# Towards Systematic Reliability Assessment: A Multi-Criteria Decision Framework for Modeling Heat Pump Systems

Ahmed Qarqour[1], Sahil-Jai Arora[2], Gernot Heisenberg[3]

[1,2]*Bosch Thermotechnik GmbH, 73243 Wernau (Neckar), Germany*

[1,3]*Institute of Information Science, Technical University of Applied Sciences Cologne, 50678 Cologne, Germany*

*Ahmed.qarqour@de.bosch.com*

## Abstract

Reliability assessment is critical to ensure the performance, availability, and safety of heat pump systems. This requires modeling strategies that reflect both component-level behavior and system-wide interactions. While physics-based, data-driven, and hybrid methods each offer unique strengths, selecting the right approach remains unsolved. This is especially evident in modern heat pump systems with tightly coupled components, fragmented supply chains, and heterogeneous levels of physical insight. Although IoT adoption enables operational data collection, such data remains often unstructured and lacks failure labeling, which limits its value for modeling. These challenges highlight the need for structured guidance in selecting suitable reliability strategies. To address this, a structured and scalable decision framework is proposed to support transparent, context-aware reliability modeling. The approach begins with system-level risk prioritization and applies five Key Decision Indicators (KDIs) to assign appropriate modeling or estimation strategies for each component. This includes both in-depth modeling for risk components and simplified estimation for passive ones. Applied to a real-world air-to-water heat pump system, the framework enables traceable modeling decisions adapted to data availability, physical knowledge, forecasting needs, and cost-efficiency. It offers practitioners a systematic pathway to tailor reliability modeling across complex systems and constrained development environments.

## 1. Introduction

Ensuring stable and reliable operation is essential for the success of heat pump systems in everyday use (International Energy Agency, 2022). These systems are expected to perform reliably year-round, avoiding failures that cause discomfort, costly repairs, or downtime. Long-term functionality supports user satisfaction, operational safety, and high energy efficiency across the system's lifetime (Brudermueller et al., 2025). Reliability is also key to achieving broader goals – lowering $CO_2$ emissions, advancing the energy transition, and enabling climate-neutral heating (IEA, 2022). However, realizing these goals at scale is challenged by limited resources and the wide variety of system configurations, which demand increasingly cost-conscious and selective approaches to reliability assessment.

Reliability assessment is the systematic process of evaluating how well a system performs its intended function over time without failure (O'Connor & Kleyner, 2012). In engineering domains such as heat pumps, it supports performance prediction, maintenance planning, and long-term operational safety (IEA, 2025). Various modeling strategies are used for reliability assessment, typically categorized as physical, data-based, or hybrid approaches. Physical models are based on system equations and known failure mechanisms, offering transparency when detailed physical knowledge is available. Data-based models rely on historical records, such as maintenance logs or performance reports, together with sensor measurements of operational variables. These approaches often do not require deep insight into the system's internal structure. Hybrid approaches combine physical and data-based elements, allowing for flexibility in complex systems with partial knowledge (Lei et al., 2018).

Despite their individual strengths, physical, data-based, and hybrid modeling approaches face notable limitations in complex engineering systems. Physical models struggle with scalability when interactions, control layers, or failure mechanisms become difficult to fully represent – particularly in large systems where detailed physics for all components is impractical (Khan et al., 2024). Data-based models rely on high-quality datasets and often lack interpretability, which limits their trustworthiness in safety-critical contexts (Doshi-Velez & Kim, 2017). Hybrid approaches, though increasingly explored by researchers, inherit the challenges of both –

demanding extensive domain knowledge, careful integration design, and thorough validation. Moreover, they lack standardized procedures for systematically combining physical insights with data-driven components, which limits their broader applicability (Lee et al., 2015; Lei et al., 2018).

Selecting an appropriate reliability modeling approach remains a core challenge in the context of complex technical systems (Khan et al., 2024). Heat pumps involve numerous interacting components and dynamic operating conditions (Fischer & Madani, 2017). In parallel, sensors and smart controls enable the collection of large volumes of operational data, often heterogeneous in structure and quality (IEA, 2022). The interplay between data variability and system complexity further complicates the selection of appropriate reliability modeling approaches (Lazarova-Molnar et al., 2017).

Recent research has addressed system complexity through hybrid modeling approaches and integrated reliability assessment frameworks (Lei et al., 2018; Khan et al., 2024). While these efforts enhance modeling capabilities, they often lack structured guidance for selecting appropriate modeling strategies. In most cases, model selection still depends on fixed criteria – such as physical insight or data availability – without a systematic evaluation process (Ma et al., 2024). As detailed in related work (Section 2), existing literature lacks a structured and adaptable framework for guiding the selection of reliability modeling strategies. Current methods offer limited support for aligning modeling choices with specific system characteristics, data constraints, and practical evaluation needs. This gap becomes increasingly critical as engineering systems grow in complexity and data-driven techniques become more prevalent. Simultaneously, practitioners are under increasing pressure to allocate limited resources efficiently across large and heterogeneous system landscapes.

This paper introduces a multi-criteria decision framework that combines technical system understanding, data characteristics, and practitioner needs to guide the structured selection of reliability modeling approaches. It defines a set of evaluation criteria – such as the criticality of system components, availability of physical knowledge and operational data, forecast requirements, and cost efficiency. The framework aims to increase transparency, reduce decision uncertainty, align modeling depth with reliability objectives, and support consistent model selection across diverse applications. Its applicability is demonstrated through a real-world air-to-water heat pump system.

## 2. RELATED WORK TOWARDS RELIABILITY MODELING

Modern heat pump systems pose growing challenges for reliability modeling. Technically, they feature multiple interacting components, advanced control logic, and diverse configurations (Fischer & Madani, 2017). From a data perspective, they generate large volumes of operational data via smart sensors, yet this data is often heterogeneous, unevenly distributed, or context-specific (IEA, 2022). These combined factors complicate the selection of suitable modeling approaches – especially under limited resources and time constraints common in practical settings.

Beyond the heating sector, valuable insights can be drawn from reliability strategies developed for other complex engineered systems. While these domains operate under different constraints, examining their adaptability to heat pump systems can reveal transferable principles. For instance, health management architectures in industrial systems and aerospace support system-level analysis by integrating diagnostic and prognostic models to monitor critical components (Khan et al., 2024). However, these architectures are tailored for high-assurance environments with extensive expert modeling and validation resources, limiting their transferability to more resource-limited sectors like residential heating. Diagnostics-driven prognostics frameworks provide formal procedures for condition assessment by leveraging detailed failure mode libraries and diagnostic tests to infer system health (Lei et al., 2018). Yet, their dependence on comprehensive diagnostic models and domain-specific knowledge constrains their adaptability to systems with less mature diagnostic ecosystems. In the heating sector, recent research on sensor-driven classification logic shows that statistical feature extraction enables state estimation without requiring full physical modeling (Qarqour et al., 2024). However, this approach lacks methodology to inform decisions between alternative modeling strategies based on system complexity or data availability. Across domains, many of these frameworks rely on fixed modeling paradigms and struggle to adapt dynamically to heterogeneous data sources or evolving system complexities.

In response to these challenges, several studies – including Ikwan, Sanders, and Haddad (2020) – have explored multi-criteria decision-making (MCDM) approaches to prioritize system components or failure modes, helping reduce complexity and support structured resource allocation. Common MCDM approaches, such as the Analytic Hierarchy Process (AHP) (Saaty, Vargas, & St, 2022) and the Preference Ranking Organization METHod for Enrichment Evaluations (PROMETHEE) (Brans & Vincke, 1985), focus on ranking technical importance and risk, but lack recommending modeling strategies tailored to these rankings. For example, AHP quantifies expert judgments to generate hierarchical importance scores but does not extend to guiding model selection, due to limitations such as ranking inconsistencies, rank reversals, and its emphasis on preference rather than prescriptive modeling (Khan & Ali, 2020). Arora and Rabe (2023) applied MCDM to assess predictive maintenance readiness in residential heating systems, integrating user requirements and technical readiness factors to rank component priorities. While their approach captures relevant dimensions, it does not link prioritization to the choice or adaptation of specific reliability modeling approaches.

Therefore, existing MCDM approaches support prioritization but provide limited guidance for matching reliability modeling strategies to component characteristics, available insights, or practitioner constraints. To address this gap, this paper introduces a structured decision-making framework that explicitly links component prioritization with the selection of appropriate reliability modeling strategies – while accounting for system-specific constraints, data availability, and practical requirements.

## 3. FRAMEWORK FOR SELECTING RELIABILITY MODELING APPROACHES

To support structured reliability modeling in heat pump systems, a systematic approach is required to translate domain complexity and diverse system conditions into justifiable modeling decisions. The core idea of the proposed framework is to establish and apply a set of key decision indicators (KDI) that transparently and consistently guide this process. Although heat pump configurations differ in their heat source or circuit design (e.g., air-to-water, air-to-air, or ground-source), the underlying reliability mechanisms and decision needs remain comparable. The core KDIs – such as forecast requirement, data availability, and cost-efficiency – capture generic decision-making dimensions and are therefore system-agnostic within the heat-pump domain. This consistency allows the same framework logic to be applied across different heat-pump types without modification of the indicator definitions. This also demonstrates the framework's potential applicability to other industrial appliances. However, system complexity remains the key distinguishing factor. To ensure practical consistency in applying the framework, the assessment of each component followed a predefined KDI-based logic to ensure traceable outcomes. Expert inputs were aligned through brief calibration discussions, providing a consistent interpretation basis without disclosing proprietary scoring details.

Without such structured criteria, modeling choices risk being misaligned with practical needs. They may overfit theoretical assumptions, overlook key system constraints, or fail to deliver actionable insights. For example, selecting a data-driven approach without addressing forecasting needs may miss essential requirements for anticipating failures. Likewise, relying solely on available measurements while ignoring physical knowledge can hinder diagnostic transparency – often a critical factor in engineering practice (Muhammad et al., 2025). The proposed indicators mitigate these risks by grounding modeling decisions in both the operational context and application goals – rather than defaulting to assumptions, preferences, or convenience.

The proposed framework consists of three main steps – (1) risk-based prioritization of components at the system level, (2) structured selection of suitable reliability modeling approaches at the component level, and (3) a reassessment mechanism that enables trigger-based updates of modeling decisions under evolving system conditions. Throughout this paper, KDI refers to the general concept of Key Decision Indicators, whereas specific indicators (*KDI 1* – *KDI 5*) are italicized to distinguish them in the text. Figure 1 illustrates the three-step decision flow of the proposed framework, which combines system-level prioritization, structured model-selection logic, and a reassessment mechanism to maintain valid recommendations over time.

The first step identifies the components with the highest contribution to overall system reliability using *KDI 0* – Risk Prioritization. These risk components are then selected for in-depth modeling.

For each risk component, a suitable modeling strategy is selected using four indicators (*KDI 1* to *KDI 4*):

- Forecast Requirement (*KDI 1*): Captures the specific output expected from the reliability model based on practical application needs.
- Physical Knowledge (*KDI 2*): Assesses the availability of domain or physics-based understanding relevant to the component.

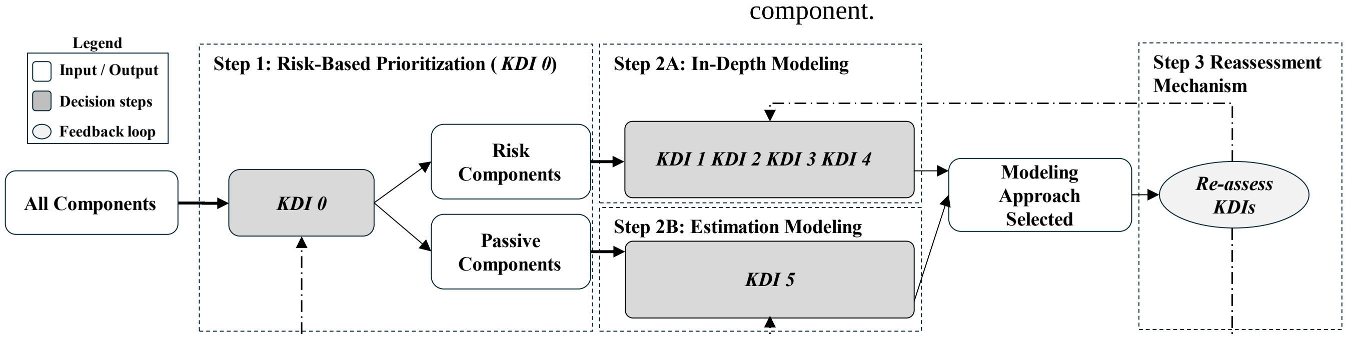


**Figure 1.** Three-step framework with integrated feedback. Step 1 (*KDI 0*) prioritizes components by reliability risk. Step 2A (*KDI 1 – 4*) selects suitable modeling strategies for prioritized components, and Step 2B (*KDI 5*) assigns estimation strategies for the remaining ones. Step 3 enables trigger-based reassessment to keep modeling decisions valid under changing system conditions.

- Data Availability (*KDI 3*): Evaluates whether sufficient operational or failure-related data exist to support the modeling task.
- Cost-Efficiency (*KDI 4*): Prioritizes options that offer an effective balance between model performance and implementation effort.

Once the risk components have been addressed, the framework applies Feasibility of Estimation (*KDI 5*) to all remaining passive components. This indicator evaluates whether a meaningful reliability estimation can be derived based on existing test data, expert knowledge, or operational experience. It ensures that all system elements are systematically considered – even those that do not justify detailed modeling – thereby reducing uncertainty and supporting a more complete system-level reliability assessment. After deployment, Step 3 – the reassessment mechanism – enables trigger-based review of the KDIs to ensure that modeling recommendations remain valid under changing system conditions.

In the following subsections, each KDI is introduced and its role demonstrated through a structured assessment applied to an air-to-water heat pump system manufactured by Bosch Home Comfort Group. The application was carried out in collaboration with domain experts to ensure practical relevance. However, detailed justifications cannot be disclosed due to confidentiality and competitive constraints. The purpose of this contribution is to present a decision framework for selecting reliability modeling approaches. This is illustrated by showing how each KDI contributes to the overall decision process and how the indicators are logically connected to support consistent and justifiable modeling choices.

### 3.1. Risk-Based Prioritization (Step 1: *KDI 0*)

The first step of the framework applies *KDI 0* to identify components with the highest contribution to overall system reliability risk. This prioritization focuses in-depth modeling where it offers the greatest value, reducing complexity and avoiding unnecessary analysis of low-impact parts. As a result, *KDI 1* to *KDI 4* are applied only to a manageable subset of prioritized components, while the remaining ones are addressed more efficiently in later stages.

*KDI 0* evaluates each component based on three key inputs. Two of them are derived from field data: the observed field failure rate ($\lambda$) and the associated failure cost ($C_f$). These values are normalized and combined into a single risk score ($R_s$) that quantifies each component's potential impact on overall system reliability. To identify components requiring in-depth modeling, a threshold is applied to the $R_s$. This threshold is determined using a pareto-based approach (O'Connor & Kleyner, 2012), which selects the top-ranking components that collectively represent the majority of the cumulative system risk (Step 1 in Figure 1). Components exceeding this threshold are classified as risk; those below it are classified as passive.

In addition to the quantitative $R_s$-based selection, the third key input serve as corrective input and is included to ensure operational relevance. The Early Degradation Alert ($\alpha_d$) promotes components to the risk group if they were flagged early in the system's lifecycle due to confirmed degradation or urgent service interventions – even if they fall below the $R_s$. threshold.

Components that either exceed the $R_s$ threshold or are marked with an $\alpha_d$ flag proceed to in-depth modeling approach selection using *KDI 1* to *KDI 4*. All remaining passive components are addressed in later stages using *KDI 5*.

#### 3.1.1. *KDI 0* in Practice

In this contribution, the air-to-water heat pump system use case assesses 28 components – each directly influencing overall system reliability – using the logic of *KDI 0*.

The procedure begins by normalizing $\lambda$ and $C_f$ for each component. These inputs are then combined into a single $R_s$ using equal weighting, as shown in Eq. (1). This weighting reflects expert judgment and ensures a balanced emphasis on both likelihood of failure and potential economic impact.

$$R_{s,i} = 0.5\ \frac{\lambda_i}{max\ (\lambda)} + 0.5\ \frac{C_{f,i}}{max\ (C_f)} \qquad (1)$$

To classify components, they are ranked in descending order based on their $R_s$. A pareto-inspired threshold is applied, selecting the top 20% of components with the highest individual $R_s$ values. Rather than targeting a fixed cumulative percentage, this approach served primarily as a practical dimensionality-reduction heuristic to narrow the modeling focus. Independently of their $R_s$ rank, components marked with an $\alpha_d$ are flagged for risk relevance.

As shown in Figure 1, the outcome of Step 1 is the classification of components into two groups. Any component that either exceeds the $R_s$ threshold or are flagged with $\alpha_d$ are classified as risk and proceed to the next modeling stage *KDI 1* to *KDI 4*. All other components are classified as passive and are later addressed using *KDI 5*.

#### 3.1.2. Interpreting *KDI 0* Results

The application of *KDI* 0 to the air-to-water heat pump system resulted in a clear classification of components into two groups. Out of 28 assessed components, six were identified as risk – either by exceeding the $R_s$ threshold or being flagged with an $\alpha_d$. The remaining 22 components were classified as passive.

The prioritization outcome was reviewed and validated with domain experts, who confirmed that the selected components reflect key service drivers and critical operational risks observed in the field. While this risk-based reduction enables focused modeling and supports efficient resource allocation, it also implies that components with lower immediate risk are treated with simplified methods. This trade-off is accepted to preserve scalability and practical applicability but may introduce residual uncertainty that must be acknowledged in system-level interpretations.

With this subset of risk-relevant components defined, the framework proceeds to determine suitable reliability modeling strategies at the component level, as described in the following sections.

### 3.2. In-Depth Modeling Decision Logic (STEP 2A: *KDI 1* to *KDI 4*)

Following the system-level prioritization in Section 3.1, six components were classified as risk component and selected for in-depth reliability modeling (Step 1 in Figure 1). To determine a suitable modeling approach for each, the framework applies four structured decision indicators – *KDI 1* to *KDI 4* – which assess forecast requirements, physical knowledge, data availability, and cost-efficiency (Step 2A in Figure 1). Together, these indicators provide a systematic basis for selecting between physics-based, data-driven, or hybrid modeling strategies.

Rather than relying on a single criterion, the decision logic integrates all four indicators to ensure a context-sensitive modeling choice. Figure 2 illustrates this logic, showing how the assessment begins with evaluating practitioner needs (*KDI 1*) and component characteristics – specifically domain knowledge (*KDI 2*) and operational data availability (*KDI 3*). If these three indicators align in pointing to the same modeling type, the corresponding strategy is adopted. In cases of divergence, *KDI 4* is used to resolve trade-offs by considering modeling performance and implementation effort.

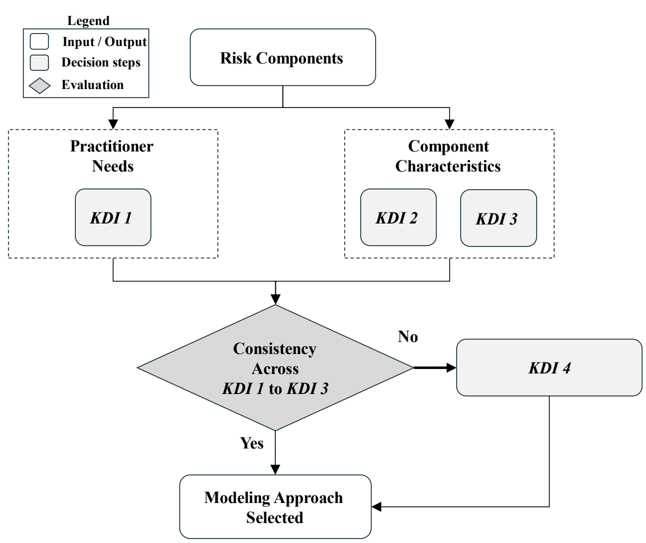


Figure 2. Decision logic for selecting a modeling approach for risk components. Practitioner needs (*KDI 1*) and component characteristics – domain knowledge (*KDI 2*) and data availability (*KDI 3*) – are first assessed. If these align, the corresponding modeling type is chosen; otherwise, *KDI 4* resolves trade-offs.

The following subsections (3.2.1 to 3.2.5) present each indicator in detail, outlining its purpose, role in the decision process, and how it was applied in the current use case.

#### 3.2.1. Forecast Requirements (*KDI 1*)

*KDI 1* determines the forecast depth required for each risk-relevant component, ensuring that the selected modeling strategy delivers outputs aligned with practitioner needs. Domain experts were consulted using a predefined and standardized assessment logic. The three assessment questions ($Q_1$ to $Q_4$) and their mapping to forecast levels are shown in Table 1. In the table, $Q_i$ denotes that Question $i$ is answered "Yes" (True), and $\neg Q_i$ denotes that it is answered "No" (False).

In the current use case, all six risk-relevant components showed different levels of forecast depth required by practitioners. These requirements were derived from expert discussions considering factors such as maintenance planning needs, the criticality of early failure detection, and compliance with safety or regulatory standards. Based on the defined mapping logic, each component was classified into a high, medium, or low forecast depth and then linked to a suitable modeling strategy. Components with high forecast requirements – such as those demanding predictive alerts, behavioral insights, and a physical understanding of failure mechanisms – were matched to physics-based models. Medium-level components required predictive alerts and behavioral insights but did not demand a physical understanding of failure causes, while low-level components required only predictive alerts for maintenance planning.

While *KDI 1* defines the forecast requirements, it does not assess whether sufficient physical understanding (*KDI 2*) or adequate data availability (*KDI 3*) exist to enable the required modeling strategy. The following subsections address these two aspects in sequence, beginning with *KDI 2*.

#### 3.2.2. Physical Knowledge (*KDI 2*)

While *KDI 1* defines the required forecast output based on practitioner needs, *KDI 2* evaluates the level of physical understanding available for each component, ensuring that the selected modeling strategy is grounded in what can realistically be explained rather than only predicted. To operationalize this indicator, domain experts were consulted using a predefined and standardized assessment logic. The four assessment questions ($Q_1$ to $Q_4$) and their mapping to knowledge levels are shown in Table 2.

Table 1. Mapping of assessment questions ($\boldsymbol{Q_1}$ to $\boldsymbol{Q_4}$) to forecast requirement levels and corresponding modeling strategies (*KDI 1*)

| Assessment Questions | Level Mapping (L) | Modeling Implication | | |
|---|---|---|---|---|
| | | Physics-based | Hybrid | Data-driven |
| $Q_1$. Is a predictive alert required (e.g., remaining useful life)?<br>$Q_2$. Are behavioral insights needed (e.g., degradation patterns, trends)?<br>$Q_3$. Is a physical understanding of failure mechanisms essential (e.g., for interpretability or compliance)? | $L_{forcast} = High \Leftrightarrow \bigwedge_{i=1}^{3} Q_i$ | x | | |
| | $L_{forcast} = Medium \Leftrightarrow Q_1 \land Q_2 \land \neg Q_3$ | x | x | |
| | $L_{forcast} = low \Leftrightarrow Q_1 \land \neg Q_2 \land \neg Q_3$ | x | x | x |

In the current use case, the *KDI 2* assessment revealed high diversity of system knowledge among the six risk-relevant components. Based on the defined mapping logic, each component was classified into a high, medium, or low knowledge level and then linked to a suitable modeling strategy. This ensured that components with comprehensive knowledge were matched to physics-based diagnostics, those with partial understanding to hybrid models, and those with limited knowledge to data-driven methods.

While *KDI 2* addresses the availability of physical understanding, it does not account for the availability of operational data required for model development. This aspect is covered by the next indicator, *KDI 3*.

### 3.2.3. Data Availability (*KDI 3*)

While *KDI 2* evaluates the level of physical knowledge for each component, *KDI 3* assesses the extent to which this knowledge can be supported by empirical evidence. Specifically, it determines whether sufficient data are available to enable the development and validation of the selected modeling approach. The assessment follows the same structured logic as *KDI 2* directly integrated into Table 3.

The mapping logic in Table 3 classifies each component into one of three data availability levels. A high level ($L_{data} = High$) is assigned when all four conditions are met (e.g., a component is monitored in the field, covers diverse lifecycle scenarios, includes failure cases, and is validated in laboratory settings). Medium availability ($L_{data} = Medium$ applies when most conditions are fulfilled – specifically $Q_1$, $Q_2$, and $Q_4$ are "Yes" but $Q_3$ is "No" (e.g., a component with broad operational coverage and lab validation, but lacking failure case representation). Low availability is assigned to

Table 2. Mapping of assessment questions ($\boldsymbol{Q_1}$ to $\boldsymbol{Q_4}$)) to physical knowledge levels and corresponding modeling strategies (*KDI 2*)

| Assessment Questions | Level Mapping (L) | Modeling Implication | | |
|---|---|---|---|---|
| | | Physics-based | Hybrid | Data-driven |
| $Q_1$. Is a validated physical model available for the component?<br>$Q_2$. Are the main failure causes clearly understood and documented?<br>$Q_3$. Are key influencing factors (e.g., operational/environmental) identified and documented?<br>$Q_4$. Are design or operating conditions clearly linked to failure modes? | $L_{knowledge} = High \Leftrightarrow \bigwedge_{i=1}^{4} Q_i$ | x | x | x |
| | $L_{knowledge} = Medium \Leftrightarrow Q_2 \land Q_4 \land (\neg Q_1 \lor \neg Q_3)$ | | x | x |
| | $L_{knowledge} = low \Leftrightarrow (\sum_{\boldsymbol{i=1}}^{\boldsymbol{4}} [Q_i] < \boldsymbol{2}) \lor (\neg Q_1 \land \neg Q_4)$ | | | x |

Table 3. Mapping of assessment questions ($\boldsymbol{Q_1}$ to $\boldsymbol{Q_4}$)) to data availability levels and corresponding modeling strategies (*KDI 3*)

<table>
<tr><th rowspan="2">Assessment Questions</th><th rowspan="2">Level Mapping (L)</th><th colspan="3">Modeling Implication</th></tr>
<tr><th>Physics-based</th><th>Hybrid</th><th>Data-driven</th></tr>
<tr><td rowspan="3">$Q_1$. Is the component regularly monitored during real-world operation?<br>$Q_2$. Does the available data cover a range of lifecycle scenarios and usage conditions?<br>$Q_3$. Are relevant failure scenarios captured in the operational dataset?<br>$Q_4$. Are degradation patterns replicated and validated in lab environments?</td><td>$L_{data} = High \Leftrightarrow \bigwedge_{i=1}^{4} Q_i$</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>$L_{data} = Medium \Leftrightarrow Q_1 \wedge Q_2 \wedge Q_4 \wedge \neg Q_3$</td><td>x</td><td>x</td><td></td></tr>
<tr><td>$L_{data} = low \Leftrightarrow$Any other combination of answers not covered by the High or Medium conditions</td><td>x</td><td></td><td></td></tr>
</table>

any other combination of responses, such as sparse field monitoring or incomplete coverage of operational conditions.

In the current use case, the *KDI 3* evaluation revealed a high diversity of data availability across the six risk-prioritized components. Some components qualified for the high category, supporting data-driven, hybrid, or physics-based modeling. Others fell into the medium or low categories, where missing operational failure data or fragmented monitoring limited the suitability of empirical model training.

Together with *KDI 1* and *KDI 2*, this indicator ensures that the final modeling strategy is not only aligned with forecast requirements and grounded in physical understanding, but also feasible given the system's operational observability.

### 3.2.4. Observed Consistency Across *KDI 1* to *KDI 3*

Figure 3 visualizes how the six risk components were assessed based on indicators *KDI 1* to *KDI 3*. Each component is positioned according to its system knowledge (*KDI 2*, x-axis) and data availability (*KDI 3*, y-axis). The required forecast depth (*KDI 1*) is shown by a color scale from light grey (low) to dark grey (high), and unique marker shapes distinguish components. Components requiring further evaluation through *KDI 4* are outlined with a circle. For clarity, the six risk components are labeled C1 to C6 throughout this section and in the figure.

For instance, C1 shows high data availability but low physical knowledge, and a low forecast depth – favoring a data-driven modeling approach. C2 combines medium data availability and physical knowledge with a medium forecast depth, making a hybrid approach suitable. C3 has high physical knowledge but low data availability and a high forecast requirement, favoring a physics-based approach. C6 scores high across all indicators, aligning well with physics-based modeling

Ambiguities occur in cases like C4, where high data availability and medium physical knowledge, combined with a low forecast depth, support both data-driven and hybrid approaches; here, *KDI 4* was applied to resolve the choice.

C5 represents a more constrained case: despite a medium forecast requirement, both system knowledge and data availability were insufficient to support robust modeling approach. This highlights a critical reliability knowledge gap and signals a need for targeted improvements at the system level before modeling can proceed.

Synthesizing the *KDI 1* to *KDI 3* results, four components were directly linked to a single strategy: two followed a

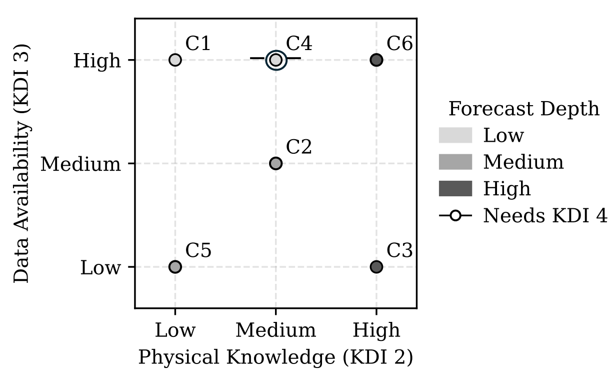


Figure 3. Mapping of six risk components (C1–C6) based on *KDI 1* to *KDI 3*. Axes show system knowledge (*KDI* 2) and data availability (*KDI 3*). Dot shading encodes forecast depth; outlined dots indicate components requiring *KDI 4* evaluation.

physics-based approach, one a data-driven approach, and one a hybrid approach. One component was suitable for both data-driven and hybrid approaches, and one lacked sufficient knowledge and data for any robust modeling. This combined evaluation identified components with aligned forecast requirements and technical feasibility, while also revealing those where modeling remains infeasible or uncertain. The latter warrant prioritization in future reliability research and system development. For components with multiple viable strategies, the final choice is determined through *KDI 4*, described in the next subsection.

#### 3.2.5. Cost-Efficiency (*KDI 4*)

*KDI 4* evaluates the cost-efficiency of implementing a given reliability modeling strategy. It becomes particularly relevant when earlier indicators (*KDI 1* to *KDI 3*) suggest multiple technically feasible options. In such cases, *KDI 4* ensures that the final decision also considers resource constraints and development effort, supporting a balanced trade-off between modeling performance and practical implementability.

In the absence of formal cost metrics, the evaluation of *KDI 4* was guided by expert input and supported by a heuristic commonly observed in industrial practice: modeling effort typically increases from data-driven, through hybrid, to physics-based approaches (Wang et al., 2019). Data-driven models often build on available operational data and existing infrastructure (Zhao et al., 2019), while physics-based models demand substantial domain knowledge, parameter calibration, and validation efforts (Hoang & Kang, 2018). Hybrid models integrate both operational data and physical principles, resulting in a moderate development effort between the generally lower demands of data-driven approaches and the higher demands of physics-based approaches. Accordingly, *KDI 4* ranks the three strategies from lowest to highest typical modeling effort as follows: (1) data-driven, (2) hybrid, (3) physics-based. This ranking served as the operational setup for *KDI 4* in the present analysis. While individual cases may vary, this relative ranking provided a practical and reproducible basis for strategy selection when resource constraints are relevant and technical options are still open.

In the current use case, *KDI 4* was applied only to Component C4. This component had been previously assessed (through *KDI 1* to *KDI 3*) as eligible for both data-driven and hybrid modeling strategies. Applying the cost-efficiency logic, the data-driven approach was preferred, as it satisfied the forecast requirement with minimal development overhead. By integrating this step, the framework ensured that modeling recommendations not only reflect technical validity but are also realistically implementable under real-world constraints.

### 3.3. Feasibility of Estimating Reliability (Step 2B: *KDI* 5)

Following the in-depth modeling decisions for risk-relevant components, the framework addresses the remaining components classified as passive (Step 1 in Figure 1). These components, while not major drivers of overall system reliability risk, must still be considered to ensure a complete reliability assessment at the system level. This is achieved through estimation-based strategies, which offer lower modeling precision but maintain coverage across the full component landscape (Azam et al., 2014).

*KDI 5* guides the selection of appropriate estimation strategies for passive components (Step 2B in Figure 1). In contrast to the earlier indicators that focused on modeling depth and feasibility, *KDI 5* addresses the residual set of components and matches them with a lightweight, context-appropriate estimation method. The selection is informed by three practical system-level criteria: the availability of test infrastructure, expert knowledge from related systems, and historical field data on failure occurrences. The mapping logic for *KDI 5* is summarized in Table 4. It translates the availability of resources into appropriate estimation types.

In the current use case, all 22 passive components were assessed using this logic. Most were selected to simplified data-based estimation, supported by operational data on failure behavior. Components lacking sufficient data but with expert familiarity were selected to expert-based estimation. A smaller subset, for which structured test procedures exist, was mapped to test-based estimation. Unlike in the in-depth modeling strategy selection, these estimation strategies are not ranked in terms of preference; their selection depends solely on the applicable boundary conditions for each passive component.

By applying *KDI 5*, the framework maintains a consistent and traceable reliability representation across all components. While these estimations do not carry the same predictive power as full models, they enable system engineers to maintain awareness of potential failure behavior. Furthermore, components relying on simplified estimates can be flagged for future data collection or modeling

Table 4. Mapping of available resources (*KDI 5*) to suitable estimation strategies for passive components

| Available Resources | Proposed Estimation Strategy |
|---|---|
| Test procedures and measurement equipment | Test-based estimation |
| Expert knowledge of similar components | Expert-based estimation |
| Historical field failure data | Simplified data-based estimation |

enhancement efforts. This ensures that system-level reliability remains quantifiable and improvable, even in the presence of limited resources.

### 3.4. Reassessment Mechanism (Step 3)

The framework includes a reassessment mechanism to keep modeling decisions aligned with evolving system conditions (Step 3 in Figure 1). After deployment, selected approaches are trigger-based reassessed when relevant changes occur in model behavior, operational performance, or available data. In practice, this does not imply continuous recalculation but rather a manual or periodic review when such changes become apparent. If deviations are detected, the corresponding KDIs are revisited: component-level updates feed into Step 2A/2B (*KDI 1 – 5*), while system-level changes can prompt an update of Step 1 (*KDI 0*). This mechanism enables periodic adaptation of modeling strategies under real operating conditions without disclosing proprietary performance details.

## 4. DISCUSSION AND FUTURE WORK

The revised framework demonstrates how complex reliability modeling decisions can be operationalized in a more transparent, structured, and context-aware manner by applying a three-step logic complemented by a reassessment mechanism. The first step – system-level risk-based prioritization (*KDI* 0) – enables engineers to identify components that most critically affect overall system reliability. These risk components are prioritized for in-depth modeling, while components with lower impact – referred to as passive components – are addressed through estimation-based strategies. The second step applies *KDI 1 – 5* to guide the assignment of appropriate modeling strategies for each component. For risk components, the decision logic ensures that selected approaches align with forecasting needs, physical knowledge, data availability, and cost-efficiency. This supports assigning data-driven, physics-based, or hybrid modeling approaches. For passive components, the fifth *KDI* guides the assignment of simplified estimation strategies – based on available failure statistics, component test procedures, or expert judgment. This inclusive design avoids overlooking components that do not warrant detailed modeling but still require reliability estimates, thereby enabling a comprehensive system-level assessment. The third step – reassessment mechanism – enables trigger-based updates of modeling decisions when system conditions change over time. By clearly separating prioritization from modeling logic and maintaining a dedicated reassessment mechanism, the framework supports traceable modeling decisions that reflect practical engineering goals and real-world constraints – without relying on fixed rules or single-method assumptions.

The applicability of the decision framework was further demonstrated by reviewing an earlier case study (Qarqour et al., 2024), in which a Random-Forest-based, data-driven model was used to predict control faults in the inverter of an air-to-water heat pump. Although this study preceded the formalization of the framework, a retrospective KDI-based assessment led to the same reliability-approach selection, confirming the logic of the decision process. Notably, the model successfully identified the inverter – later confirmed as a reliability-critical component in field data – illustrating the framework's applicability while preserving confidentiality regarding proprietary parameters.

Applying the framework to an air-to-water heat pump system revealed several practical insights. First, components with similar criticality scores required different modeling strategies due to variations in forecasting needs, data availability, and physical understanding. This highlights the importance of context-aware decisions beyond risk ranking alone. Second, some components were suitable for more than one modeling approach, allowing cost-efficiency to guide the final selection. This flexibility supports tailoring model complexity to available development resources. Third, in constrained cases, the framework exposed situations where modeling could not proceed due to missing data and limited physical insight. This points to a critical reliability knowledge gap and the need for targeted system improvements before modeling becomes viable. These observations form the basis for reflecting on the framework's current limitations and potential directions for refinement.

### 4.1. Limitations and Methodological Refinement

While the proposed decision framework demonstrated promising results in the pilot use case, several limitations should be acknowledged to guide future refinement. The KDIs, while structured, still rely on expert interpretation, which may vary depending on individual judgment or organizational context. This subjectivity introduces the risk of inconsistent modeling decisions across applications. However, the prioritization methodology remains valid in the present study, as all assessments were conducted within the same company and engineering department, ensuring a consistent reference framework despite potential evaluator bias. To enhance consistency, future studies should formalize expert evaluation through structured calibration sessions or consensus-oriented methods (Okoli & Pawlowski, 2004), such as Delphi-style reviews, and may also integrate quantitative weighting schemes to reduce evaluator bias and make implicit trade-offs – such as cost efficiency – more transparent. Moreover, the current study does not yet include a continuous feedback mechanism to verify whether selected strategies remain valid once implemented. Although the reassessment mechanism (Step 3) has been conceptually integrated, its practical effectiveness remains to be validated through real-world implementation. Further work should examine how the reassessment logic performs under operational conditions and whether it effectively supports iterative improvement of modeling strategies.

### 4.2. Generalization and Future Validation

Beyond the present air-to-water case, the framework's underlying logic is domain-agnostic, as it builds on indicators that describe decision-making dimensions common to most technical systems. KDIs such as forecast requirement, data availability, and cost-efficiency are transferable across industrial domains, whereas physical knowledge may require contextual adaptation. Cooling and battery thermal-management systems serve as illustrative examples, as they involve comparable thermodynamic processes and reliability trade-offs but operate under different boundary conditions. Transferring the framework to these domains would mainly require redefining reference variables and validation criteria while keeping the overall decision logic unchanged.

Building on the framework's findings, future work should apply the selected strategies to prioritized components and assess whether they produce reliable, interpretable results. Building on this retrospective validation, future work should apply the selected reliability-modeling approaches to prioritized components within real heat-pump systems. This next step will serve as a direct validation of the framework's decision logic under operational conditions. This implementation step will not only validate individual modeling choices but also test the framework's integration potential at the system level, for example via Reliability Block Diagrams or similar compositional methods (Hasan et al., 2015).

Expanding the framework's use to other system types, coupled with uncertainty handling through confidence scoring or consensus-based expert review, will further strengthen its robustness and generalizability. Future applications should also incorporate structured consensus or calibration steps to align expert assessments and refine the interpretation of each KDI over time. Together, these steps will support broader adoption and refinement of the framework in reliability-oriented modeling.

## 5. CONCLUSION

Modern heat pump systems include components with different levels of risk, data availability, and system understanding, making reliability modeling increasingly complex. This complexity is compounded by the absence of comprehensive decision-making frameworks, highlighting the need for structured approaches to guide model selection. This work introduced a structured decision framework that supports the transparent selection of reliability modeling strategies across such systems. The framework combines system-level risk prioritization with five KDIs to align modeling decisions with practitioner needs, available physical knowledge and data, and resource constraints. By separating prioritization from modeling logic, it enables traceable, context-aware decisions for both risk and passive components. Applied to an air-to-water heat pump system, the framework demonstrated its ability to differentiate suitable modeling strategies, highlight knowledge gaps, and support system-level reliability assessment. While the approach showed practical value, it remains subject to limitations – including reliance on expert judgment, the absence of a built-in validation mechanism, and an initial focus on a single system type. Future work should operationalize the assigned strategies in real systems, extend the framework to other domains, and incorporate uncertainty handling and consensus-driven validation. Together, these steps will support broader adoption and refinement of the framework in reliability-oriented modeling.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the valuable collaboration and support received throughout this research. Special thanks are extended to the engineering and data analytics teams of Bosch Home Comfort Group for their contributions during the development and validation of the decision framework. This work was conducted as part of a doctoral research project funded through a collaborative program between Bosch Home Comfort Group and the Technical University of Applied Sciences Cologne, with the shared goal of advancing reliability modeling in sustainable heating technologies.

## REFERENCES

Arora, S. J., & Rabe, M. (2023). Predictive maintenance: Assessment of potentials for residential heating systems. *International Journal of Computer Integrated Manufacturing, 36*(12), 1876–1900. https://doi.org/10.1080/0951192X.2023.2204471

Azam, M., Ghoshal, S., Deb, S., Pattipati, K., Haste, D., Mandal, S., & Kleinman, D. (2014). Integrated diagnostics and time to maintenance for complex engineering systems. In *Proceedings of the IEEE Aerospace Conference* (pp. 1–10). IEEE. https://hdl.handle.net/10945/44461

Brans, J. P., & Vincke, P. (1985). A preference ranking organisation method: (The PROMETHEE method for multiple criteria decision-making). *Management Science, 31*(6), 647–656. https://doi.org/10.1287/mnsc.31.6.647

Brudermueller, T., Potthoff, U., Fleisch, E. (2025). Estimation of energy efficiency of heat pumps in residential buildings using real operation data. *Nature Communications, 16*, 2834. https://doi.org/10.1038/s41467-025-58014-y

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint*, arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608

Fischer, D., & Madani, H. (2017). On heat pumps in smart grids: A review. *Renewable and Sustainable Energy Reviews, 70,* 342–357. https://doi.org/10.1016/j.rser.2016.11.182

Hasan, W., Ahmed, S., Tahar, S., & Hamdi, M. S. (2015). Reliability block diagrams based analysis: A survey. *AIP Conference Proceedings, 1648*(1), 850129. https://doi.org/10.1063/1.4913184

Hoang, T., & Kang, H.-J. (2018). A survey on deep learning based bearing fault diagnosis. *Neurocomputing, 335.*, Article 403–420. https://doi.org/10.1016/j.neucom.2018.06.078

International Energy Agency. Heat pumps. Paris, France: Author. Retrieved August 13, 2025, from https://www.iea.org/energy-system/buildings/heat-pumps

International Energy Agency. (2022). *The future of heat pumps*. Paris, France: Author. https://www.iea.org/reports/the-future-of-heat-pumps

Ikwan, F., Sanders, D., & Haddad, M. (2020). A combined AHP–PROMETHEE approach for intelligent risk prediction of leak in a storage tank. *International Journal of Reliability, Risk and Safety: Theory and Application, 3*(2), 55–61. https://doi.org/10.30699/IJRRS.3.2.7

Khan, A. U., & Ali, Y. (2020). Analytical hierarchy process (AHP) and analytic network process methods and their applications: A twenty-year review from 2000–2019. *International Journal of the Analytic Hierarchy Process, 12*(3). https://doi.org/10.13033/ijahp.v12i3.822

Khan, S., Yairi, T., Tsutsumi, S., & Nakasuka, S. (2024). A review of physics-based learning for system health management. *Annual Reviews in Control*, *57*, 100932. https://doi.org/10.1016/j.arcontrol.2024.100932

Lazarova-Molnar, S., Mohamed, N., & Shaker, H. R. (2017). Reliability modeling of cyber-physical systems: A holistic overview and challenges. In *2017 IEEE International Workshop on Measurement and Networking of Cyber-Physical Systems (MSCPES)* (pp. 1–6). IEEE. https://doi.org/10.1109/MSCPES.2017.8064536

Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters, 3*, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing, 104,* 799–836. https://doi.org/10.1016/j.ymssp.2017.11.016

Ma, Z., Jørgensen, B. N., & Ma, Z. G. (2024). A systematic data characteristic understanding framework towards physical-sensor big data challenges. *Journal of Big Data, 11*, 84. https://doi.org/10.1186/s40537-024-00942-5

Muhammad, D., Salman, M., Inan Keles, A., & Bendechache, M. (2025). All diagnosis: Can efficiency and transparency coexist? An explainable deep learning approach. *Scientific Reports, 15*. https://doi.org/10.1038/s41598-025-97297-5

O'Connor, P. D. T., & Kleyner, A. (2012). *Practical reliability engineering* (5th ed.). Chichester, UK: Wiley.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations, and applications. *Information & Management, 42*(1), 15–29. https://doi.org/10.1016/j.im.2003.11.002

Qarqour, A., Arora, S. J., Heisenberg, G., Rabe, M., & Kleinert, T. (2024). Utilizing data analysis for optimized determination of the current operational state of heating systems. In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KDIR)* (pp. 200–209). SciTePress. https://doi.org/10.5220/0012876200003838

Saaty, T., Vargas, L., & St, C. (2022). *The analytic hierarchy process*. Springer.

Wang, Y., Zhao, Y., & Addepalli, P. (2019). Remaining useful life prediction using deep learning approaches: A review. *Procedia Manufacturing, 49,* 67–73. https://doi.org/10.1016/j.promfg.2020.06.015

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050