at the same temporal resolution. Consequently, an output-normalised energy-efficiency indicator or a reliable predictive baseline model cannot be established from the available data. The methodological problem is therefore to create meaningful within-resource comparisons without treating low energy use as proof of high efficiency. This study addresses that problem through local benchmarking. For each resource, a benchmark reference is derived from comparatively low-energy active-like windows, and large positive benchmark gaps are used to identify periods for further investigation. The flagged windows are then compared with available stop-event and stop-detail records. The benchmark is an empirical screening reference rather than an ISO 50006 energy baseline, an engineering optimum, or evidence of energy inefficiency, and the stop-related findings are interpreted as candidate explanations rather than confirmed causes.

1.3 Research aim and question

The aim of this study is to develop and apply a benchmarking-based procedure for identifying, classifying, and supporting the interpretation of within-resource energy-performance variation under limited industrial data conditions. The procedure is evaluated through an industrial case study using the three Volvo-provided datasets.

This study addresses the following research question: How can energy-performance variation be identified, classified, and interpreted in manufacturing systems when production-context data are incomplete?

1.4 Research limitation

This study is limited to screening and interpreting within-resource energy-performance variation in one manufacturing line under constrained data availability. The January 2025 datasets contain energy records and stop-related logs but do not contain validated production output, good-part counts, product mix, cycle time, or complete machine-state labels at the same temporal resolution. Consequently, the study does not estimate thermodynamic, technical, or production-normalised energy efficiency, and variation in recorded energy use is not interpreted as variation in efficiency. The empirical analysis is restricted to 30-minute windows and internal local benchmarking within each resource. Cross-machine and cross-factory rankings are outside the scope because the monitored resources have different functions and energy scales. The benchmark represents a screening reference derived from historically lower-energy active-like windows; it is not a causal model, a certified energy baseline, or a target validated against production output. High benchmark-gap windows therefore indicate periods for further investigation, not proven inefficiency or quantified energy-saving potential. The one-month case period and incomplete stop coding further limit causal interpretation and generalisability.

2

Methodology

2.1 Overall Research approach

This research examines energy-performance variation in industrial manufacturing systems under limited data conditions. As illustrated in Figure 2.1, the study follows a structured research pathway that combines a literature review with an industrial case study. The core research process consists of a literature review, assessment framework development, industrial case application, and results interpretation.

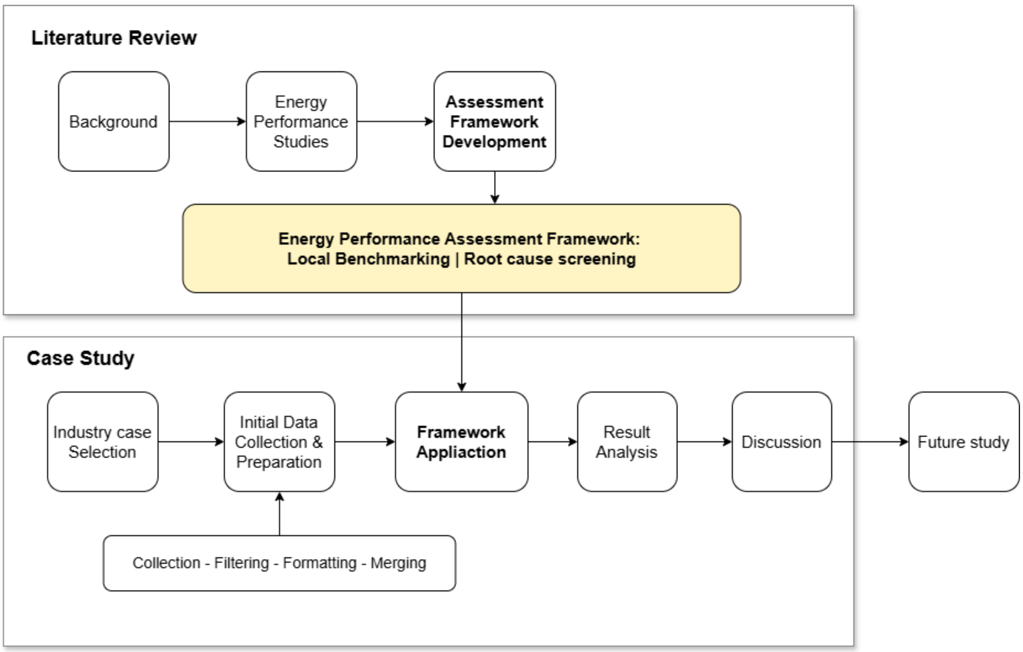

Figure 2.1: Overall Methodological approach

The first stage establishes the theoretical foundation through a structured literature review. Research on energy-performance assessment, energy baselines, benchmarking, and manufacturing energy management was reviewed to identify relevant concepts, assessment methods, and data requirements. Particular attention was given to the relationship between method selection and data availability. The review therefore provided the basis for choosing an assessment approach consistent with the characteristics and limitations of the industrial dataset.

The second stage focuses on the industrial case study. It includes data collection, cleaning, formatting, and integration of energy-consumption records and stop-event logs. A benchmarking-based assessment procedure is then applied to identify windows with high energy use relative to each resource's local reference. Variation classification is used to distinguish differences in gap magnitude, temporal distribution, and stop relation.

The final stage interprets the results and evaluates the practical applicability of the framework. The observed variation profiles are examined for candidate operational explanations, while unresolved windows are used to identify data gaps. The case-study findings inform recommendations for future data integration and further development of manufacturing energy-performance assessment.

2.2 Literature study method

The literature review followed the PRISMA framework [5]. As shown in Figure 2.2, study selection comprised identification, screening, eligibility assessment, and final inclusion.

Identification
Scopus records identified: $n = 154$

↓

Duplicate removal
Duplicates removed: $n = 2$; records remaining: $n = 152$

↓

Title and abstract screening
Records retained for full-text assessment: $n = 142$

↓

Eligibility and inclusion
Studies included in the final review: $n = 126$

Figure 2.2: PRISMA-based literature review process.

Search strategy and identification: Scopus was used as the primary database, and a Boolean search string was designed around three dimensions: energy performance, manufacturing, and benchmarking. The initial search returned 154 records. After two duplicates were removed, 152 records remained.

Title and abstract screening: Titles and abstracts were screened to exclude studies outside the manufacturing domain, such as architecture and transportation, and records without accessible full text. After this stage, 142 articles were retained for full-text assessment.

Full-text assessment and eligibility: The remaining articles were reviewed in full against predefined inclusion criteria derived from the research objectives. Studies with limited relevance to manufacturing energy-performance assessment or insufficient methodological detail were excluded. The final review sample comprised 126 studies.

Analysis and synthesis: The included studies were analysed to identify common concepts, methodological approaches, and data requirements in industrial energy-performance assessment. The synthesis was organised around three themes that support the objectives of this study:

1. **Energy-performance assessment frameworks:** approaches used to define and structure assessment in manufacturing systems, including assessment objectives, system boundaries, performance indicators, reference conditions, and the integration of energy-management principles into manufacturing operations.

2. **Assessment and benchmarking methods:** methods used to evaluate energy performance, including performance indicators, energy baselines, benchmarking, and comparative assessment techniques. The review examined how these methods identify performance variation and support decision-making under different industrial conditions.

3. **Data requirements and influencing factors:** production and operational variables required to construct valid indicators and reference conditions, including output, product mix, cycle time, machine state, and stop information. This theme was used to assess how data availability constrains method selection and the strength of the resulting claims.

2.3 Benchmarking-based four-step methodological framework

This study develops a benchmarking-based framework for screening energy-performance variation in a manufacturing system under data constraints. The purpose is to provide a transparent procedure for identifying periods that differ from a resource's own local reference and for screening candidate explanations in the available stop-related records. The framework does not estimate production-normalised energy efficiency and does not use a predictive baseline model as its primary assessment route.

Manufacturing energy analysis normally requires energy data to be integrated with operational information such as stop events, machine states, cycle time, product type, and production output [24]. Model-based baselines likewise require a suitable indicator and relevant explanatory variables to construct a defensible reference [23]. The Volvo case provides 30-minute energy records and two stop-related datasets, but not validated output quantity, cycle time, product type, or complete machine-state records. A predictive baseline could therefore produce numerical estimates without supporting reliable interpretation. Local benchmarking is adopted as the main route because it matches the available evidence and keeps the comparison within each resource.

The proposed framework consists of four steps, as shown in Figure 2.3. First, the available energy and stop-related data are cleaned and integrated into a common resource-window dataset. Second, a screening indicator is defined and comparable active-like windows are constructed under the limitations of the available data. Third, local benchmarking is used to identify high benchmark-gap windows.

Fourth, the identified high-gap windows are classified and screened using the available stop-related variables.

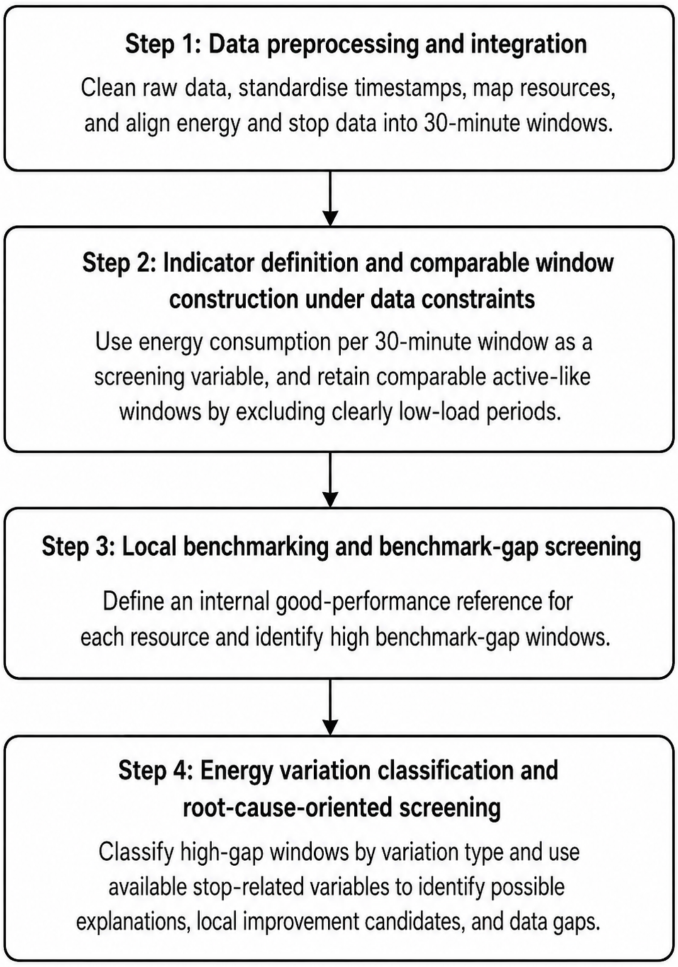

Figure 2.3: Four-step benchmarking-based framework for energy performance variation assessment.

Step 1: Data preprocessing and integration

The three raw operational datasets were supplied by Volvo Trucks. `Based-on Device_January2025.xlsx` contains 30-minute energy records by monitored location; `Stop log_January2025.xlsx` contains event start and end times and stop durations; and `Stop details (without cause log)_January2025.xlsx` contains stop categories, alarm information, and shift-related fields. The study did not generate or supplement these raw records. The energy dataset is interval based, whereas the two stop datasets are event based.

Because these data sources have different temporal structures, the timestamps are first cleaned and standardised. Resource names, work areas, and energy locations are also standardised to support data matching. A mapping between work areas and energy locations is then created so that stop-related records can be linked to the corresponding energy measurement points.

For stop events that overlap more than one 30-minute interval, the event duration is allocated to each time window according to the actual overlap duration. This avoids assigning a complete stop event to a single timestamp when the event continues across several windows. After this processing, each row in the final dataset represents one resource in one 30-minute window. The integrated dataset contains energy consumption and stop-related variables, such as total stop duration, fault duration, waiting duration, and uncoded stop duration. This resource-window structure forms the basis for the following steps of the analysis.

Step 2: Indicator definition and comparable window construction under data constraints

The second step defines the screening indicator and constructs comparable windows under the limitations of the available data. In this case, detailed production context is limited. The available dataset does not include reliable output quantity, cycle time, product type, or detailed machine-state records. Therefore, the analysis does not use an output-normalised indicator. Instead, energy consumption per 30-minute window is used as the screening variable:

$$Y_{r,t} = E_{r,t} \tag{2.1}$$

In Equation 2.1, $Y_{r,t}$ is the recorded energy use of resource r in 30-minute window t. It is a screening variable, not an output-normalised EnPI and not a measure of efficiency. It is used only to identify windows that are high relative to the same resource's local reference.

To avoid comparing clearly different operating conditions, comparable windows are constructed before benchmarking. Since real machine-state labels are not available, an energy-state proxy is used to remove clearly low-load or non-comparable periods. For each resource, the energy distribution is analysed to identify a low-load centre, an active-like centre, and an active threshold. Windows below this threshold are treated as clearly low-load or non-comparable periods and are excluded from the main benchmarking analysis. The remaining windows are regarded as comparable active-like windows.

This step does not replace real machine-state classification. However, it reduces the influence of obviously different operating states and makes the following local benchmarking analysis more meaningful under the available data conditions.

Step 3: Local benchmarking and benchmark-gap screening

The third step establishes a local benchmark and identifies high benchmark-gap windows. This is the main quantitative step of the framework. Instead of comparing resources with external standards or with other factories, the method compares each resource with its own historical good-performance behaviour. This within-resource comparison reduces the influence of differences in machine type, process function, and energy scale.

For each resource r, a good-performance subset G_r is selected from its comparable active-like windows. It contains the lowest 20% of screening-variable values for that resource:

$$G_r = \{i : Y_{r,i} \leq Q_{0.20}(Y_r)\}\,. \tag{2.2}$$

The resource-specific local benchmark is then defined as the median of this subset:

$$B_r = \text{median}(Y_i, i \in G_r) \tag{2.3}$$

In Equation 2.3, B_r is the internal benchmark for resource r. The median is used instead of a single minimum to reduce sensitivity to isolated low values. The 20% share is a pragmatic screening parameter rather than a universal definition of good performance; it retains a group of lower-energy observations while limiting the influence of the centre of the active-like distribution. The benchmark remains conditional on the active-like proxy and is not interpreted as a theoretical optimum or an ISO 50006 energy baseline.

After the local benchmark is defined, the absolute benchmark gap is calculated for each window:

$$\Delta B_{r,t} = Y_{r,t} - B_r \tag{2.4}$$

Equation 2.4 shows how much the actual energy consumption exceeds the local benchmark. Since different resources have different energy scales, a relative benchmark gap is also calculated:

$$g_{r,t} = \frac{Y_{r,t} - B_r}{B_r} \tag{2.5}$$

Positive values in Equation 2.5 indicate that the window is above the local benchmark. To identify the most relevant high-energy windows, a resource-specific threshold is applied. In this study, windows above the 90th percentile of positive relative benchmark gaps within each resource are defined as high benchmark-gap windows:

$$g_{r,t} \geq Q_{90,r} \tag{2.6}$$

The output of this step includes the local benchmark, absolute and relative benchmark gaps, high benchmark-gap windows, and resource-level metrics such as mean relative gap and total recorded energy above the local benchmark. The latter is a descriptive difference and must not be reported as achievable energy savings.

Step 4: Energy variation classification and root-cause-oriented screening

The fourth step describes the high benchmark-gap windows and screens candidate explanations using the available stop-related data. It does not confirm a root cause. The purpose is to distinguish variation profiles and to test how much of the flagged behaviour is represented in the stop records.

For each resource, the analysis reports the number and share of retained windows flagged as high gap, the magnitude of the relative gap, the distribution of flagged windows across dates, hours, and available shift categories, and the share with recorded stops. Because the high-gap rule is the resource-specific 90th percentile, approximately one tenth of positive-gap windows will be flagged by construction. The raw number of flags is therefore not used on its own as evidence that one resource varies more frequently than another. Severity and temporal concentration are interpreted separately.

Variation is described as smaller or larger according to relative-gap magnitude and as temporally concentrated or dispersed according to whether flagged windows recur in the same dates, hours, or shift categories. These descriptions are empirical profiles rather than universal classes. Stop-related windows can be examined using fault, waiting, and uncoded-stop categories; no-stop high-gap windows instead indicate that the available stop records do not explain the observed variation.

The stop-relation screen divides high benchmark-gap windows into windows with and without an overlapping stop record. The resulting shares show the coverage of the stop data, not causality. A high with-stop share permits more detailed candidate screening, whereas a high no-stop share indicates a need for additional production-context variables.

For the subset of high-gap windows with recorded stops, the study compares stop-related variables between high-gap windows and normal reference windows. The comparison is expressed as:

$$\Delta X_j = \overline{X}_{j,A} - \overline{X}_{j,N} \tag{2.7}$$

In Equation 2.7, the first group represents high benchmark-gap windows, the second group represents normal reference windows, and the compared variable represents a stop-related variable, such as fault duration, waiting duration, or total stop duration. A positive value means that the variable is higher in high-gap windows and may be considered as a local candidate explanation. If the value is close to zero or negative, the variable cannot be treated as a general explanation for the high-energy behaviour.

The stop detail data are then used to identify local improvement candidates, such as waiting-related stops, equipment-fault-related stops, and uncoded stop categories. These candidates are interpreted as screening-level findings, not as confirmed root causes. For high-gap windows without recorded stops, the result is treated as a data-gap diagnosis. In such cases, further root-cause confirmation would require additional production-context data, such as output quantity, cycle time, product type, shift information, and detailed machine-state records.

The final output of this step includes variation types, stop-related local candidates, and data gaps that should be addressed in future data collection and improvement work.

3

Results

3.1 Results of the literature study

3.1.1 Overview

Figure 3.1 shows the chronological distribution of the 126 studies included in the final review from 2000 to 2026. Publication activity on manufacturing energy performance increased over this period, particularly after 2012. The annual count decreased around 2022 and subsequently rose towards 2025.

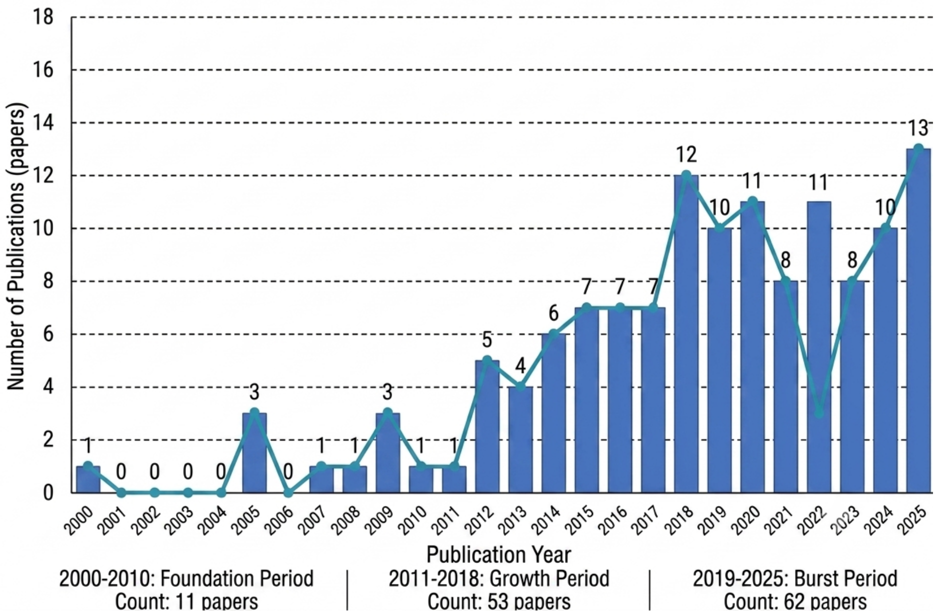

Figure 3.1: Publication Trends of Selected Studies by Year

Geographically, the Fig 3.2 shows that the United States, Italy, China, and Germany are the most active contributors, which reflects the strong industrial focus of this research area.

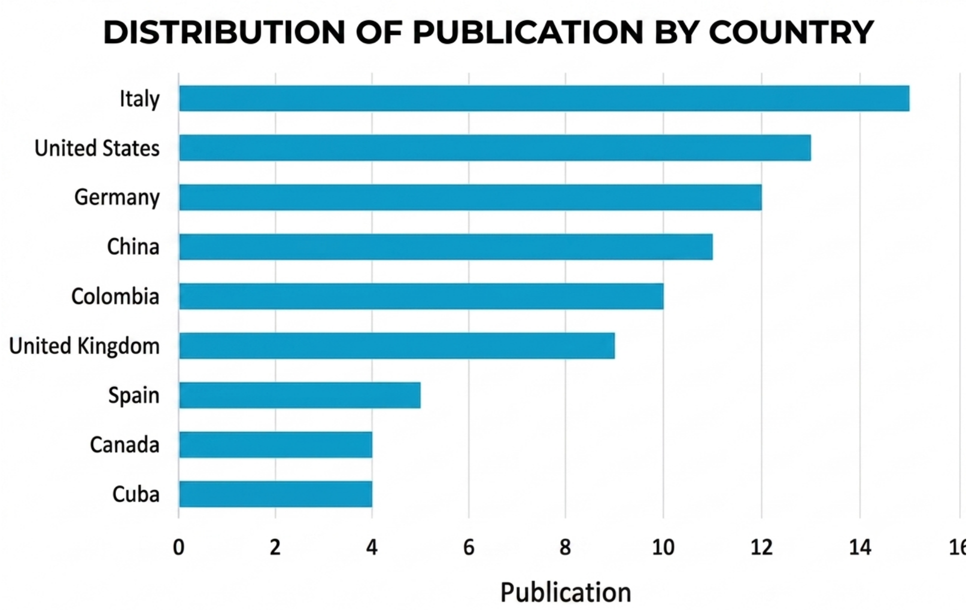

Figure 3.2: Distribution of Publication by country

Key authors include Introna, V., Martini, F., and Santos, V. S., while major funding bodies include the National Natural Science Foundation of China and the European Commission. In addition, as shown in Figure 3.3, Engineering (28.3%) and Energy (23.5%) are the dominant subject areas, followed by Environmental Science (14.3%).

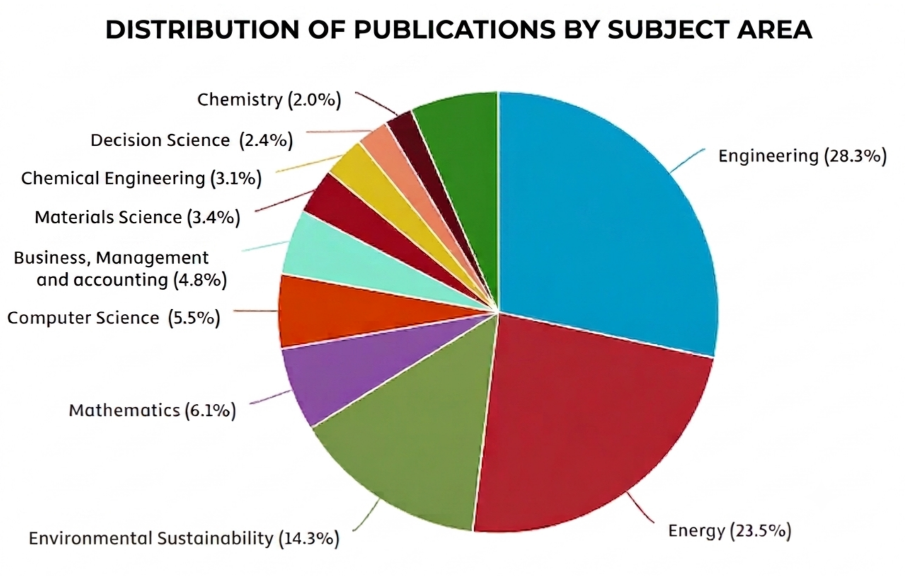

Figure 3.3: Publication Distribution by Subject Area

Overall, the literature confirms that research on energy benchmarking, performance indicators, and efficiency assessment methods is well established, providing a relevant foundation for the present thesis.

3.1.2 Main Themes Identified in the Literature

To support the development of the case-study methodology, the reviewed literature was grouped into three thematic categories: energy performance assessment frameworks, assessment and benchmarking methods, and data requirements and influencing factors.

The first category concerns energy performance assessment frameworks. Most studies describe energy-performance assessment as a structured process involving the definition of system boundaries, the selection of performance indicators, the establishment of reference conditions, and the interpretation of energy-consumption behaviour. Several studies adopt principles derived from energy-management standards such as ISO 50001, where energy performance is assessed through continuous monitoring and comparison against defined baselines[14, 15]. Across the literature, energy consumption is generally analysed in relation to operating conditions and production activities rather than as a standalone measure. This perspective highlights the importance of considering contextual factors when evaluating industrial energy performance.

The second category covers assessment and benchmarking methods. EnPIs are among the most frequently reported assessment tools and are commonly used to monitor energy consumption over time or compare performance between resources, processes, etc. The reviewed studies indicate that different types of EnPIs are applied depending on the assessment objective and the available data[16] As summarised in Table 3.1,

Table 3.1: EnPI Classification

Type	**Expression**	**Description**	**Advantages**	**Limitations**
1.Absolute energy-based	$EnPI = Energy$	Total energy consumption	Stable data, easy to compute	Not normalized, limited comparability
2.Energy efficiency (SEC)	$EnPI = \frac{Energy}{Output}$	Energy per unit of output	Comparable, suitable for efficiency evaluation	Requires reliable output data
3.State-based normalization	$EnPI = \frac{Energy}{Time_{state}}$	Energy per unit time in a given state	Enables state-level analysis (e.g., idle, processing)	Depends on accurate state classification

they are commonly used indicators include absolute energy-based, output-normalised, and state-based measures, each with different data requirements and areas of application. The literature also suggests that indicator selection should be aligned with the characteristics of the available dataset and the purpose of the assessment. The main considerations influencing EnPI selection are summarised in Table 3.2,

Table 3.2: EnPI Selection Criteria

Condition	Recommended EnPI	Description
Output data unavailable or unreliable	Type 1 (Energy)	Ensures computability
Output data available and aligned	Type 2 (SEC)	Suitable for efficiency comparison
Focus on operating states	Type 3 (State-based)	Supports state-level analysis
Multi-dimensional analysis	Combination of EnPIs	Improves analysis coverage

Baseline-based approaches are also widely applied, where historical energy-consumption data are used as reference values for identifying deviations and performance changes. In addition, a number of studies employ statistical and data-driven methods, including regression models and machine-learning techniques, to predict expected energy consumption under varying operating conditions. However, these methods generally require detailed production and process data. In situations where such information is unavailable, benchmarking approaches are frequently reported as practical alternatives[17, 18]. Rather than predicting expected energy consumption from multiple explanatory variables, benchmarking evaluates performance by comparing observations across different time periods, operating conditions, or resources. Internal benchmarking, in particular, uses historical data from the same system as a reference and has been applied to identify abnormal energy-consumption patterns, performance deterioration, and opportunities for improvement[18]. As a result, benchmarking is commonly reported in studies where energy consumption data are available but detailed production information is limited.

The third category concerns data requirements and influencing factors. The reviewed studies consistently indicate that the validity of an EnPI, baseline, or benchmark depends on the availability and temporal alignment of relevant production-context variables. Output quantity, product mix, cycle time, machine state, shift pattern, and stop information are repeatedly identified as factors that influence recorded energy use and the interpretation of performance variation [3, 21, 20]. When these variables are unavailable or incomplete, output-normalised indicators and predictive baselines may confound operational changes with changes in energy performance. This theme therefore establishes data availability as a method-selection criterion and supports the use of bounded internal comparison when detailed production context cannot be reconstructed.

3.1.3 Literature-Based Method Selection

The literature review showed that energy-performance assessment in manufacturing usually includes three main elements: defining EnPIs, setting a reference or baseline for comparison, and explaining deviations by using operational-context information. These principles are shown in energy-management standards and previous industrial energy-assessment studies[3, 21].

Several methods have been used to apply these principles. When detailed production information is available, output-normalised indicators and predictive baseline models are often used to assess energy performance under different production conditions. Internal benchmarking methods use historical operating data to set performance references. These methods have been widely used to monitor energy behaviour, identify deviations, and find opportunities for improvement in industrial systems[21].

Based on these findings and the available case data, local benchmarking was selected as the primary assessment route. The three operational datasets were first integrated at a common resource–time-window level. A screening variable and comparable active-like windows were then defined before comparison. For each resource, an internal benchmark was constructed from lower-energy observations within those windows, after which absolute and relative benchmark gaps were calculated. Finally, the high-gap windows were described by gap magnitude, temporal distribution, and stop relation, and the stop-detail records were used to screen candidate explanations. This sequence applies the assessment principles identified in the literature without implying that the available energy values constitute an efficiency measure.

3.2 Benchmarking Framework Application

This section reports the application of the four-step framework to the three January 2025 datasets provided by Volvo Trucks. The results follow the same order as the methodology: data integration, construction of comparable active-like windows, local benchmarking, and variation classification with stop-related screening.

All benchmark gaps reported below are departures from a resource's own local energy reference. They are not output-normalised efficiency losses. Likewise, recorded energy above the benchmark is a descriptive screening quantity and not an estimate of achievable savings.

Step 1: Data Preprocessing and Integration Results

The Volvo energy dataset contained 12,695 rows of 30-minute observations, the stop-log dataset contained 1,865 rows, and the stop-detail sheet contained 1,753 event rows. Timestamps and resource identifiers were standardised before the event durations were allocated to overlapping 30-minute windows.

Six resources were linked consistently across the energy and stop-related datasets: Robotcell, Robotcell_Bin_picking, Borrmaskin, Kuggfrasmaskin, Gradmaskin, and Skavmaskin. Each integrated row contains the resource, 30-minute energy value, and any overlapping stop duration and stop-detail information.

The resulting resource–time-window dataset provides the common unit of analysis used in Steps 2–4. Unmatched or unavailable production variables were not imputed, because doing so would create unsupported operating context.

Step 2: Comparable-Window Selection Results

Before local benchmarking, comparable active-like windows were selected for each resource based on its own energy distribution. Figure 3.4 shows the low-load centre, active-like centre, and active threshold estimated for each resource.

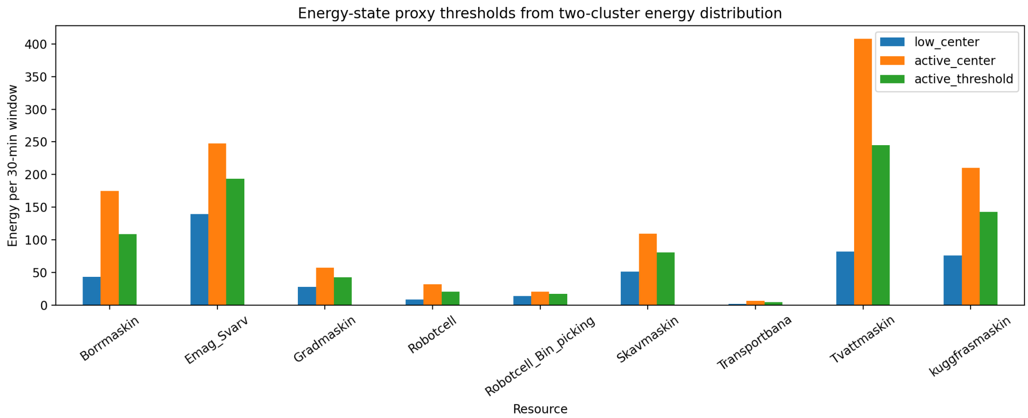

Figure 3.4: Energy-state proxy thresholds for comparable-window selection.

The results show clear differences in energy level between resources. Kuggfrasmaskin and Borrmaskin have the highest active-like energy centres, which is consistent with their role as machining resources. In contrast, Robotcell and Robotcell_Bin_picking have much lower active-like centres, which reflects their handling and auxiliary functions.

The separation between low-load and active-like centres also shows that the energy distribution can be used to distinguish lower-load windows from more active-like windows. After this selection, the following benchmarking analysis was conducted only on the retained comparable windows.

Step 3: Local Benchmarking and High Benchmark-Gap Results

Local benchmarking was first examined at the individual resource level. Figure 3.5 shows the local benchmarking result for Robotcell. The blue line represents the actual energy consumption per 30-minute window, the orange line represents the local benchmark, and the marked points indicate high benchmark-gap windows.

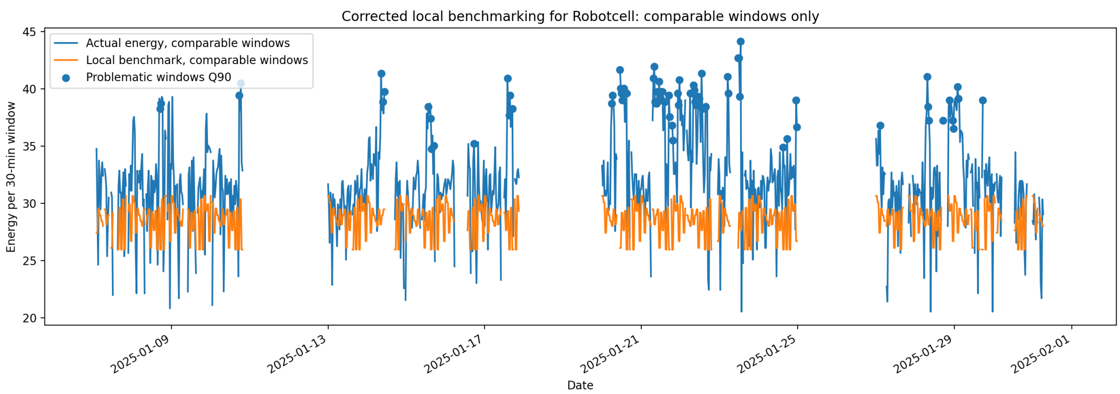

Figure 3.5: Local benchmarking result for Robotcell.

For Robotcell, several windows were clearly above the local benchmark.

These high benchmark-gap windows were especially visible in the later part of January. This indicates that Robotcell had several periods where its energy consumption was noticeably higher than its own internal benchmark level.

Figure 3.6 reports the number of windows flagged by the resource-specific 90th-percentile rule. Gradmaskin has the largest count and Robotcell_Bin_picking the smallest. These counts should not be interpreted as a direct comparison of recurrence: the rule selects approximately the upper decile within each resource, so count differences mainly reflect the number of retained comparable windows and possible ties at the threshold.

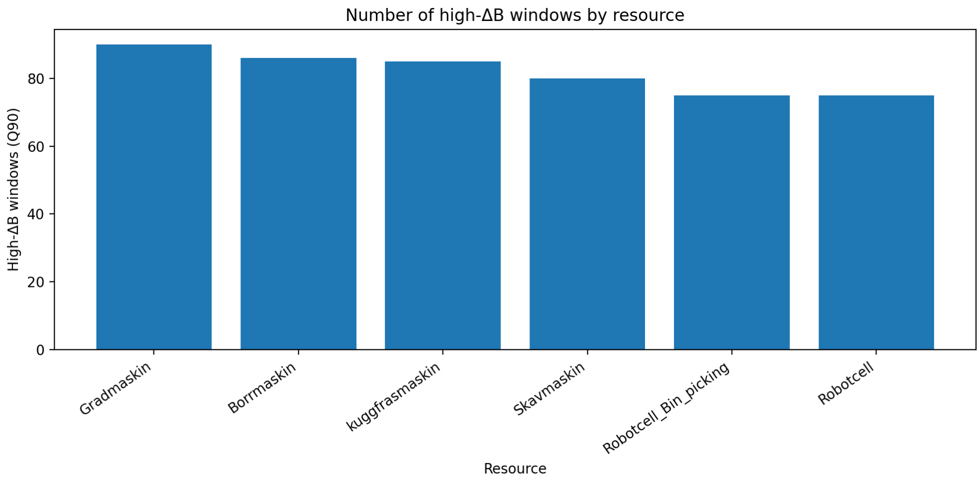

Figure 3.6: Number of high benchmark-gap windows by resource.

However, the number of high-gap windows alone does not show the magnitude of the deviation. Therefore, the mean relative benchmark gap was also calculated. Figure 3.7 shows that Robotcell had the highest mean relative benchmark gap, at approximately 47%. Borrmaskin had the second highest value, at approximately 33%. Skavmaskin and Kuggfrasmaskin showed medium-level relative gaps, while Gradmaskin had the lowest mean relative gap.

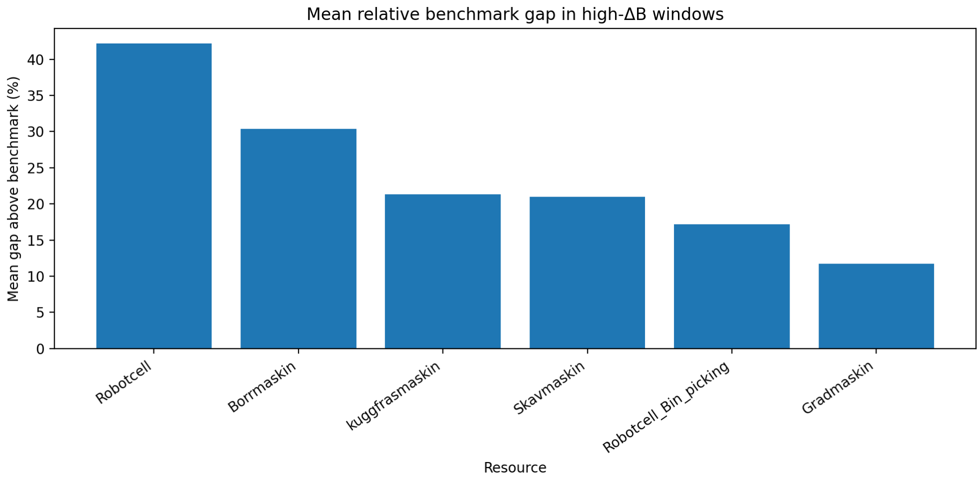

Figure 3.7: Mean relative benchmark gap in high-gap windows by resource.

The severity result is more informative than the raw flag count. Robotcell has the largest mean relative gap, whereas Gradmaskin has the lowest. The figure therefore distinguishes the magnitude of the screened deviation without claiming that the resources with more retained flags are intrinsically less efficient.

The accumulated recorded energy above the local benchmark was also calculated. As shown in Figure 3.8, Borrmaskin had the largest total, followed by Kuggfrasmaskin and Skavmaskin. Robotcell had the largest relative gap, but its accumulated difference was lower than that of the main machining resources. This quantity is used for prioritisation only and is not interpreted as recoverable energy savings.

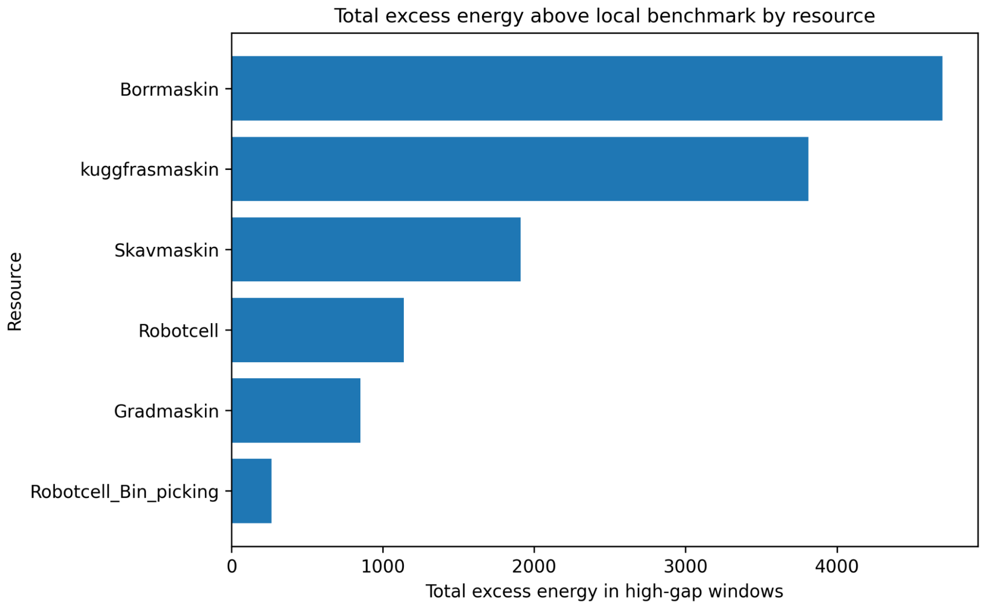

Figure 3.8: Total recorded energy above the local benchmark by resource.

Taken together, the local benchmarking results identify different prioritisation perspectives. Robotcell warrants attention because of its large relative deviation, while Borrmaskin has the largest accumulated energy difference above its local benchmark. Kuggfrasmaskin combines a medium relative gap with a comparatively large accumulated difference. The high flag count for Gradmaskin is not used as evidence of higher recurrence because the percentile rule fixes the screening share by design.

Step 4: Variation Classification and Candidate-Explanation Screening Results

The next result concerns the relationship between high benchmark-gap windows and recorded stop events. Figure 3.9 divides the high benchmark-gap windows into two groups: windows with recorded stops and windows without recorded stops.

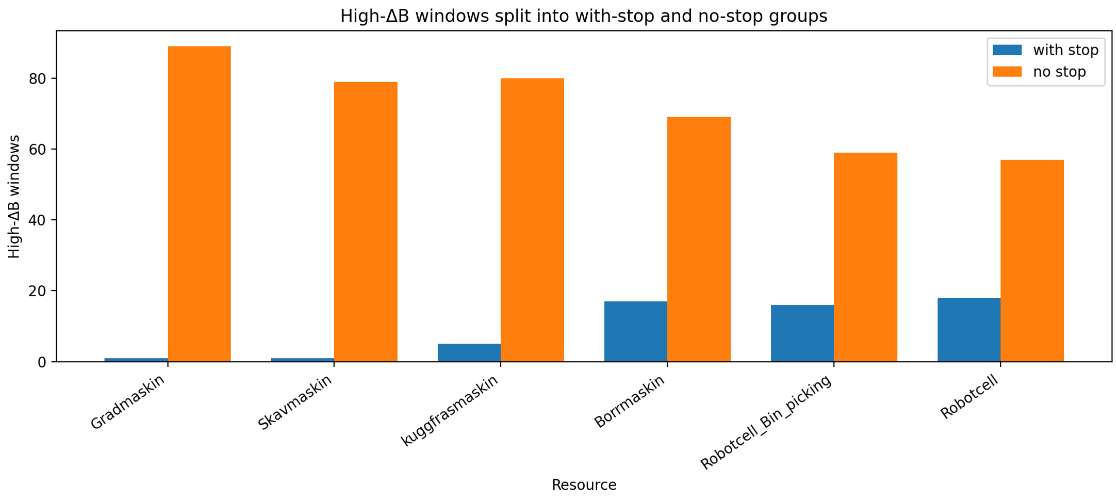

Figure 3.9: High benchmark-gap windows split by stop occurrence.

The results show that the high benchmark-gap windows did not follow one single stop-related pattern. For Gradmaskin, Skavmaskin, and Kuggfrasmaskin, most high-gap windows were found in the no-stop group. For Borrmaskin, Robotcell, and Robotcell_Bin_picking, the number of high-gap windows with recorded stops was higher than for the other resources, but no-stop windows still remained visible.

Recorded stop occurrence therefore covers only part of the high benchmark-gap pattern. Some flagged windows overlap recorded stops, whereas others have no overlapping stop event. Co-occurrence does not by itself establish that the stop caused the energy difference.

For the high-gap windows with recorded stops, the most visible stop-category candidates were summarised. Figure 3.10 reports stop duration within high benchmark-gap windows by resource and category.

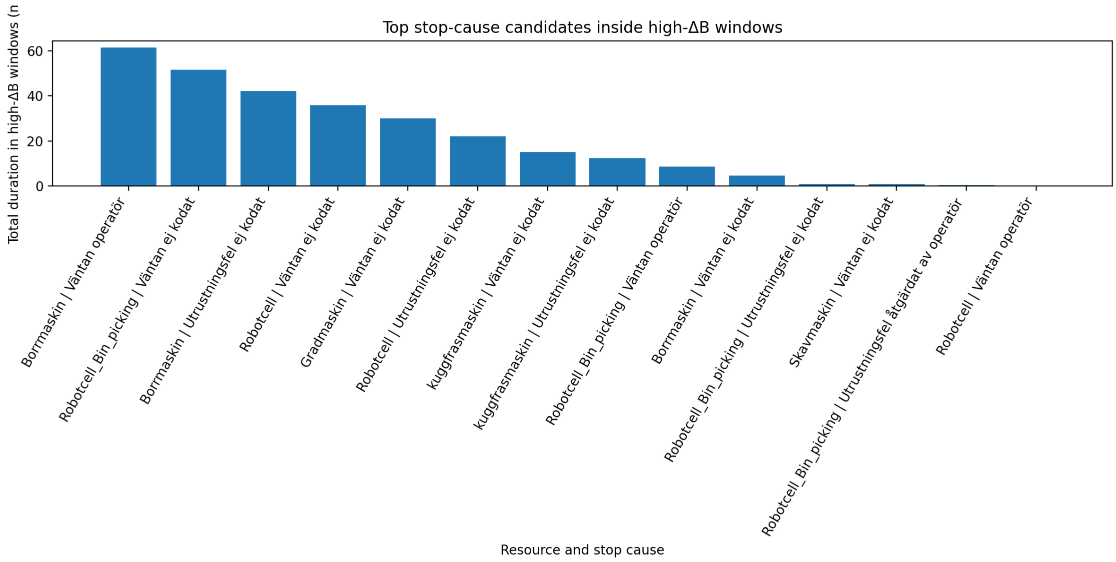

Figure 3.10: Stop-category candidates within high benchmark-gap windows.

The most visible categories were waiting for personnel, uncoded waiting, uncoded equipment faults, and waiting for operators. Robotcell_Bin_picking had the largest stop duration associated with waiting for personnel. Kuggfrasmaskin and Robotcell also showed waiting-related categories, while Borrmaskin showed both waiting-related and equipment-fault categories.

These categories are candidate explanations for the subset of flagged windows with recorded stops; they are not confirmed root causes. The large no-stop share for several resources shows that the stop records do not represent the full operating context required for causal interpretation.

Table 3.3 summarises each resource by relative-gap severity, accumulated energy above the local benchmark, and stop relation. Frequency labels are deliberately excluded because the resource-specific percentile rule fixes the flagging proportion.

Table 3.3: Summary of high benchmark-gap patterns by resource

Resource	**Benchmark-gap profile**	**Stop-related pattern**	**Main result**
Robotcell	Largest mean relative gap	Mixed stop/no-stop pattern	Highest relative benchmark gap; priority resource for further analysis.
Borrmaskin	Second-largest mean relative gap; largest accumulated energy above benchmark	Mixed stop/no-stop pattern	Largest accumulated energy above benchmark; waiting and equipment-fault candidates are visible.
Kuggfrasmaskin	Medium relative-gap severity; comparatively large accumulated energy above benchmark	Mostly no-stop high-gap windows	Medium relative gap and comparatively large accumulated energy above benchmark.
Skavmaskin	Medium relative-gap severity	Mostly no-stop high-gap windows	High-gap windows mainly occurred without recorded stops.
Gradmaskin	Lowest mean relative-gap severity	Mostly no-stop high-gap windows	Large flag count reflects the larger retained comparison set, not proven higher recurrence.
Robotcell Bin Picking	Medium relative-gap severity	Partly stop-related	Waiting for personnel was the main visible stop-related candidate.

The available month of data showed visible temporal concentration for selected resources, including several Robotcell flags in the later part of January. However, a one-month case period and incomplete production-state information are insufficient to establish stable regular-versus-irregular behaviour. Temporal concentration is therefore reported as an exploratory pattern rather than a deterministic classification.

3.3 Summary of Application Results

The results show that the four-step analysis procedure can transform the January 2025 energy and stop-related records into comparable resource-window results. Based on this structure, different high benchmark-gap patterns were identified across the selected resources.

The resource-level results prioritise Robotcell, Borrmaskin, and Kuggfrasmaskin for different reasons. Robotcell has the largest mean relative benchmark gap. Borrmaskin has the largest accumulated recorded energy above its benchmark. Kuggfrasmaskin combines a medium relative gap with a comparatively large accumulated difference. Gradmaskin's larger flag count is not interpreted as more frequent variation because the upper-decile screening rule and the number of retained windows determine that count.

Within the flagged windows that overlap stop records, waiting-related and uncoded stop categories are the most visible operational candidates. These categories occur for Robotcell_Bin_picking, Kuggfrasmaskin, Robotcell, and Borrmaskin, while Borrmaskin also shows equipment-fault categories. The findings support targeted follow-up, but they do not confirm that the recorded stops caused the benchmark gaps.

Overall, the results demonstrate that local benchmarking can screen and prioritise within-resource energy variation under limited data conditions. The stop-related analysis adds candidate operational context for a subset of windows and, equally importantly, identifies where the available data are insufficient. The findings do not constitute a full assessment of energy efficiency.

4
Discussion

This chapter interprets the findings from the benchmarking-based energy-variation screening. It focuses on the meaning of the local benchmark, differences in benchmark-gap magnitude and accumulated energy above the reference, the coverage of the stop-related data, practical implications for Volvo Trucks, and the interpretation boundaries of the case study.

The results show that local benchmarking distinguishes several resource-level variation profiles. Some resources have larger relative gaps, while others have greater accumulated recorded energy above their local reference. The stop screen adds operational context for a subset of flagged windows, but it does not establish causality. The framework should therefore be understood as a screening and prioritisation procedure rather than a final diagnosis.

4.1 Methodological Value and Research Contribution of the Framework

A predictive energy baseline was not adopted as the main method because the available case data omit validated output, cycle time, product type, and complete machine-state variables. Baseline methods can support comparison when relevant explanatory variables are reliable [23]; under the present conditions, however, model residuals would be ambiguous and could not support stronger conclusions than the observed data. Local benchmarking is better aligned with the case evidence because it constructs a transparent reference within each resource, flags windows above that reference, and then examines whether the available operational records overlap those windows.

The four-step procedure operationalises the three themes identified in the literature review. In relation to energy-performance assessment frameworks, Steps 1 and 2 make the system boundary, unit of analysis, screening indicator, and comparability rules explicit before performance variation is evaluated. In relation to assessment and benchmarking methods, Step 3 constructs a resource-specific internal reference and quantifies both relative and accumulated benchmark gaps rather than assuming that different resources share a common operating level. In relation to data requirements and influencing factors, Step 4 uses the available stop records to screen candidate operational context and retains high-gap windows without recorded stops as unresolved cases. Missing production context is therefore treated both as a constraint on method selection and as a result that defines the data required for stronger follow-up analysis.

The research gap addressed by this study is the limited methodological guidance between high-level energy-performance assessment frameworks and data-intensive baseline or efficiency models when production-context data are incomplete. The contribution is not a new universal benchmark, an output-normalised EnPI, or a causal diagnosis. It is a bounded and auditable procedure that combines data integration, construction of comparable active-like windows, local benchmarking, variation classification, stop-overlap screening, and data-gap diagnosis while keeping the conclusions consistent with the available evidence.

The case application further shows why this procedural contribution matters. The mean relative benchmark gap and accumulated recorded energy above the benchmark produce different resource priorities: the former reflects the proportional magnitude of deviation within flagged windows, whereas the latter reflects the aggregate recorded energy difference across those windows. The stop-overlap analysis then separates windows with a recorded candidate context from windows that remain unexplained by the available stop data. Benchmarking is thus extended from a descriptive ranking exercise into an investigative sequence that indicates where to look, which resources or periods merit attention, and which additional variables are needed next. This supports industrial decision-making without presenting association as causation or energy variation as measured energy efficiency.

4.2 Practical Implications for Industrial Energy Analysis

The framework supports resource prioritisation from more than one perspective. Robotcell should be examined because it has the largest mean relative benchmark gap. Borrmaskin should be prioritised because it has the largest accumulated recorded energy above its local benchmark. Kuggfrasmaskin combines a medium relative gap with a comparatively large accumulated difference. Gradmaskin's larger flag count is not treated as evidence of more frequent variation because the resource-specific upper-decile rule largely determines that count.

Second, the stop-related results provide concrete directions for shop-floor investigation. Waiting-related categories appeared several times in high benchmark-gap windows. This suggests that waiting for personnel, waiting for operators, and uncoded waiting should be checked more carefully. For Borrmaskin, equipment-fault-related candidates should also be reviewed. These results do not directly provide final root causes, but they help narrow down where engineers should start looking.

The results also guide data-collection priorities. Moving from screening to stronger causal confirmation requires energy data to be aligned with cycle time, output quantity, product type, shift information, machine-state records, and more detailed stop coding. May et al. emphasise the need to connect manufacturing energy indicators with production and management practices [21], while Bandaru et al. demonstrate the value of integrating energy and operational data in discrete manufacturing analysis [24].

Fourth, the results show that non-productive or idle-related behaviour should not be ignored. Previous research has shown that operational methods, such as controlling underutilised or idle equipment, can reduce manufacturing equipment energy consumption [22]. In the present case, waiting-related stop candidates and no-stop high-gap windows both suggest that energy use outside clear productive activity deserves attention. For industrial users, this means that improvement work should not only focus on machining energy during production, but also on how energy is maintained during waiting, standby, interruptions, and auxiliary operation.

Overall, the practical value of the framework is that it turns raw energy and stop data into a structured decision-support result. It helps identify which resources should be prioritised, what type of benchmark-gap pattern each resource shows, which stop-related categories are visible, and what additional data would improve the next analysis step.

4.3 Interpretation Boundaries and Case Study Limitations

The findings should be interpreted within the boundaries of the available case data. First, the energy data were analysed at a 30-minute window level. This resolution is useful for observing resource-level energy variation, but it may smooth short state changes, short stops, or short peaks. As a result, some short-duration events may not be clearly separated in the analysis.

Second, the analysis mainly used energy records and stop-related records. Since output quantity, product type, detailed cycle time, and full machine-state labels were not connected at the same window level, the high benchmark-gap windows should be interpreted as deviations from an internal energy reference. They should not be directly treated as output-normalised performance problems. This is an important boundary of the interpretation.

Third, the comparable-window selection was based on an active-like proxy derived from the energy distribution. This was a practical solution for the available data, but it is still a proxy rather than a direct machine-state label from the control or production system. If producing, waiting, idle, setup, and fault states were available at the same time resolution, the comparable-window definition could be improved.

The explanatory value of stop-related screening depends on coding quality and coverage. Uncoded waiting and uncoded equipment-fault categories indicate where to investigate but are too broad to constitute confirmed causes. More detailed stop coding and validation against maintenance and production records would strengthen the operational interpretation.

Finally, the case study covers one month of data from one production line. The results are therefore most suitable for interpreting the energy-variation patterns and resource priorities in this specific case. A longer time period and additional production lines would be needed to test how stable and transferable the findings are.

These limitations define the boundary between screening and causal confirmation. The framework provides a structured first step for identifying high benchmark-gap windows and prioritising investigation. Moving beyond that step requires richer production-context data and more detailed operational records.

4.4 Summary of Discussion

The discussion shows that the proposed framework is most useful as a structured screening and prioritisation approach. Local benchmarking helps identify resources and windows that deviate from internal reference behaviour, instead of simply ranking machines by absolute energy consumption. This makes the results more meaningful for a production line where resources have different functions and normal energy levels.

The resource-level interpretation implies different follow-up actions. Robotcell should be examined for its large relative deviation. Borrmaskin should be prioritised because of its accumulated energy difference and visible stop-related candidates. Kuggfrasmaskin requires additional production-context information because many flagged windows had no overlapping stop record.

Stop-related data provided useful operational evidence, especially for waiting-related and equipment-fault-related candidates. However, the analysis also showed that stop records alone cannot explain all high benchmark-gap windows. This supports the need to combine energy data with richer production information, such as machine state, cycle time, output quantity, product type, and shift information.

Overall, the benchmarking-based route is the primary contribution of the study. It supports transparent shop-floor screening, resource prioritisation, and data-gap diagnosis under the available industrial conditions. It is not intended to establish energy inefficiency or a final causal diagnosis, but to provide a defensible basis for targeted follow-up and improved future data collection.

5 Conclusion

This study developed and applied a benchmarking-based framework for screening within-resource energy-performance variation under limited industrial data conditions. Three operational datasets provided by Volvo Trucks were integrated into a common 30-minute resource–time-window structure. Comparable active-like windows were then selected, local benchmarks were constructed, and high benchmark-gap windows were screened against the available stop records.

The case application shows that the selected resources have different variation profiles. Robotcell has the largest mean relative benchmark gap, whereas Borrmaskin has the largest accumulated recorded energy above its local benchmark. Kuggfrasmaskin combines a medium relative gap with a comparatively large accumulated difference. These quantities support prioritisation, but they do not demonstrate energy inefficiency or estimate achievable savings.

Stop-related categories provide candidate operational context for a subset of flagged windows. Waiting-related and equipment-fault categories are visible for selected resources, while many high-gap windows have no overlapping stop record. The available stop data therefore support targeted follow-up but are insufficient for general causal conclusions.

The contribution can be stated explicitly in relation to the three themes identified in the literature review. For energy-performance assessment frameworks, the four steps translate boundary definition, indicator selection, reference construction, and interpretation into an applied sequence. For assessment and benchmarking methods, the study demonstrates how resource-specific local references and complementary relative and accumulated gap measures can be used to prioritise different forms of performance variation. For data requirements and influencing factors, it treats incomplete context as both a method-selection constraint and an analytical result: stop overlap provides candidate context, while unexplained high-gap windows reveal the variables required for stronger follow-up.

The study therefore addresses a procedural research gap in energy-performance assessment: the lack of an auditable route for moving from available interval energy and stop-event data to defensible shop-floor screening when aligned output and machine-state data are incomplete. Its methodological contribution is a bounded four-step procedure rather than a new universal benchmark, predictive baseline, or efficiency metric. Its empirical contribution is the demonstration that relative and accumulated benchmark gaps can lead to different resource priorities and that the available stop records explain only part of the flagged variation.

Stronger assessment will require aligned data on output, good parts, product type, cycle time, shift, and machine state. Until such data are available, the framework should be used for screening, prioritisation, and data-acquisition planning rather than as a complete energy-efficiency assessment or root-cause diagnosis.

6 Future Work

Future work should first strengthen the production context available for each energy window. Output quantity, good-part counts, cycle time, product type, shift, and machine-state labels should be recorded at the same or a compatible temporal resolution as the energy data. These variables would make it possible to distinguish variation associated with production load and operating state from variation that warrants operational investigation. They would also support output-normalised EnPIs, which are not justified by the present dataset.

The construction of comparable windows should then be validated against actual machine states. The active-like proxy used in this study is a practical response to missing state labels, but it cannot distinguish producing, setup, waiting, idle, and fault states with engineering certainty. Direct state data would reduce the risk of comparing unlike operating conditions and improve the interpretation of the local benchmark.

The robustness of the screening rules should also be tested. Future applications should compare alternative good-performance shares, such as 10%, 20%, and 30%, and alternative high-gap thresholds, such as the 85th, 90th, and 95th percentiles. Stability in resource ranking and variation profiles across these choices would provide stronger evidence that the results do not depend on one arbitrary parameter setting. Temporal concentration should be evaluated over longer periods so that recurring shift, weekday, seasonal, and maintenance-related patterns can be distinguished from isolated events.

The possible explanations from the stop screen require independent validation. Stop logs should be combined with maintenance records, operator notes, production schedules, and engineering review. Uncoded waiting and equipment-fault categories should be refined before they are used for operational decisions. Applying the framework to additional months, lines, and factories would also test the stability and transferability of the local benchmarking procedure.

A predictive baseline model should be considered only after the relevant production variables are available and model validity has been demonstrated. At that stage, model-based deviations could complement the local benchmark rather than replace it. The immediate priority is richer, better-aligned data and a validated comparison structure, not a more complex algorithm. This follows the broader manufacturing energy-management literature, which emphasises integration of energy indicators with production and operational context [21, 24].

Bibliography

[1] International Energy Agency (2026) *Industry – Energy System*. Paris: International Energy Agency. Available at: https://www.iea.org/energy-system/industry

[2] International Organization for Standardization (2018) *ISO 50001:2018 Energy management systems – Requirements with guidance for use*. Geneva: ISO. Available at: https://www.iso.org/standard/69426.html

[3] International Organization for Standardization (2023) *ISO 50006:2023 Energy management systems – Evaluating energy performance using energy performance indicators and energy baselines*. Geneva: ISO. Available at: https://www.iso.org/standard/79367.html

[4] Trianni, A. and Cagno, E. (2017) 'Dealing with barriers to energy efficiency and SMEs: Some empirical evidences', *Energy Efficiency*, 10(2), pp. 1–20. doi:10.1007/s12053-016-9450-y.

[5] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., et al. (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, 372, n71. doi:10.1136/bmj.n71.

[6] Makhadmeh, Z. *et al.* (2020) 'Hierarchical aggregation of energy performance indicators in continuous production processes', *Applied Energy*, 264, p. 114709. doi:10.1016/j.apenergy.2020.114709.

[7] Thollander, P. *et al.* (2018) 'Benchmarking energy performance of industrial small and medium-sized enterprises using an energy efficiency index'.

[8] Thiede, S. (2012) *Energy Efficiency in Manufacturing Systems*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-25914-2.

[9] Duflou, J. R., Sutherland, J. W., Dornfeld, D., Herrmann, C., Jeswiet, J., Kara, S., Hauschild, M., and Kellens, K. (2012) Towards energy and resource efficient manufacturing: A processes and systems approach. *CIRP Annals*, 61(2), 587–609. doi:10.1016/j.cirp.2012.05.002.

[10] Andrijevskaja, J., and Volkova, A. (2025) Industrial energy use, efficiency, and savings: Methods for quantitative analysis. *Energy Efficiency*, 18(7), 76. doi:10.1007/s12053-025-10367-5.

[11] Patterson, M. G. (1996) What is energy efficiency? Concepts, indicators and methodological issues. *Energy Policy*, 24(5), 377–390. doi:10.1016/0301-4215(96)00017-1.

[12] Giacone, E., and Mancò, S. (2012) Energy efficiency measurement in industrial

processes. *Energy*, 38(1), 331–345. doi:10.1016/j.energy.2011.11.054.

[13] Trianni, A., and Cagno, E. (2012) Dealing with barriers to energy efficiency and SMEs: Some empirical evidences. *Energy*, 37(1), 494–504. doi:10.1016/j.energy.2011.11.005.

[14] Gopalakrishnan, B., Ramamoorthy, K., Crowe, E., Chaudhari, S., and Latif, H. (2014) A structured approach for facilitating the implementation of ISO 50001 standard in the manufacturing sector. *Sustainable Energy Technologies and Assessments*, 7, 154–165. doi:10.1016/j.seta.2014.04.006.

[15] Kanneganti, H., Gopalakrishnan, B., Crowe, E., Al-Shebeeb, O., Yelamanchi, T., Nimbarte, A., Currie, K., and Abolhassani, A. (2017) Specification of energy assessment methodologies to satisfy ISO 50001 energy management standard. *Sustainable Energy Technologies and Assessments*, 23, 121–135. doi:10.1016/j.seta.2017.09.003.

[16] Beisheim, B., Krämer, S., and Engell, S. (2020) Hierarchical aggregation of energy performance indicators in continuous production processes. *Applied Energy*, 264, 114709. doi:10.1016/j.apenergy.2020.114709.

[17] Couder, J., and Verbruggen, A. (2003) Technical efficiency measures as a tool for energy benchmarking in industry? *Energy & Environment*, 14(5), 705–724. doi:10.1260/095830503322663410.

[18] Andersson, E., Arfwidsson, O., and Thollander, P. (2018) Benchmarking energy performance of industrial small and medium-sized enterprises using an energy efficiency index: Results based on an energy audit policy program. *Journal of Cleaner Production*, 182, 883–895. doi:10.1016/j.jclepro.2018.02.027.

[19] Kim, S., and Joo, Y. (2025) Energy-based key performance indicator for energy-intensive manufacturing processes: Application to steel casting. *Energy*, 317, 134543. doi:10.1016/j.energy.2025.134543.

[20] Mengistu, A. T., and Panizzolo, R. (2023) Analysis of indicators used for measuring industrial sustainability: A systematic review. *Environment, Development and Sustainability*, 25(3), 1979–2005. doi:10.1007/s10668-021-02053-0.

[21] May, G., Stahl, B., Taisch, M., and Kiritsis, D. (2017) Energy management in manufacturing: From literature review to a conceptual framework. *Journal of Cleaner Production*, 167, 1464–1489. doi:10.1016/j.jclepro.2016.10.191.

[22] Mouzon, G., Yildirim, M. B., and Twomey, J. (2007) Operational methods for minimization of energy consumption of manufacturing equipment. *International Journal of Production Research*, 45(18–19), 4247–4271. doi:10.1080/00207540701450013.

[23] Yao, D., Zhang, X., Wang, K., Zou, T., Wang, D., and Qian, X. (2016) An energy efficiency evaluation method based on energy baseline for chemical industry. *Mathematical Problems in Engineering*, 2016, 7087393. doi:10.1155/2016/7087393.

[24] Bandaru, S., Barrera Diaz, C. A., Ng, A. H. C., and Hanson, L. (2026) Identifying energy bottlenecks in manufacturing systems through an integrated dashboard. *IOP Conference Series: Materials Science and Engineering*, 1342(1), 012060. doi:10.1088/1757-899X/1342/1/012060.

[25] Tranfield, D., Denyer, D., and Smart, P. (2003) Towards a methodology for developing evidence-informed management knowledge by means of system-

atic review. *British Journal of Management*, 14(3), 207–222. doi:10.1111/1467-8551.00375.

A

Appendix: Supplementary figures for benchmarking and root-cause screening

This appendix provides supplementary figures used to support the benchmarking and root-cause-oriented analysis in the case study.

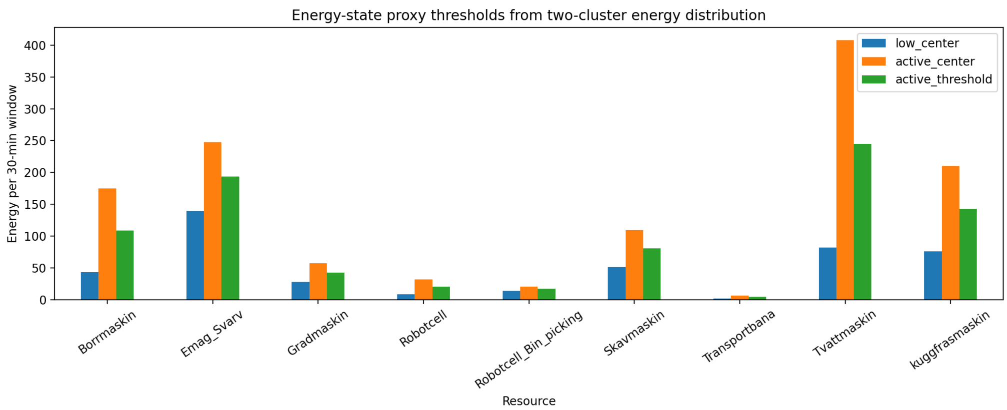

Figure A.1: Energy-state proxy thresholds derived from a two-cluster energy distribution

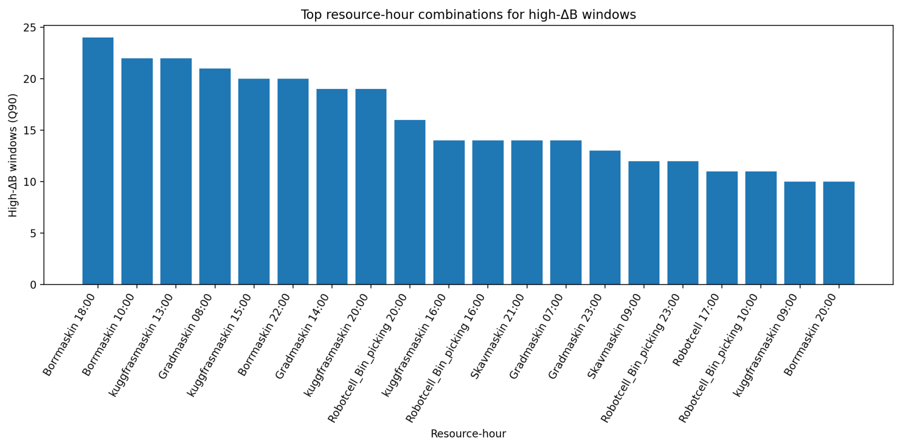

Figure A.2: Top resource-hour combinations for high-ΔB windows

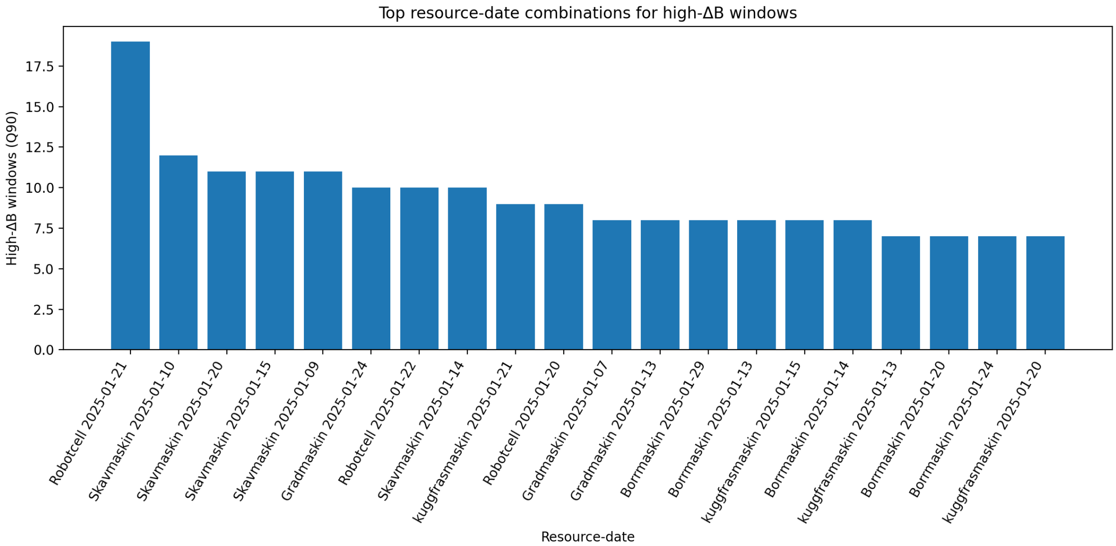

Figure A.3: Top resource-date combinations for high-ΔB windows

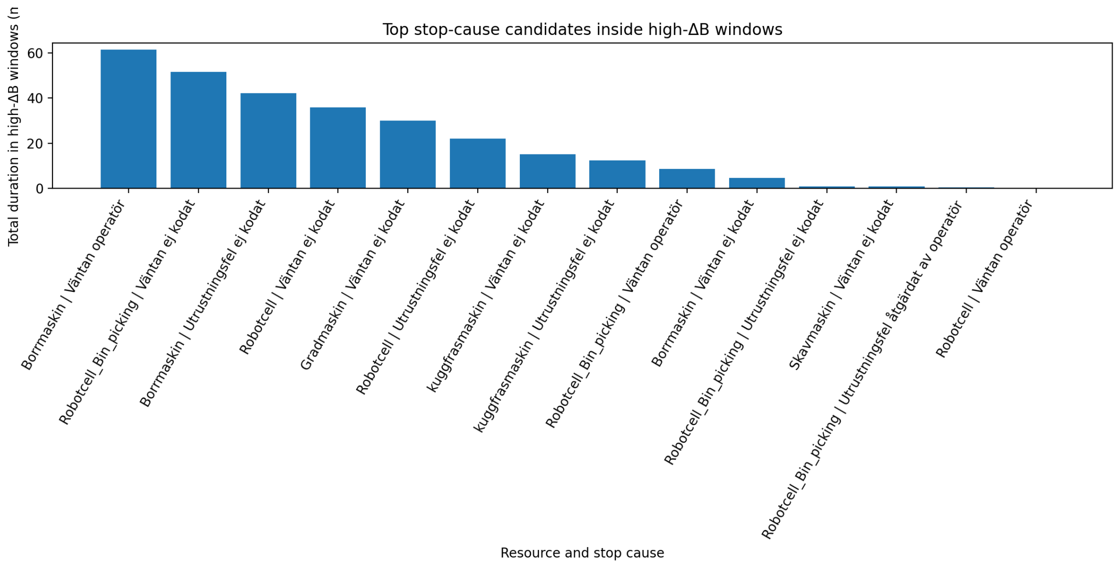

Figure A.4: Top stop-cause candidates inside high-ΔB windows

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY