Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p28964-cd

A Data-Driven Methodology for Forecasting Long-Term Mission Profiles of Household Appliances from Limited Usage Data

Enrico Belmonte
Electrolux Italia S.p.A., Porcia, Italy, enrico.belmonte@electrolux.com
Martin Neumann
AB Electrolux, Stockholm, Sweden, martin.neumann@electrolux.com
Ian Marsh
Industrial Systems, RISE, Research Institutes of Sweden, AB, Sweden. E-mail: ian.marsh@ri.se

The derivation of mission profiles for household appliances is a critical step in the product development process, as accurate knowledge of consumer usage patterns enables engineers to prevent both under- and over-design. The increasing availability of connectivity in domestic appliances has created new opportunities to replace traditional assumptions, often based on surveys or expert judgment, using data-driven insights. However, understanding long-term usage behaviour remains challenging because available connectivity datasets typically cover periods much shorter than the target service life of the products. This paper presents a forecasting methodology for deriving long-term mission profiles from limited historical data, typically one year, considering univariate usage distribution. The proposed approach aims to overcome current data-length limitations and establish robust, scalable methods for predicting design-goal mission profiles in the context of connected household appliances.

Keywords: Reliability, Product Lifetimes, Prediction, Connectivity

1. Introduction

Deriving representative mission profiles is a key task in reliability engineering of home appliances. Connected devices provide detailed insight into usage behavior and patterns. However, available datasets often contain only limited early-life observations. This paper presents a distribution-based forecasting methodology to extrapolate long-term cumulative usage from short-term data. The method explicitly models seasonality, stochastic variability, and user persistency. Monthly usage is modeled with a seasonal Gamma distribution estimated by moment matching. User heterogeneity is captured through a multiplicative persistency factor.

The approach is validated by forecasting 30-month cumulative usage using only the first 12 months of data. Predictions are compared against observed ground truth. Validation is performed at the distribution level using variance decomposition and the 1-Wasserstein distance. Results show that early usage data captures the dominant source of long-term variability. This enables accurate distributional forecasts.

2. Literature research

The frequency and patterns of usage of domestic appliances have been extensively investigated in the literature, particularly for washing machines and dishwashers. These studies aim to quantify how often appliances are used and how usage varies across households, thereby providing empirical benchmarks for real-world operation.

A large share of this literature is based on survey data and observational studies, which report appliance usage at household level. These studies consistently show that usage varies significantly across households, while average usage levels remain relatively stable across different populations and contexts [1], [2]. Consistent findings emerge across studies. Appliance usage shows relatively stable average levels across populations, while varying systematically with factors such as household size and composition [3], [4]. Overall, these results indicate that usage patterns have been widely characterized and that household composition is one of their primary drivers.

In addition to survey-based approaches, usage has also been analysed using service and repair data. These studies provide indirect estimates of usage intensity based on failure observations but are inherently biased toward failed appliances and do not capture the continuous evolution of usage over time [5]. Beyond the estimation of usage frequency, recent work in reliability engineering has highlighted limitations of traditional mission-profile representations. Classical mission profiles are often expressed as aggregated summaries, such as histograms of single stress variables, which may neglect temporal dependence, interactions among variables, and the ordering of usage conditions over time. More recent approaches advocate richer statistical representations capable of preserving lifecycle structure, identifying user quantiles, and distinguishing regular from outlying usage patterns [6]. This perspective is particularly relevant for modern connected products, where field data are available at increasing temporal resolution.

Taken together, these studies show that appliance usage is well characterized in terms of average levels and variability across households. However, important limitations remain. Survey-based methods rely on self-reported data and provide limited temporal resolution, while service data capture only failure events and do not reflect full usage trajectories. At the same time, aggregated mission-profile representations may obscure the temporal structure and heterogeneity that are relevant for reliability analysis.

Most importantly, existing studies focus either on describing observed usage levels or on improving mission-profile representations, but they do not address the problem of forecasting long-term cumulative usage from limited early-life observations. In particular, they do not provide a statistical framework to infer long-horizon usage distributions from short observation windows while preserving heterogeneity across appliances.

The present work addresses this gap by leveraging connected-device data to construct a statistical framework that separates usage intensity from temporal variability and enables extrapolation to long-term cumulative usage distributions from limited observation windows.

3. The Dataset

The dataset used in this study consists of connectivity-derived operational data from Electrolux washing machines. Connected appliances continuously transmit status information during operation, enabling the reconstruction of usage histories at the level of individual cycles. Data collection is performed with user consent during device registration and complies with applicable data protection regulations. All records are pseudonymized prior to analysis.

A key characteristic of connected-device data is its asynchronous and unbalanced structure: appliances can enter or leave the observation window at any time, and the duration of available histories varies across units. In addition, data transmission may be affected by temporary connectivity losses, leading to missing, duplicated, or delayed messages.

To ensure statistical consistency, a subset of several thousand appliances was selected based on minimum observation length, requiring at least 30 months of recorded activity. Data quality filtering was applied to remove appliances exhibiting persistent communication issues or incomplete usage reconstruction.

Following preprocessing, the data were aggregated at monthly resolution, yielding for each appliance a time series of cycle counts. This aggregation level represents a trade-off between retaining temporal structure (seasonality), and achieving robustness against high-frequency noise and data irregularities. The resulting dataset consists of appliance-level monthly usage trajectories, which form the basis for the statistical modelling framework described in the following section.

4. Statistical approach

Monthly appliance usage is observed as the number of completed cycles per appliance. Each appliance is associated with a time series of monthly counts, denoted as $X_{i,m}$, where i indexes the appliance and m the month of operation.
The objective is to characterize the stochastic behaviour of monthly usage at appliance level in a form that supports extrapolation to longer horizons. Let the cumulative usage over T months be defined as:

$$S_{i,T} = \sum_{m=1}^{T} X_{i,m} \qquad (1)$$

where $S_{i,T}$ represents the total number of cycles performed by appliance i over T months. Only the first $H < T$ months are used for model estimation, with $H = 12$ in this study.
Monthly usage is modelled through a multiplicative decomposition that separates temporal effects from appliance-specific behaviour. Specifically, the number of cycles per month is assumed to follow a Gamma distribution:

$$X_{i,m} \sim \text{Gamma}(k_m, \theta_m \cdot Z_i) \qquad (2)$$

where k_m and θ_m are month-of-year–specific shape and scale parameters capturing seasonal variability, and Z_i is an appliance-specific persistency factor representing user-dependent usage intensity. This formulation decouples the temporal structure of usage from cross-sectional heterogeneity across appliances. Figure 1 illustrates the fitted Gamma distributions for selected months of appliance operation, highlighting the ability of the model to capture right-skewed usage patterns and month-to-month variability.

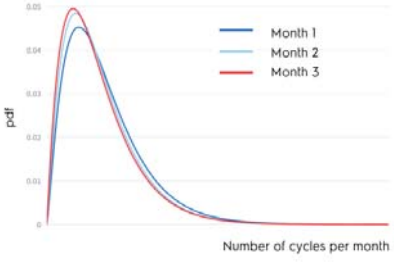

Figure 1. Estimated probability density functions of the number of cycles per month for Months 1–3 of appliance operation.

For a Gamma distribution with zero location parameter, the expected value of monthly usage is:

$$\mathbb{E}[X_{i,m}] = k_m \cdot \theta_m \cdot Z_i \qquad (3)$$

which shows that the mean usage can be expressed as the product of a seasonal component and an appliance-specific scaling factor.
The parameters k_m and θ_m are estimated independently for each calendar month using the method of moments. Denoting by μ_m and σ_m^2 the empirical mean and variance of monthly usage across appliances, the estimators are:

$$\hat{k}_m = \frac{\mu_m^2}{\sigma_m^2}, \hat{\theta}_m = \frac{\sigma_m^2}{\mu_m} \qquad (4)$$

This approach preserves the first two empirical moments and provides stable parameter estimates under noisy and discretized observations typical of connectivity data.
User heterogeneity is captured through the persistency factor Z_i, which is estimated from the first H months of observed usage as:

$$Z_i^{(H)} = \frac{\sum_{m=1}^{H} X_{i,m}}{\sum_{m=1}^{H} \mathbb{E}[X_m]} \qquad (5)$$

This definition normalizes individual usage by the expected population-level behaviour over the same period, isolating appliance-specific intensity from seasonal effects. By construction, this formulation satisfies the consistency condition:

$$\mathbb{E}[Z_i] \approx 1 \qquad (6)$$

ensuring that the decomposition preserves aggregate usage levels.
The coefficient of variation of the persistency factor,

$$\text{CV}(Z) = \frac{\sigma_Z}{\mu_Z} \qquad (7)$$

is used as a dimensionless measure of heterogeneity across the appliance population. Values of Z_i above or below unity correspond to systematically higher or lower usage intensity relative to the population average.
The modelling framework is based on the assumption that the persistency factor is stable over time, such that appliance-specific behaviour observed during the initial period H is representative of longer-term usage patterns. Under this assumption, the joint specification of seasonal parameters and persistency factors provides a complete statistical description of monthly usage at both population and appliance level.

5. Forecasting Long-Term Usage Distributions

5.1 Aggregation and Variance Decomposition

The forecasted cumulative usage over horizon T is defined as the aggregation of monthly contributions:

$$S_T = \sum_{m=1}^{T} X_m \qquad (8)$$

where the appliance index is omitted for clarity.
To characterize the variability of cumulative usage, the law of total variance is applied with respect to the persistency factor Z:

$$\text{Var}(S_T) = \mathbb{E}[\text{Var}(S_T \mid Z)] + \text{Var}(\mathbb{E}[S_T \mid Z]) \qquad (9)$$

The two terms correspond to fundamentally different sources of uncertainty.
The first term, $\mathbb{E}[\mathrm{Var}(S_T \mid Z)]$, captures the contribution of stochastic variability in monthly usage. Conditional on Z, the monthly variables are independent with variance determined by the Gamma model, and therefore this term scales linearly with the horizon T.
The second term, $\mathrm{Var}(\mathbb{E}[S_T \mid Z])$, captures the effect of user heterogeneity. Using the multiplicative structure introduced in Section 4, the conditional expectation of cumulative usage can be written as:

$$\mathbb{E}[S_T \mid Z] = Z \cdot \sum_{m=1}^{T} \mu_m \qquad (10)$$

which implies:

$$\mathrm{Var}(\mathbb{E}[S_T \mid Z]) = \mathrm{Var}(Z) \cdot \left(\sum_{m=1}^{T} \mu_m\right)^2 \qquad (11)$$

This term grows proportionally to the **square of the horizon**, reflecting the persistent nature of user-specific intensity.
In contrast, the conditional variance term can be expressed as:

$$\mathbb{E}[\mathrm{Var}(S_T \mid Z)] = \sum_{m=1}^{T} \mathbb{E}[\,\mathrm{Var}(X_m \mid Z)] \qquad (12)$$

which scales linearly with T, as it reflects independent month-to-month fluctuations.
This decomposition leads to an important structural result: as the forecast horizon increases, the relative contribution of stochastic monthly variability diminishes, while the contribution of user heterogeneity becomes dominant. In other words, long-term uncertainty in cumulative usage is primarily driven by persistent differences in appliance usage intensity rather than by short-term fluctuations.
This result provides the theoretical basis for focusing on accurate estimation of the persistency factor and its dispersion, as these determine the shape and spread of long-term usage distributions.

5.2 Forecasted Distribution Construction

Using the decomposition introduced in Section 5.1, the cumulative usage distribution at horizon T is approximated by a Gamma distribution whose moments are matched analytically.
The expected cumulative usage is given by:

$$\mathbb{E}[S_T] = \sum_{m=1}^{T} \mu_m \qquad (13)$$

where μ_m denotes the mean usage for calendar month m.
The variance of cumulative usage follows directly from the variance decomposition and can be expressed as:

$$\mathrm{Var}(S_T) = \sum_{m=1}^{T} \sigma_m^2 \; + \; \mathrm{CV}(Z)^2 \cdot \left(\sum_{m=1}^{T} \mu_m\right)^2 \qquad (14)$$

where σ_m^2 is the month-specific variance and $\mathrm{CV}(Z)$ is the coefficient of variation of the persistency factor.
The first term represents the aggregation of stochastic month-to-month variability, while the second term captures the contribution of persistent user heterogeneity, which scales with the square of the cumulative mean. This formulation is consistent with the quadratic growth identified in Section 5.1 and provides a closed-form approximation of the variance at any horizon T.
Matching the first two moments yields a Gamma approximation of the cumulative usage distribution:

$$S_T \approx \mathrm{Gamma}(k_T, \theta_T) \qquad (15)$$

with parameters obtained as:

$$k_T = \frac{\mathbb{E}[S_T]^2}{\mathrm{Var}(S_T)}, \theta_T = \frac{\mathrm{Var}(S_T)}{\mathbb{E}[S_T]} \qquad (16)$$

This approach provides a tractable approximation of long-term usage distributions without requiring explicit simulation of individual appliance trajectories, while retaining the effects of both seasonal variability and user heterogeneity.

5.3 Monte Carlo Simulation

Monte Carlo simulation is used to numerically propagate the stochastic model defined in Section 4 and to validate the analytical approximation derived above.
For each appliance, simulations are performed by generating synthetic monthly usage sequences according to the fitted Gamma model:

$$X_m^{(s)} \sim \mathrm{Gamma}(k_m, \theta_m \cdot Z_i) \qquad (17)$$

where Z_i is the estimated persistency factor and s indexes the simulation run. Monthly values are then aggregated to obtain simulated cumulative usage:

$$S_T^{(s)} = \sum_{m=1}^{T} X_m^{(s)} \qquad (18)$$

Repeating this procedure across multiple simulations and appliances yields an empirical distribution of cumulative usage at horizon T.
The simulation serves two purposes. First, it provides a numerical reference for the full distribution implied by the stochastic model, including higher-order effects not captured by moment matching. Second, it enables validation of the Gamma approximation by comparing simulated and analytically derived distributions in terms of moments and overall shape.

6. Validation Strategy

6.1 Data Coverage and Sampling Limitations

While the sample selection process was designed to minimize bias it had limitation due to the structure of the data. Since multiple years of data where needed for the analysis more recent appliances where not represented. While the sampling was agnostic to location and therefore, climate and cultures, its likely that the European market is over represented. It makes up a large portion of the overall population especially for the older models.

6.2 Temporal Hold-Out Design

Model validation is conducted using a temporal hold-out strategy applied at the appliance level. The training and validation datasets contain the same appliances. However, validation is restricted to appliances with at least 30 months of observed usage, so that their long-term cumulative usage can be used as ground truth. Months 1–12 are used exclusively for training, months 1–30 provide the ground-truth cumulative distribution. The appliance population is kept fixed, and training and validation are separated along the time axis. This avoids population bias and information leakage, and reflects a realistic forecasting scenario where only early-life data are available at prediction time.

6.3 Distribution-Level Validation Metrics

Validation is conducted at the distribution level, consistent with the objective of estimating population-level usage rather than individual appliance trajectories.
The evaluation focuses on three complementary aspects.
First, the stability of user heterogeneity is assessed through the coefficient of variation of the persistency factor, $\mathrm{CV}(Z^{(H)})$, computed at increasing observation horizons ($H = 12,24,30$ months). This analysis verifies the assumption that usage intensity can be reliably inferred from early observations.
Second, central tendency and tail behaviour are evaluated through direct comparison of forecasted and observed cumulative usage distributions. In particular, the mean and selected percentiles (P50, P90, P95) are used to assess the accuracy of both typical and high-usage regimes.
Third, overall distributional agreement is quantified using the 1-Wasserstein distance between forecasted and observed cumulative usage distributions. To enable scale-independent interpretation, the distance is normalized by the mean observed cumulative usage, yielding a dimensionless measure of discrepancy.
Together, these metrics provide a consistent validation framework that captures both structural properties of the model and its ability to reproduce the full distribution of long-term usage.

7. Results and Discussion

The empirical results provide consistent evidence in support of the proposed forecasting approach. First, the coefficient of variation of the persistency factor, $CV(Z)$, stabilizes after the first year of usage (Table 1)

Table 1. CV(Z) at different observation horizons for the same appliance population.

Months	CV(Z)
12	0.542
24	0.533
30	0.531

This indicates that user heterogeneity is observable early in the appliance lifetime and can be reliably estimated from short-term data, without requiring long observation periods.
Second, direct comparisons between forecasted and observed cumulative usage distributions at 30 months show good agreement in both central tendency and tail behavior. In particular, the forecasted mean and key percentiles (P50, P90) closely match their observed counterparts, indicating that the model accurately captures both typical and high-usage regimes (Table 2)

Table 2. Forecasted vs. Observed Cumulative Usage at 30 Months (normalized values).

	Observed S30	Forecasted S30	Relative difference
Mean	100	98	-0.26
p50	90	87	-0.43
p90	174	173	-0.12

Third, the normalized 1-Wasserstein distances between forecasted and observed 30-month cumulative usage distributions is below 5%. This confirms overall distributional agreement and provides a compact summary measure consistent with the mean and percentile comparisons. Finally, the comparison of the forecasted and observed probability density functions reveals close alignment in overall shape, including the location of the mode and the decay of the right tail. This qualitative agreement confirms that the proposed model reproduces not only summary statistics but also the full distributional structure of cumulative usage.

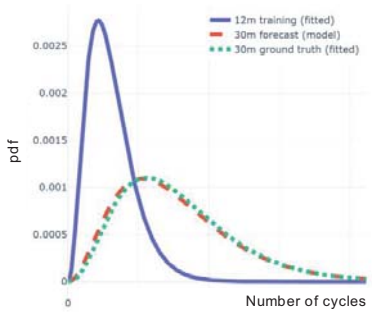

Figure 2. Probability density functions of cumulative usage after 12 months (S_{12}), observed cumulative usage after 30 months (S_{30}), and forecasted 30-month cumulative usage ($\hat{S}_{30}$)

8. Conclusions

This paper presents a data-driven methodology for forecasting long-term appliance usage distributions from short-term observations. By combining a multiplicative stochastic model of monthly usage with an explicit treatment of user heterogeneity and an analytically derived aggregation framework, the approach enables accurate estimation of cumulative usage distributions over extended horizons.
The validation results demonstrate that key model assumptions, in particular the stability of usage intensity, hold empirically and support reliable long-term extrapolation. The proposed framework provides a practical and statistically grounded basis for constructing mission profiles and informing reliability engineering decisions using connected-device data.

References

1. Alt, T., Boivin, D., Altan, M., Kessler, A., Schmitz, A., & Stamminger, R. (2022). Exploring consumer behaviour in automatic dishwashing: A quantitative investigation of appliance usage in six European countries. *Tenside Surfactants Detergents, 60*(2), 106–116. https://doi.org/10.1515/tsd-2022-2488

2. Hook, I., Schmitz, A., & Stamminger, R. (2018). Dishwashing behaviour of European consumers with regard to the acceptance of long programme cycles. *Energy Efficiency, 11*, 1627–1640. https://doi.org/10.1007/s12053-017-9539-y

3. Kruschwitz, A., Karle, A., Schmitz, A., & Stamminger, R. (2014). Consumer laundry practices in Germany. *International Journal of Consumer Studies, 38*(2), 265–277. https://doi.org/10.1111/ijcs.12091

4. Lewitschnig, H. J., Mayrhofer, M., & Filzmoser, P. (2025). A new view to mission profiles. In *2025 Annual Reliability and Maintainability Symposium (RAMS)* (pp. 1–6). IEEE. https://doi.org/10.1109/RAMS48127.2025.10935003

5. Horstmann, S. L., Hüppe, C., Geppert, J., & Stamminger, R. (2019). Socio-demographic differences in washing-up behaviour in Germany. *Tenside Surfactants Detergents, 56*(6). https://doi.org/10.3139/113.110655

6.

7. Tecchio, P., Ardente, F., & Mathieux, F. (2019). Understanding lifetimes and failure modes of defective washing machines and dishwashers. *Journal of Cleaner Production, 215*, 1112–1122.