Building Service Life Revised: A Statistical Survival Analysis of 100.000 Demolitions in Denmark

O. S. Droob[1], M. F. Nybroe[1], N. S. Larsen[1], B. F. Nielsen[1] C. Rode[2], R. A. Svedin[2], N. Aagaard[3]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

[2] Department of Environmental and Resource Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

[3] Department of Civil and Mechanical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

Corresponding author: nisi@dtu.dk

Abstract. The construction industry is currently one of the sectors with the largest contribution to global warming, resource depletion and waste generation. Initiatives to reduce environmental impact are hampered by high construction rates and a large existing building stock that needs to be renovated. Avoiding demolition of existing buildings through service life extension measures is one method to reduce construction rates, but today we have little knowledge about what is being demolished, why and when, and to what extent these buildings can provide building mass and materials to cover our future construction needs. This paper provides new data-driven insights into the service life of the existing Danish building stock based on statistical survival models and an extensive collection of data on 106,908 buildings demolished in the period from 2010 to 2024. For the demolished buildings with use categories Housing, Agriculture & Production, and Transport & Commerce, we found area-weighted median service lives of 87 years, 55 years and 50 years respectively. Using Turnbull's algorithm, we were able to incorporate censored building data, increasing our data foundation to 2,753,327 existing and demolished buildings available for estimation. This leads to slightly higher median service life estimates of 91 years for Housing-buildings, 71 years for the Agriculture & Production-buildings and 65 years for the Transport & Commerce-buildings. These estimates exceed the current Danish regulatory LCA observational periods, suggesting a need for policy revision. The models clearly show that service life varies significantly by building type, highlighting the need for a better understanding of the underlying mechanisms governing why some buildings are demolished, and others are not.

1 Introduction

The Danish building sector is a significant contributor to national CO_2e emissions. Reducing the Danish building sector's climate impact requires a data-driven understanding of refurbishment and demolition dynamics, but data about the building stock are still scattered and underutilized. Many studies have previously attempted to estimate building service lives, but only a few have used real-life data sets of any considerable scale and the results rarely align, with estimates of Danish building service life ranging from 55 to 196 years. Table 1 presents service life estimates for Danish and international residential buildings based on 9 papers, chosen among a larger body of articles to represent the existing diversity of methods and geography.

Paper	Country	Method	Service life estimate (years)
Aagaard et al., 2013[1]	Denmark	Assuming regression rate	Mean:120
Østergaard et al., 2018[2]	Denmark	Regression on Danish building stock (2009-2015)	Median:55
Andersen, 2023[3]	Denmark	Needleman's formula and BYGB12	Median:196
Jensen et al., 2022[4]	Denmark	BBR data on demolished buildings 2010-2021	Mean:85
Andersen and Negendahl,2023[5]	Denmark	Generalized logistic service life prediction model	Mean:91-99
Rincón et al., 2013[6]	Spain	Dwelling stock census comparison	Mean:80
Liu et al., 2014[7]	China	Hedonic modelling	Mean:34
Kornmann and Queisser, 2012[8]	Switzerland	Dwelling stock census comparison	Mean:180
Bradley and Kohler, 2007 [9]	Germany	Historic regional data analysis	Median:300+

Table 1: Overview of the existing literature on service life modeling for Danish and international buildings, showing little consensus.

In this study, the dataset presented in [5] is extended temporally to cover the period from 2010 to 2024 and is presented for all standing Danish buildings in 2024 as well as all buildings demolished between 2010 and 2024. Furthermore, we present a novel method for statistical survival analysis tailored to datasets containing both uncensored (demolished between 2010 and 2024) and right-censored (standing) observations, but not left-censored observations (demolished before 2010). This methodology enables an estimation of survival functions that includes additional data, offering insights for both life cycle assessments and policy formulation.

2 Methodology

This study aims to provide service life estimates of Danish housing buildings based on empirical observations, where the term service life is used to describe the total time from when the building is constructed until it is demolished.

The data on Danish demolitions used in this study is limited to the observational period from 2010 to 2024. Buildings demolished within this period are referred to as uncensored observations. Existing buildings are referred to as right-censored observations, where an exact construction year is known and 2024 serves as a lower limit on demolition year. Left-censored observations where a construction year is known as well as an upper limit on demolition year are not present in the data. Classic methods for survival curve estimation, such as the Kaplan-Meier estimator [10] and the Expectation-Maximization algorithm (EM) [11] cannot incorporate left-censored data, which can lead to bias. We therefore introduce the Turnbull survival curve estimator [12] that incorporates both left-, right- and uncensored data in its estimation process, providing more balanced estimates. As we don't have access to left-censored data (could be obtained from reliable yearly construction data prior to the observational period), we introduce a simple way of creating synthetic left-censored data given the uncensored data and a parametric assumption on the survival curve. A brief visualization of the overall workflow is shown here:

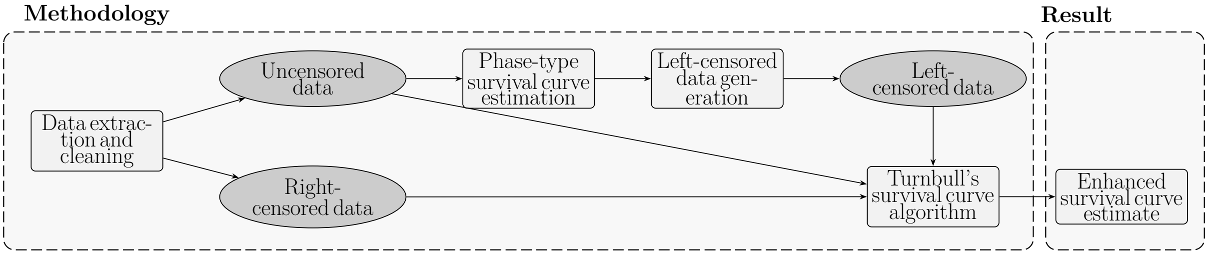


Figure 1: Workflow from data cleaning to survival curve estimate

2.1 Data extraction and cleaning

The raw data for this study is sourced from the Danish Building Registry (BBR) [13], a national building registry administered by the Danish Ministry of Taxation. The BBR is highly useful due to its centralized structure, national scope and high granularity; however, as the registry relies partially on citizen-reporting, data entry errors are to be expected. A systematic assessment of data quality was undertaken and erroneous entries were removed. For the modeling variables in this study, the uncertainty introduced by citizen-reporting is not believed to affect the robustness of the analysis. The dataset was constructed by merging data from [5] with the publicly accessible BBR registry containing records about all standing Danish buildings as well as demolitions from 2017 to 2024.

After removing outbuildings, the resulting dataset consists of $106,908$ buildings demolished from 2010 to 2024 as well as $2,687,072$ existing buildings in Denmark in June 2024. The data variables used in this paper are presented in Table 2.

Variable	**Description**	**Data type**
`ID`	Unique ID for each building	UUID
Demolished	Indicator for whether the building was demolished 2010 - 2024	binary
`Demolition year`	Year the building was demolished (if relevant)	integer
`Construction Year`	Year the building was constructed	integer
`Age`	Service life of demolished building or age of standing building	integer
`Area`	Building area	float (m^2)
`Use-category`	Current building use, categorized according to [1]	categorical
`Refurbishment year`	Year building was refurbished or "Not refurbished"	integer/string

Table 2: Description of building dataset variables

2.2 Phase-type survival curve estimation

The generation of left-censored data requires a model of the uncensored data's survival function. The phase-type distribution is chosen for this estimation, as it has proven applications within survival analysis [14]. The distribution describes the total sojourn time of a stochastic process through a pre-defined number of phases, each phase with an exponentially distributed sojourn time. This makes it a natural choice for service life data, where the phases can be interpreted as distinct states of decay.

Using the uncensored observations, we employ the EM-algorithm [11] from the R-package `matrixdist` implemented in [15] to obtain a model for a phase-type distribution with a general structure for 10 phases. We use random initialization of the parameters for the EM-algorithm and implement a convergence criteria using thresholds for both max-iterations and minimum improvement pr. iteration. Using goodness-of-fit measures and visual inspection the model is found to be a good fit to the uncensored data.

2.3 Left-censored data generation

Right-censored and uncensored data are far more tractable than left-censored data, as left-censored data constitutes data on demolition events before the observational period began. However, left-censored data can be obtained by assuming that the survival curve of the uncensored data follows some defined parametric distribution. In section 2.2 we introduce the phase-type distribution and use it to model the survival curve of the uncensored service lives. The underlying assumption here is that the service life distribution of the uncensored data also describes the service life distribution of the left-censored data.

We denote the number of buildings built in year c which are demolished before 2010 as $N_{c,d}$. We denote the number of existing buildings in 2010 built in year c as $N_{c,ss}$. We then estimate $N_{c,d}$ as

$$N_{c,d} = \frac{N_{c,ss}}{S(2010-c)} - N_{c,ss}$$

where we impose a phase-type assumption on the survival curve S:

$$N_{c,d} = \frac{N_{c,ss}}{\boldsymbol{\pi} \mathrm{e}^{\mathbf{T}(2010-c)} \mathbf{1}_p} - N_{c,ss}$$

We impute left-censored buildings from the year 1570 and onwards, as 1570 is the construction year of the oldest building in the dataset. For the residential buildings, this imputation yields a dataset consisting of 73.0% synthetic left-censored service lives, 26.3% right-censored service lives and 0.7% uncensored data. For Agriculture & Production-buildings, the proportions are 81.4%, 17.2% and 1.4% and for Transport & Commerce-buildings 72.8%, 26.2% and 1.0%. Our estimate that roughly 20-30% of buildings built since 1570 are still existing today seems within reason, but these proportions have not been proposed before and opportunities for validations are few.

2.4 Turnbull's survival curve algorithm

Turnbull's algorithm is a non-parametric unbiased estimator for survival curve estimation developed for settings with left-, right- and uncensored observations [12]. Given a censored dataset, it computes discrete survival probabilities for each observed age, incorporating information from all three types of censoring. The algorithm is based on an initial survival curve estimate, made using a simpler estimator such as Kaplan-Meier [10], which incorporates uncensored and right-censored data. The Turnbull algorithm incorporates the left-censored data iteratively until a convergence criterion is reached. The algorithm can be applied directly on the synthetic dataset created in Section 2.3.

3 Results

3.1 Building uses

Each observation is associated with a building use-category and subcategory. For the use-categories Agriculture & Production, Housing, and Transport & Commerce, the complete service life distributions are shown in Figure 2.

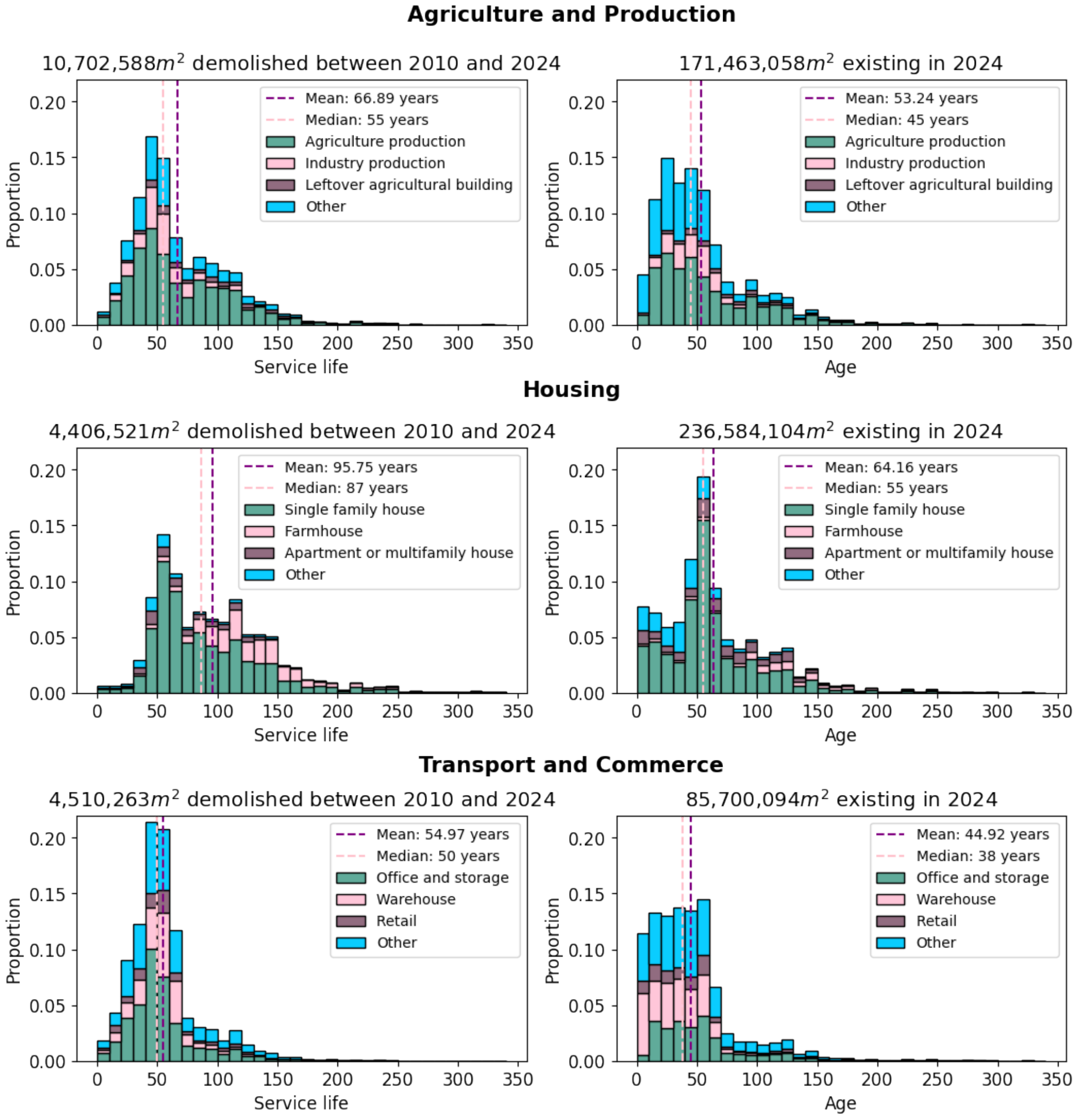


Figure 2: Service life and age distributions for use-categories, colored according to the three most frequent subcategories and weighted based on area. The distributions on the left illustrate the service life of the building stock demolished from 2010 to 2024, while the distributions on the right illustrate the age of the existing building stock in 2024.

Agriculture & Production shows a distinct bimodal service life histogram of the demolished building stock with peaks at 40 to 50 years and 80 to 90 years. The largest subcategory, Industry production, consists of buildings used for commercial production of agriculture or raw material extraction. The peak among agricultural buildings aged 40 to 50 years can be explained by the popularization of hobby farms, where the older farms are more classic agricultural properties. The existing Agriculture & Production buildings exhibit a higher proportion of younger buildings with one wide peak at 10-60 years of age. For the oldest buildings in this category, buildings used for production are not as frequent, with agricultural subcategories accounting for most of the tail.

Among the demolished residential buildings, the service life histogram is wide and flat with a right-skew, showing that residential buildings have been built for centuries and have long service lives. It peaks at 50 to 60 years of service life and again at 110 to 120 years, potentially indicating a mixture distribution. Single-family houses account for the vast majority of demolitions, but for housing buildings with a service life over 100 years, farmhouses account for an almost equal proportion of the demolitions. These older farm properties are prone to demolition due to shifting demographics and a decrease in the number of farmers. The age distribution for the standing residential building stock is rather flat with a big peak between 50 and 60 years of age, the peak consisting primarily of single family housing and apartment buildings. Like with the demolished buildings, the existing farmhouses are mostly above 100 years old, indicating changing housing preferences.

The demolished buildings from the use-category Transport & Commerce have a service life histogram with a distinct peak at 40 to 60 years. The demolitions mainly consist of office and storage buildings. Among the existing building stock a bigger proportion of warehouse buildings are found, though office and storage buildings still account for substantial parts of the age distribution. Both the existing and demolished Transport & Commerce-buildings show more rapid declines after their peak than the other distributions. While agricultural and residential buildings have been in societal demand for centuries, the need for warehouses, offices and retail stores has risen rapidly along with the overall economic growth and consumer buying power in the Danish economy from the 1950s and 60s and forward.

It is clear that the three examined use-categories exhibit heterogeneous building and demolition dynamics, a heterogeneity which should be accounted for in lawmaking and building planning.

3.2 Refurbishment rates and effect on service life extension

The variable "Refurbishment year" indicates whether a building was substantially refurbished and the year it happened. A substantial refurbishment is defined as one where the value of the improvements amounts to at least 15% of the building's value prior to the improvements [13].

Table 3 shows the area-weighted proportion of the building stock that was refurbished in each building use-category, the area-weighted average time before refurbishment as well as the area-weighted median service life of the refurbished and non-refurbished building stock demolished 2010-2024.

	Agriculture & Production	**Housing**	**Transport & Commerce**
% Refurbished	28.8%	39.8%	44.1%
Average time before refurbishment	36.5 years	62.9 years	31.9 years
Median refurbished service life	57 years	93 years	54 years
Median non-refurbished service life	53 years	83 years	45 years

Table 3: Statistics on refurbishment based on the building stock demolished 2010-2024, for each building use category. The statistics are weighted based on the area of each building in m^2.

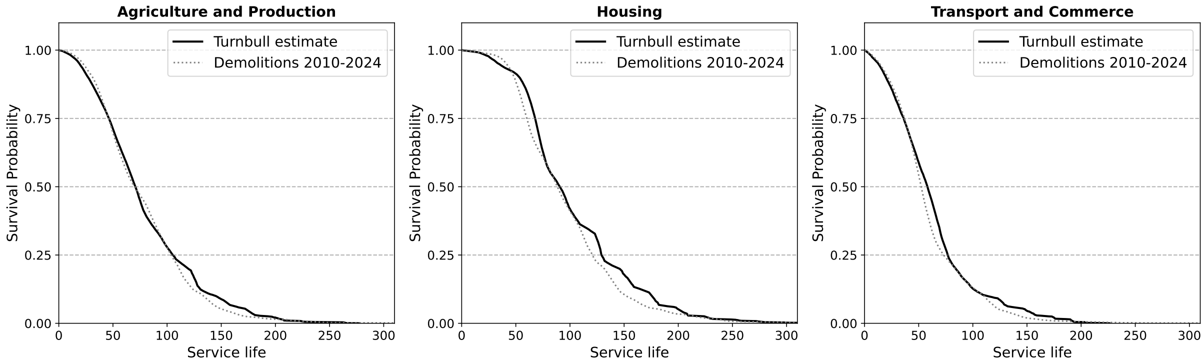
We note that this comparison is subject to survivorship bias. The refurbished buildings are prone to longer service lives, as newer buildings are less suited for refurbishment. The significantly different service life distributions don't necessitate causation. The difference could be caused by other correlated dynamics such as a generally higher level of care for refurbished buildings or wealthier building owners.

For Housing and Transport & Commerce, around 40% of the demolished building stock is refurbished, while only 29% of the demolished building stock used for Agriculture & Production undergo refurbishment. The average time before refurbishment varies among the building use-categories - with demolished residential buildings on average being refurbished later in their service life compared to the other uses. The demolished Transport & Commerce and Agriculture & Production building stock both have an average time before refurbishment approximately halfway through their service life.

For all categories the refurbished buildings have higher median service lives than the non-refurbished buildings, with the gap size varying among the use-categories. For residential and commercial buildings the average service life extension is 10 and 9 years, respectively. For Agriculture & Production the service life extension is only 4 years, coinciding with this category being the least frequent subject of refurbishment - possibly indicating that use-categories with longer potential service life extension are more prone to undergo refurbishment.

3.3 Survival curve and service life estimates

For each use-category, a survival curve is estimated using Turnbull's algorithm and shown in Figure 3. As the algorithm assumes the observations to be independent of each other, the analysis is based solely on complete buildings and does not take the area into account.



	Demolitions 2010-2024	Turnbull estimate
Q1	46	46
Median	70	71
Q3	104	105

	Demolitions 2010-2024	Turnbull estimate
Q1	61	68
Median	89	91
Q3	121	129

	Demolitions 2010-2024	Turnbull estimate
Q1	37	36
Median	53	58
Q3	73	77

Figure 3: Upper: Turnbull survival curves estimated from data, including the synthetic left-censored data. The dotted line shows the empirical service life distribution for the uncensored buildings. Lower: The service life quartile values, computed on the uncensored data for demolitions from 2010 to 2024 and computed using all the data in Turnbull's algorithm

The Turnbull survival curve estimates are relatively smooth functions until service lives of approximately 100 years, where the data is sparser. As the Turnbull algorithm does not impute values, when $N_{c,d}$ becomes sufficiently small, the predictions increment less smoothly. For example, the reported construction years for older buildings have a slight tendency towards round 10-year increments (1940,1930 etc.), meaning that the data itself is not completely smooth.

The estimated survival curve in the category Agriculture & Production shows a slightly flattened S-shape, with a median service life of 71 years. Comparing the estimated survival curve to the survival curve obtained directly from the uncensored data shows equal or higher estimates of the quartiles. This means that incorporating the censored data extended the expected service life for this category compared to the straight-forward computation of median service life from only the demolished buildings.

For the residential buildings, the algorithm estimates a survival probability until age 50 of 91.6% and to age 100 of 41.9%, and a median service life of 91 years. The survival curve approximates the S-shape seen in the uncensored service life data, although it consistently estimates higher quartiles than the ones observed for houses demolished 2010 to 2024. The longer estimated service lives are a result of incorporating the right- and left-censored data.

For the buildings with use-category Transport & Commerce, the Turnbull estimator predicts slightly lower survival probabilities than seen in the uncensored demolitions from 2010 to 2024. However, from around 50 years of service life, the Turnbull estimate is consistently higher than the survival probabilities observed among the uncensored observations, suggesting again that the estimations made with the inclusion of left- and right-censored data project higher survival probabilities and longer service lives than using just the uncensored data. This is also shown by the median and 3rd quartile for service life being estimated to 58 and 77 respectively, compared to the demolished buildings having median and 3rd quartile service lives of 53 and 73.

4 Conclusion

This study began by advocating for more data-driven research on building service lives, particularly focusing on how bias affects empirical computations of survival curves. The proposed model ties together phase-type modeling, generation of synthetic left-censored data and Turnbull's algorithm to incorporate censored data in the estimation of survival curves for buildings. The survival curves, modeled on a dataset consisting of 2,753,327 Danish demolished and existing buildings, estimate slightly higher service lives than the observed demolitions, suggesting that the inclusion of censored data can provide new insights into building service lives.

The study lays the foundation for future work in building service life estimation, presenting data for 100,000 Danish demolition cases. It shows the heterogeneity in service life distributions among different building use categories, with most building use categories lasting at least as long as the current LCA observational period requires.

This data-driven approach can provide insights, but cannot fully capture the diverse range of factors influencing building demolition decisions. Dynamic processes such as shifts in housing demand, aesthetic preferences, technological advancements, and economic factors all affect demolition decisions. The study does not address the non-stationarity in Danish construction activity caused by building booms. A possible path for future work would be to utilize the hazard rates of buildings over their service life to alleviate some of the bias inferred by the non-stationarity.

References

[1] Aagaard, N. J. et al. *Levetider af bygningsdele ved vurdering af bæredygtighed og totaløkonomi.* SBI forlag, 2013.

[2] Østergaard, N. et al. *Data Driven Quantification of the Temporal Scope of Building LCAs.* 2018.

[3] Andersen, M. L. *Karakteristika for huse der rives ned med henblik på nybyggeri.* 2023.

[4] Jensen, J. O. et al. *Nedrivning af enfamiliehuse: Omfang og årsager.* 2022.

[5] Andersen, R. and Negendahl, K. “Lifespan prediction of existing building typologies”. In: *Journal of Building Engineering* 65 (2023).

[6] Rincón, L. et al. “Service life of the dwelling stock in Spain”. In: *The International Journal of Life Cycle Assessment* 18 (June 2013).

[7] Liu, G. et al. “Factors influencing the service lifespan of buildings: An improved hedonic model”. In: *Habitat International* 43 (2014), pp. 274–282.

[8] Kornmann, M. and Queisser, A. “Service life of the building stock of Switzerland”. In: *Mauerwerk* 16 (2012), p. 210.

[9] Bradley, P. E. and Kohler, N. “Methodology for the survival analysis of urban building stocks”. In: *Building Research & Information* 35.5 (2007), pp. 529–542.

[10] Kaplan, E. L. and Meier, P. “Nonparametric Estimation from Incomplete Observations”. In: *Journal of the American Statistical Association* 53.282 (1958), pp. 457–481.

[11] Asmussen, S., Nerman, O., and Olsson, M. “Fitting Phase-Type Distributions via the EM Algorithm”. In: *Scandinavian Journal of Statistics* 23 (1996).

[12] Turnbull, B. W. “The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 38.3 (1976), pp. 290–295.

[13] Vurderingsstyrelsen. *"BBR"*. `https://bbr.dk/forside`. Accessed:20/08/2024. 2025.

[14] Bladt, M. and Nielsen, B. F. *Matrix-Exponential Distributions in Applied Probability.* Springer, 2017.

[15] Bladt, M., Yslas, J., and Müller, A. *Matrixdist: Statistics for Matrix Distributions.* R package version 1.1.9. Comprehensive R Archive Network (CRAN), 2023.