15TH IEA HEAT PUMP CONFERENCE
MAY 26–29, 2026 | HOFBURG
VIENNA 2026
DECARBONISATION THROUGH INNOVATION

Common failure analysis of domestic heat pumps and air conditioners

T. Oltersdorf[a*], M. Drews[a], L. Schnabel[a], H. Fugmann[a]

[a]Fraunhofer ISE, Heidenhofstr. 2, 79110 Freiburg, Germany

Abstract

The fate and lifetime of products is key to circular economic concepts. Even though circular concepts are planned to be adopted in the close future in several economic areas like the European Economic Area, there is a knowledge gap on the fate of heat pumps (HPs) and air conditioners (ACs) after their commissioning. Only few studies are available in which a relevant database was collected or available to be studied. Either to review the root causes for failures leading either to servicing or the dismantling of a system or to gather information on the estimated useful life (EUL) of HPs or ACs.
As part of the German HP resilience project in this work, a new approach was chosen to gather information and collect data able to fill this knowledge gap. Internet platforms dealing with used products were monitored over years to build a specific database for HPs and ACs. The dataset comprises data on used markets of almost all EU member states. Currently, about several ten thousand ads of used HP and AC products were identified and classified. The database contains still functional but also used HPs and about 10% of totally damaged HPs. This database is exploited for the first time. The first knowledge gain aims for the analysis of selling reasons and root cause analysis for failures.
The dataset, current findings, and its further potential use is presented. Further challenges to deploy this dataset for lifetime analysis is discussed.

Keywords: failure analysis, second-hand market data

1. Introduction

A precise knowledge about the stock of existing heating equipment and for HPs is fundamental to estimate the state of decarbonization and its decarbonization potential . Knowing the estimated useful lifetimes of HPs helps to understand and supports modelling the stock of equipment. Does newly sold equipment substitute existing fossil-fueled boilers or is new equipment just used to replace already existing but worn-out heat pumps? [1] . The used market data are a source of information that was rarely used, and it is worth investigating its potential for contributing to a better understanding of HPs and ACs lifetimes that could be used for such decarbonization potential studies. In this study, the data acquisition focused on heat pumps, air conditioners and pool heat pumps for domestic use.

State-of-the-art estimation of end of useful life (EUL) of equipment is rare. For residential central ACs and HPs there are studies from the U.S. which builds on Lutz et al.'s Weibull-based lifetime modelling of national survey data and on US DOE Technical Support Documents (TSD) for Residential Central ACs and HPs, which operationalize these distributions into typical service lifetimes (e.g. on the order of two decades for central AC and somewhat shorter for HPs) in rulemakings [2,3]. Recent DNV[1] studies, notably Project 73 Track D for Massachusetts and the CPUC Group A EUL work for California, extend this by combining snapshot age data

* Corresponding author. Tel.: +49 761 4588 5239
E-mail address: thore.oltersdorf@ise.fraunhofer.de
[1] DNV: Den Norske Veritas

from on-site market characterization and technician/program surveys with AHRI/DOE shipment series in survival models to derive EULs for both mature AC and emerging HP stocks [4,5]. For saturated markets, DOE-style methods infer retirement patterns from long shipment histories, stock estimates, and Weibull survival fractions, whereas in unsaturated or rapidly growing HP markets, cohort-based approaches are required that reconstruct installation-year profiles and survival functions from observed age distributions [6,7]. Because technician or used-market datasets are inherently biased convenience samples, such analyses must embed them in an explicit survey framework with appropriate case weights, treatment of missing manufacture dates, separation of age-at-failure from age-in-place, and parametric survival modelling so that, when combined with auxiliary installation data, they approximate the properties of an ideal random sample [7,8].

2. Data extraction procedure

Used-market datasets from marketplace are therefore strongly selected, left-truncated and right-censored samples of the underlying equipment stock—items only enter the data if owners decide to sell them and tell in their ads, key variables like installation year, usage intensity and failure cause are frequently missing, and residual lifetimes are only observed for units that are still working—so lifetime models must explicitly incorporate these selection and incompleteness mechanisms, as recent auction- and second-hand-based studies do [6,9,10].

During the 1,5 years of continuous studying of these advertisements and marketplaces we have seen a large spectrum of data quality, and the content provided. This experience allows us to state that a huge share of advertisers / owners do not just state by their own layish knowledge on HVAC equipment about the state of equipment. In most cases they forward the statement as part of a technician's evaluation, cost estimates or even a quotation on the repair or replacement costs, or they state about the error code of the equipment which they offer. This decision process is usually a simple economic process if replacement or repair is economically better. It is difficult to do a final assessment on the quality of data from these used-equipment markets, with the knowledge that much of it relies on contractors' estimates and even quotes handed over to the owners of the offered equipment, compared with checklist-based information gathered during site visits to residential installations as part of large-scale studies i.e. like done by DNV on behalf of Massachusetts or other US federal states.

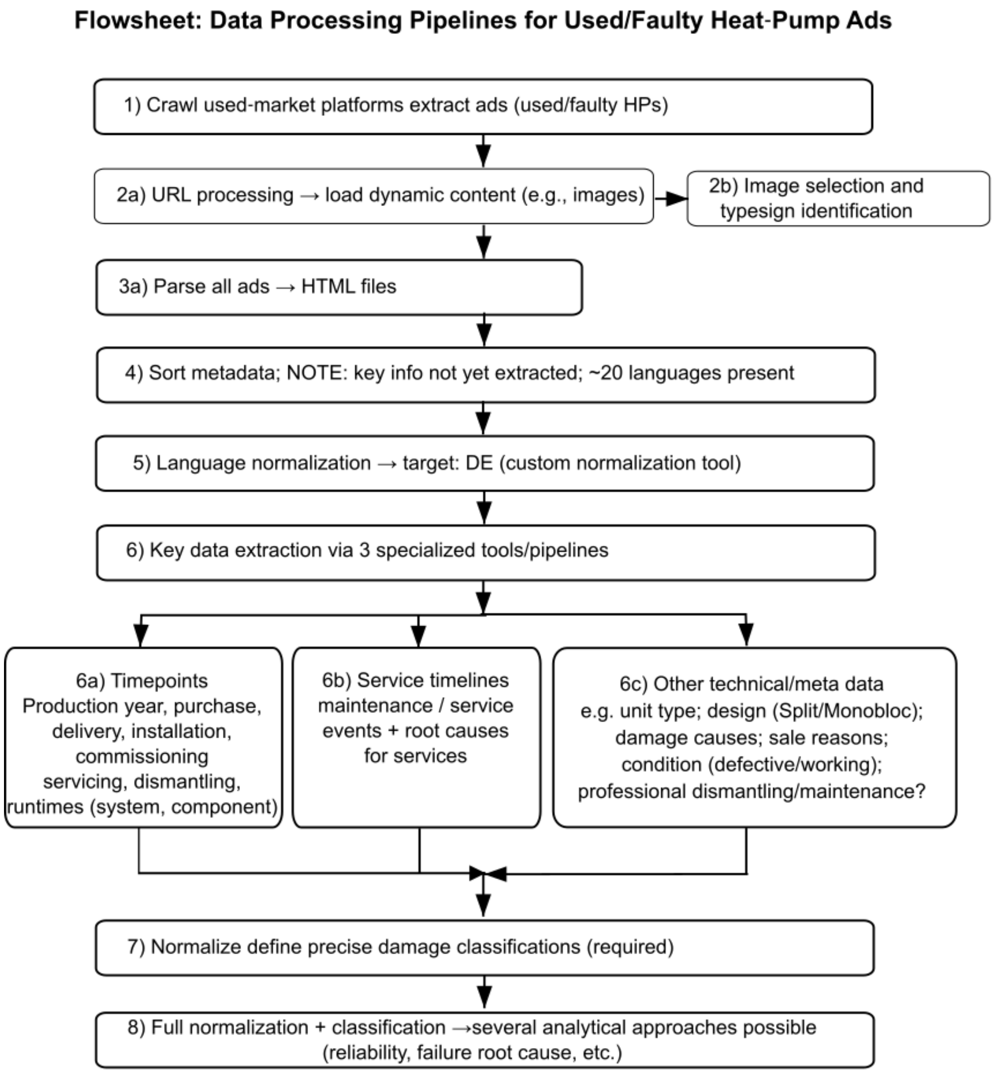
However, based on our experience, compared to official surveys with a very heterogeneous personnel base handling these surveys, often have to fight, too, with incomplete checklists, manual errors in filling out information, and comprehension issues around soft factors when invoices are not available anymore (e.g., timing differences between purchase, installation, commissioning, failure, interim storage periods, and the time before system dismantling). Furthermore, anyone offering this equipment already invested his time into preparing such an advertisement and is motivated to sell the equipment based on the state it has which is an intrinsic motivation to describe its state in a good but also honest way. On the other side any advertiser who does not publish any information or reliable data usually tend to be less attractive and will minimize his chances to sell his used equipment. For this reason, we conclude that any form of data collection can pose major challenges and requires sophisticated filtering mechanisms to build a robust, reliable data foundation.

To obtain something that is usable for cohort-based lifetime estimation, the literature relies on large numbers of observations with at least approximate age and usage information per product group, fits parametric or semi-parametric lifetime models under truncation and censoring, and validates the resulting distributions against independent survey or stock data instead of using simple descriptive statistics or a fixed, one-size-fits-all sample-size rule [6-11].

In Figure 1 the procedure of data acquisition, deduplication, structuring and normalization is described. The data were acquired for 1.5 years from May 2024 and is ongoing. After a complex deduplication and filtering strategy, it resulted into a diverse use state spectrum of HPs and ACs of about 4 000 ads of defective units of HPs and ACs from all over Europe.

For language normalization and a first run of the very heterogenous ad texts and its structuring the large-language model (LLM) GPT 5 Thinking and the translator of DeepL were used [12, 13]. The database built-up process for an effective usage of pattern recognitions was supported using GPT 5 Thinking [12]. The programming was realized with Python [14] in the Spyder IDE [15].

However, due to the complexity of the topic and the sheer size and heterogeneity of the database the preparation with hundreds of pattern recognitions and the work with large LLMs is very challenging. That is why the work needed to be focused on the creation of an automated toolchain before the database analysis



itself could be focused. Before studying the lifetime of equipment based on these used market data the authors

Figure 1: Flowsheet on the procedure of the creation of the used market database

decided to do a conceptual change and to focus in a first step on understanding better the fundamental properties of the generated dataset before a reliable lifetime model could have been developed. Here, the reasons for sales and the damage root causes were identified as two important properties that are needed to distinguish any artificial end of life and a real end of useful life of equipment.

ystematically coding the stated reasons for sale and the damage descriptions in used-equipment listings allows researchers to distinguish hard technical failures from functional obsolescence and elective early replacement, aligning observed retirements with categories such as “malfunction,” “still functional but below expectations,” and “working yet replaced,” which are known to coexist in WEEE[2] flows and HVAC[3] replacement decisions [16-18]. Leveraging preparation-for-reuse methodologies that classify discarded appliances by condition into direct-reuse, repairable, or recycling streams [17], these textual reasons can serve as noisy proxies for technical state to clean lifetime datasets (e.g., treating clear upgrade decisions as censored rather than failure events and down-weighting ads mentioning only cosmetic defects), thereby addressing the critique that many units could, in principle, have remained in service through further repair [16,18].

[2] WEEE: Waste Electrical and Electronic Equipment.
[3] HVAC: Heating, ventilation and Air-Conditioning

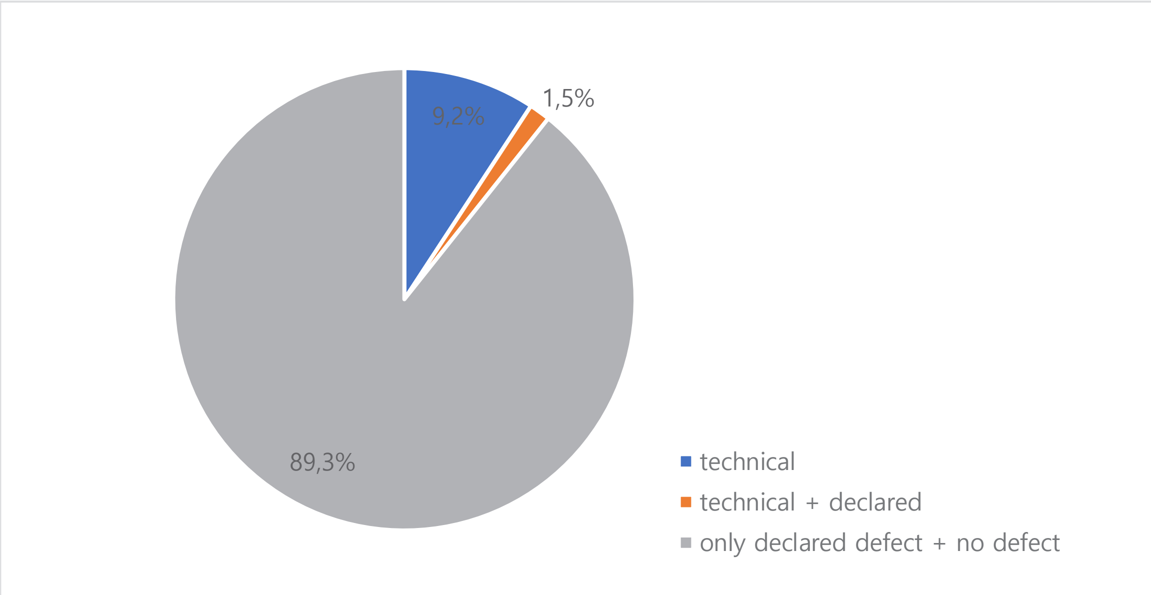


Figure 2: Share of defect types

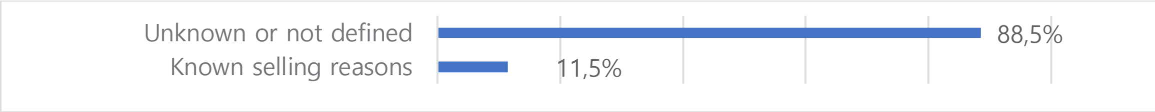
3. Results

All generated figures reflect an analysis and classification that was defined by the authors. The statements by advertisement authors or used equipment owners were very heterogenous. The authors defined a reasoning scheme and patterns for the identification of defect statements and by using the classification features of the used LLM model the spectrum of possible classification schemes was narrowed and homogenized. The final classification scheme as shown in the tables were then – in a final step – generated by the authors.

The defect type plays an important role. It is crucial to understand the difference between our categories of a “declared defective” state and the “technical defective” state of equipment in the generated database to identify the actual state of the equipment. Advertisers of used equipment tend to declare the defect state of equipment to avoid legal liability, even if this often results in a loss in the price that could be obtained. Here, the complexity of the task becomes visible to define pattern recognitions and using LLM models. In **Fehler! Verweisquelle konnte nicht gefunden werden.** the share of defect types occurring in the used market database is shown.

Only definite defective units, the ones classified as having a “technical defect”, here shown with a percentage of 10.7%, are the units that were used for the analysis of the reasons for offering the used equipment and the analysis of root failure causes for it. This is a needed measure for reducing the bias due the large number of commercial ads with new, almost new equipment that was never used.

First, we analyze and discuss the reasons for selling defective HPs and ACs by presenting a distribution of all identified sales reasons, including unknown and undefined categories. After that, we ignore unknown and broad categories and focus exclusively on the known sales reasons that are relevant for defective units.



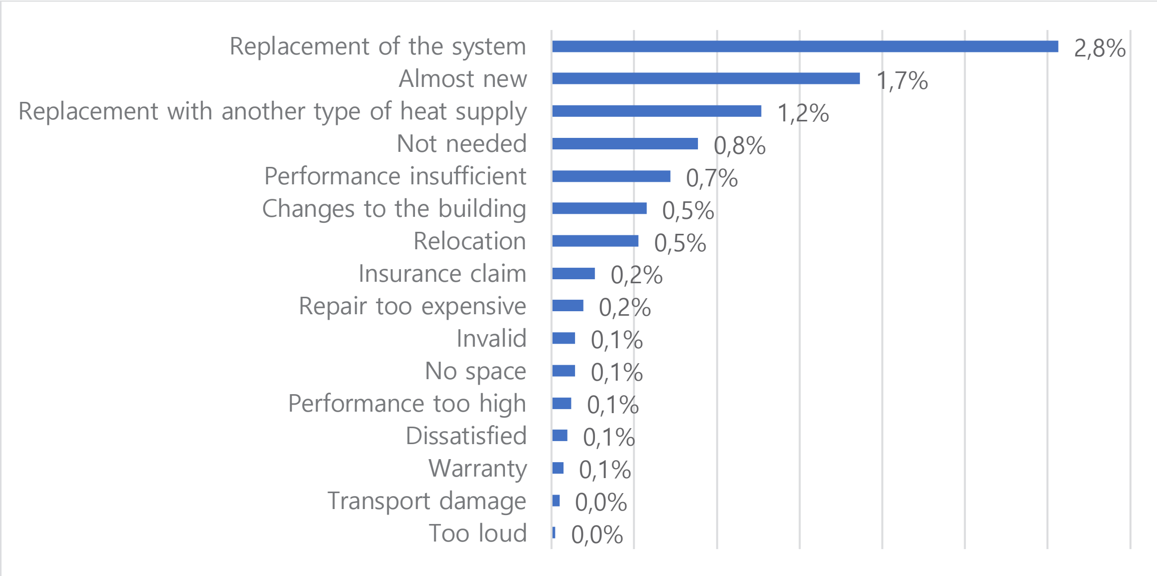


Figure 3: Survey on selling reasons of used equipment. Top: Absolute share of reasons for sale with a huge part in unknowns, not defined reasons. Bottom: detailed spectrum of selling reasons as absolute shares (4 180 out of 36 500 offers).

In Figure 3 the reasons for sales are shown. Given that used markets almost never list equipment still under warranty, the very low share of "warranty" as a reason for sale (0.1%) is likely driven more by this product selection effect than by an inherently low rate of warranty-related problems (e.g. cheap or low-cost air-conditioners that could be part of this analysis). Under this assumption, the distribution suggests that the main drivers for selling used equipment are system changes and contextual factors—such as replacement of the system (2.8%), "almost new" units (1.7%), or switching to another heat supply (1.2%)—rather than issues that would typically be handled within the warranty period. It should be mentioned that reasons for sales are not by definition singular but could be a multitude of reasons. For example, building measures, insufficient performance of the equipment or the exchange of the heating equipment are often interlinked with each other.

In Figure 4 the distribution shows that damage in this dataset is clearly dominated by the refrigeration circuit. Refrigerant leakages have a share of (2.5%) and compressor failures (2.2%). Together account for almost half of all identified causes, while many other components contribute only very small shares. In addition, the relatively high shares of unspecific categories such as "ODU" (1.6%) and "Others" (1.0%), "Other errors" (0.6%), and "Error code" (0.4%) indicate that a notable portion of failures are classified only broadly and not traced back to a single, clearly identified component. In order to keep the table transparent most of the categories were not classified together. Except 120 advertisements in which multiple errors of the used equipment was explicitly stated. And the category "Sensors/Electronics" which was compiled by the authors out of advertisements in which defective PCBs, power boards and sensors were mentioned. It would make sense to reclassify certain subgroups in a higher abstract level for a more condensed root cause list. But this should be driven based on the regime of dominating failure mechanisms which was not yet added to this data analysis.

4. Conclusions

The generation of the used market database aims for supporting and producing better knowledge of the HP and AC stock, including their estimated useful lifetimes. Accurate data on the stock and lifetimes of heat pumps and air conditioners is essential to model decarbonization potential and to determine whether new sales displace fossil-fueled boilers or merely replace worn-out heat pumps.

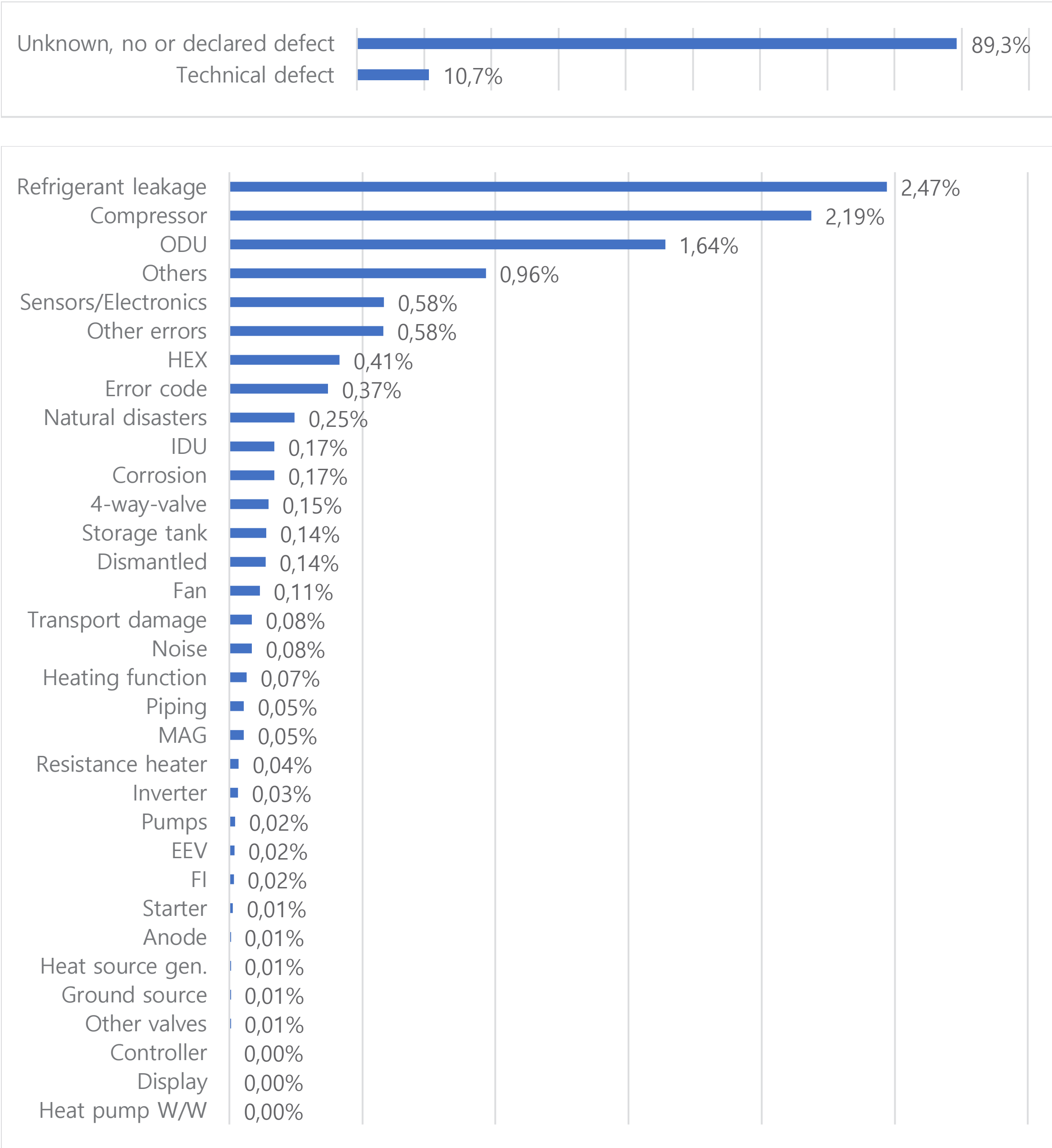


Figure 4: Survey on root causes for selling used equipment. Top: Absolute share of root cause vs. unknown/others. Bottom: detailed spectrum of root causes as absolute shares (3 940 out of 36 500 offers).

Existing EUL work is largely limited to a few studies that use Weibull-based survival models, shipment data, and technician surveys, and must carefully correct for biased, non-random samples when markets are growing fast or when using technician/used-market data. Building on this, the study designs a large-scale data extraction pipeline for European used-market platforms (including about 4 000 defective units since May 2024 which means about 200 defective units are added monthly). The pipeline involves OCR, language normalization, and LLM-supported structuring to cope with heterogeneous, incomplete listings. Instead of directly estimating lifetimes. focus on understanding the dataset's structure by systematically coding reasons for sale and damage descriptions to distinguish artificial end-of-life (e.g. elective replacement or upgrades) from true technical end-of-useful-life, thereby preparing a cleaner basis for subsequent lifetime modelling.

5. Outlook

In general, a reclassification will be implemented to be able to separate used equipment according to domestic HPs for heating, swimming pool HPs as well as ACs such that type-specific studies are possible.

Another potential improvement is for example the identification of all error codes. To be clear, with “error code” it is not meant that the advertiser stated there was an error code for the offered equipment without explicitly stating the specific error code. These include multiple statements of a low pressure alarm which indicates that there is probably a refrigerant leakage issue (and subsequently a failure mechanism). Subsequently, it is possible to improve this root cause analysis based on the underlying and available documentation of this equipment from manufacturers.

Based on this data, a foundation has been created that will facilitate the implementation of these sampled data into a lifetime modeling approach. Different time specifications for acquisition, installation, commissioning, operating time, age, and dismantling are available, and can be used. In terms of lifetime calculation in next steps these features need to be carefully implemented to improve the database generated until now to facilitate the usage of this pipeline and concept for the calculation of EUL data and widely accepted methodologies for the identification of lifetime data (e.g. BP90) [6,19]. At first statements in advertisements for relative and absolute time ranges must be added. In parallel annual and overall HP and AC stocks data of EU member states like Germany with unsaturated markets are prepared to try a cohort-based approach for calculating the lifetimes of equipment. Approaches as developed in [7] need to be implemented for such a more sophisticated analysis of the lifetime distribution in HP stocks and/or AC stocks.

In terms of reliability analysis and the goals of the project a reclassification of root causes based on failure mechanisms will be done to allow optimized strategies for accelerated aging tests.

Acknowledgements

The work on the data analysis was supported by the project HP resilience of the German Federal Ministry of Economics and energy (FKZ 03EN4058).

References

[1] Birolini A. Reliability Engineering: Theory and Practice. 7th ed. Berlin: Springer; 2013.
[2] Lutz JD, et al. Lawrence Berkeley National Laboratory report on U.S. appliance lifetime distributions based on national survey data and Weibull survival models. 2011.
[3] U.S. Department of Energy. Technical Support Document: *Energy Conservation Standards for Residential Central Air Conditioners and Heat Pumps*. U.S. Department of Energy; 2016.
[4] Goldberg, M. Project 73 Track D: *Expected Useful Life (EUL) Estimation for Air-Conditioning Equipment from Current Age Distribution*. DNV. Massachusetts evaluation study; 2018.
[5] DNV. Group *A Effective Useful Life (EUL) Studies for Residential and Commercial HVAC and Water Heating*. California Public Utilities Commission (CPUC) evaluation study; 2024.
[6] Cramer E. *Ordered and censored lifetime data in reliability: An illustrative review*. Wiley Interdiscip Rev Comput Stat 2023;15(2):e1571.
[7] Meeker WQ, Escobar LA. Statistical Methods for Reliability Data. John Wiley & Sons 1998. ISBN:9780471143285.
[8] Elliott MR, Valliant R. Inference for nonprobability samples. Statistical Science 2017;32(2):249–264. doi:10.1214/16-STS598.
[9] eBay. eBay Recommerce Report 2021. San Jose (CA): eBay Inc.; 2022.
[10] Mak M, Heijungs R. *Environmental externalities of secondhand markets—Based on a Dutch auctioning company*. Sustainability 2022;14(3):1682.
[11] Amatuni L, Yamamoto T, Baldé CP, Clemm C, Mogollón JM. *Quantifying total lifetimes of consumer products: Stochastic modelling accounting for second-hand use and establishing an open-collaborative database*. Resour Conserv Recycl 2023;197:107103.
[12] OpenAI. Models: GPT 5 Thinking. OpenAI; https://platform.openai.com/docs/models (last accessed 2026-01-14) within the FHGenie platform fhgenie.fraunhofer.de (last accessed 2026-01-12).
[13] DeepL SE. DeepL Translator. https://www.deepl.com/translator (last accessed 2026-01-12).
[14] Python Software Foundation. (2026). Python (Version 3.12.1) [Computer software]. https://www.python.org (Accessed Jan 13, 2026).
[15] Spyder-IDE Developers. (2026). Spyder: The Scientific Python Development Environment (Version 6.0.0) [Computer software]. https://www.spyder-ide.org (Accessed Jan 13, 2026).

[16] Li J, Liu L, Ren J, Duan H, Zheng L. *Behavior of urban residents toward the discarding of waste electrical and electronic equipment: a case study in Baoding, China*. Waste Management & Research 2012;30(11):1187–1197. doi:10.1177/0734242X12456728.
[17] Bovea MD, Ibáñez-Forés V, Pérez-Belis V, Quemades-Beltrán P. Potential reuse of small household waste electrical and electronic equipment: Methodology and case study. Waste Management 2016;53:204–217. doi:10.1016/j.wasman.2016.03.038.
[18] Evergreen Economics, Michaels Energy, Willems P. Early Replacement Measures Study: Final Phase II Research Report. *A report to the Regional Evaluation, Measurement and Verification Forum*, facilitated by Northeast Energy Efficiency Partnerships (NEEP); 4 November 2015.
[19] International Electrotechnical Commission (IEC). IEC 61649:2008 (2nd ed.). Reliability – Weibull analysis. Geneva: International Electrotechnical Commission 2008.