SURVIVAL FUNCTIONS OF BUILDING STOCKS AND COMPONENTS

Patrick Erik BRADLEY[1]
Ebba BUERGEL-GOODWIN[2]
Claudio FERRARA[1]
Niklaus KOHLER[1]

[1]Institut für Industrielle Bauproduktion, Universität Karlsruhe, Englerstr. 7, D-76128 Karlsruhe, Germany, {patrick.bradley, claudio.ferrara, niklaus.kohler}@ifib.uni-karlsruhe.de
[2]ECOTECH, Kapuzinerstraße 84e, A-4020 Linz, Austria, ebba.buergel-goodwin@ecotech.cc

Keywords: Building stock, right censoring, survival analysis, Kaplan-Meier estimator.

Summary

An example-based methodology is developed for estimating survival functions of large building stocks for which lifetime data is difficult to obtain, e.g. for German townships. By considering demolished as well as undemolished buildings, techniques from censored data are applied: the event considered being demolition, undemolished buildings are then right-censored, thus leading one to the Kaplan-Meier estimator for the survival function. This methodology is applied to a German middle-sized small town, where, right-censoring itself is estimated by a random sample from a cadastral plan, and a complete inventory count is performed for the demolished part of the stock. The resulting estimates for the survival functions are compared for residential and non-residential buildings by age classes. In a further step, fitting parametric survival functions allows predicting the behaviour of a building stock. The example building stock was studied in the project "Validierung eines integrierten, dynamischen Modells des deutschen Gebäudebestandes" (Validating an integrated, dynamic model of the German building stock) funded by the Deutsche Forschungsgemeinschaft 1999-2003, aiming at developing methods for collecting and analysing large-scale building stock related data.

A related study on methods towards survival analysis of building components performed by the second author for her diploma thesis, Buergel-Goodwin (2004), is described and first results are shown.

1. Introduction

Because building related data is in general hard to obtain, a methodology for dealing with sparse information has to be developed. For doing so, a township building stock was studied in a research project funded by the Deutsche Forschungsgemeinschaft (DFG), focussing on how demolition of buildings depends on age and function. Here, we describe the methods and show the results of the study. But we remark that the emphasis lies on the methodology.

Firstly, building age is given by the exact year of construction. However, this level of granularity is impractical. Therefore, age classes were defined, and the age-function classes thus obtained are assumed homogeneous subsets of the building stock. Age-class definitions in Table 1 aremade from constructional considerations of buildings of various times (Ferrara (2004), p.35 or Bischoff et al. (2005) Tabelle 1).

Table 1 Age class definitions.

Age class	1	2	3	4	5	6	7	8	9
Begin		1835	1871	1919	1934	1950	1965	1977	1995
End	1834	1870	1918	1933	1949	1964	1976	1994	today

The stock size being large, a random sample was taken from the building stock of the year 2000, and with it the function and age distribution at that time was estimated. For describing the building stock at times before 2000, knowledge about the demolished part of the building stock is needed. This was gained through a complete inventory count of the demolished buildings of the stock for the time of 1936 on, allowing to perform a survival analysis.

However, knowledge about the aging process of components is necessary for planning or management. Simply adopting literature values leads to arbitrary results (Figure 1).

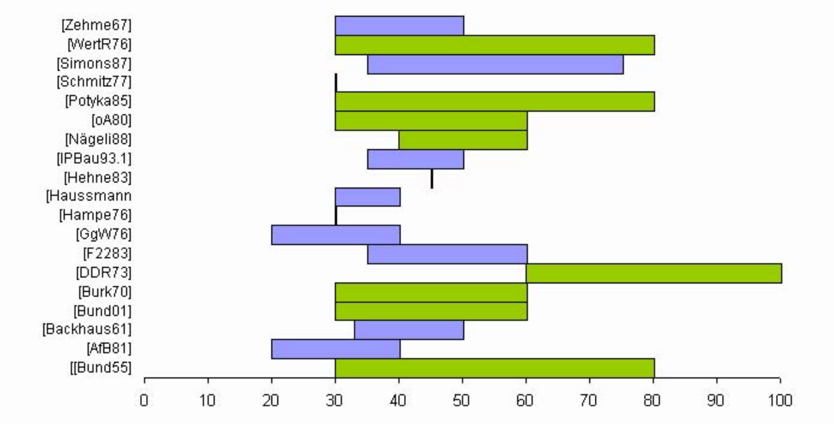

Figure 1 Values of service life of a building component in different sources, Meyer (1994).

Meyer (1994) analysed a large number of *existing* building components, combining survival functions with degradation levels for twelve components. Studying many cases, he established characteristic aging curves which associate duration with a certain state of degradation. Knowing the state of degradation and exposition allows estimating the probable residual lifetime. This partially explains the large variance in the literature.

Data of building components for our second study was gained from three non-residential buildings in three different South German towns. Again, emphasis lies on methods, not on results.

More information on the project can be found at BEVAL (2004).

2. Data acquisition for building stock

If the building stock to be analysed is already incorporated in a database, then analysis takes place right away. However, when dealing with the building stock of a German town, the first problem is to get the existing data into a form ready for analysis. Time-data of buildings exist abound, but is spread over various institutions and archives, and mostly in paper form. Among the most important sources for our first study id the archive of the building insurance ”Sparkassenversicherung Badische Gebäudeversicherung” (named BI, here) for obtaining dynamic building data on the town of Ettlingen in the Northern part of Baden (a region in the South of Germany). It is a medium-sized small town with 38,716 inhabitants at this time. Its size makes it a representative of the biggest class of German towns (about 24% of all German towns have 20,000-49,999 inhabitants) in which approximatively 48% of Germany’s residents live.

Our building stock consists of the residential (RB) and non-residential buildings (NRB) of Ettlingen excluding the smaller surrounding towns adhered to Ettlingen between 1972 and 1974.

Another important source was the digitalised cadastral plan of Ettlingen from the year 2002 showing the registered buildings in the year 2000. It will be referred to as ”the cadastral plan”, and its buildings are defined to make up the building stock of Ettlingen in that year. Of course, it does not quite represent the real building stock at the time, but seems a good enough approximation.

The cadastral plans are updated by the office of survey without retaining a copy of the state before the update, thus making a spacial time series difficult. They contain information on location, surface and function of the buildings. Ages, however, are not contained.

3. Random sampling from a cadastral plan

The problem of acquiring age data of the Ettlingen building stock (cf. the preceding section) was solved by a random sample as follows (cf. Bader et al. (2001) for details): a cadastral plan partitions the area it shows into numbered parcels, and buildings are closed polygons which do not always lie completely in one parcel—intersections with more than one parcel do occur. Since in this case, buildings are not identifiable, the cadastral plan was incorporated into a geographic information system and each building was given a unique permanent number.

In order to determine a reasonable sample size, the building stock was partitioned into function classes of known size. The gross volume (gv) was chosen to decide upon the size of the sample by using a priori information on the mean gv and its variance for each class (Bader et al. (2001), Tabelle 2.1). Sampling theory then allows to calculate confidence intervals and sample sizes (Table 2). Thus the total stock of non-residential buildings (NRB) in the year 2000 consists of 2166 buildings and does not contain the 1543 garages and 8 institutional bulidings excluded from the study. The desired sample size of non-residential buildings is calculated as 716 buildings.

Since the BI keeps building information by address, it seemed practical to collect the data of all buildings in an address containing a sampled building. However, some inconsistency occurs due to different definitions in the cadastral plan and in the BI: a building as in the BI is obtained from a building according to the cadastral plan by subdividing it along the different addresses under which it is stored in the BI. This guarantees that each BI-building lies in a unique parcel.

Another inconsistency problem arises when trying to identify sampled buildings in the cadastral plan of 2000—in particular, when a parcel contains more than one building of the same function. Then geometric information (e.g. floor plans) from the BI has to be compared with the buildings of the same parcel in the cadastral plan. In the case of Ettlingen, this work has yet to be completed. Some attention has to be given to the gap of 6 years between the levels of information in the BI and the cadastral plan. Two addresses proved to be extraordinarily complex: Pforzheimer Strasse 134 and 68-74, both in industrial areas of Ettlingen. There, only a superposition of the available plans from the BI and comparison with their files can yield correct building histories. For convenience, this was realised only for the first address. It seems that results can be improved by considering external sources on these particular building stocks (e.g. chronicles etc.). Therefore, the building stock is (re)defined by excluding the Pforzheimer Strasse 68-74 which according to the database [db] used by Bader et al. (2001) consists of 50 undemolished NRB. Likely, superposition will yield more buildings there. Table 2 shows the reduced stock and sample sizes.

Table 2 The Ettlingen building stock of the year 2000 with samples

	RB	NRB
Stock of year 2000	3084	2166
Reduced stock 2000	3084	2116
Required sample size	635	716
Sample size [db]	918	858
Reduced sample size [db]	918	808

4. The demolished buildings of Ettlingen

An inventory count of all demolitions for the period 1936-1994 was taken from the BI in two steps: First, all files in the BI were inspected upon whether they contain demolitions or not. Secondly, detailed information about demolished buildings in the addresses found in the first step was extracted. This gives the demolitions between 1936 and 1994.

For the period 1995-2000, the office of building administration (OBA) was consulted (cf. Bradley (2004) for details). Still, the exact number of demolitions between 1936 and 2000 remains unknown due to the incompleteness of the OBA records (only 47 demolitions between 1995 and 2000), and due to the loss of so-called "dead files" when the documents from the BI were archived centrally after the monopoly ceased in 1994: formerly, the files were kept locally in each town. A dead file was often created when an address was completely demolished and nothing was constructed on the address for some time. These files were kept in special sections of the local archive and could be ”revived” when new construction began. On gathering all local archives together after 1994, only files containing insured buildings were taken to the central archive.

The demolitions found thus are (Table 3) are considered as very close to a complete inventory count of demolitions occurring in Ettlingen between 1936 and 2000, and hopefully underestimate the true complete inventory count by only a small number.

Table 3 The number of demolitions between 1936 and 2000 found in the BI

RB	NRB	changing function	total
146	846	12	1004

5. Estimating the survival function of the building stock

For estimating the survival function *S(t)*, we make use of the collected data on both the demolished and undemolished part of the stock, as the survivors do have an influence on *S(t)*.

Let X_b be the random lifetime before demolition of a fixed building *b* from a stock *B*, viewed as a non-negative random variable. Further, we assume the stock as homogeneous, i.e. the family of all X_b (*b* in *B*) is assumed independent and identically distributed. Now, for each building *b* there is some positive censoring time C_b (e.g. end of study), and the value of $T_b = \min\{X_b, C_b\}$ is observed. If $X_b > C_b$, then *b* is called *right censored*, otherwise *b* is *uncensored*. In Figure 2, the length of each beam is the observed lifetime, and + indicates a right censored building.

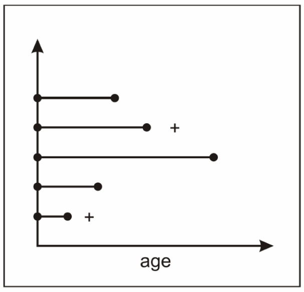

Figure 2 Illustration of censored lifetimes. The + indicates right censoring.

If further it is assumed that among the buildings of equal T_b the censored ones are a random sample (whenever covariates coincide), then one speaks of *independent* censoring. In this case one can use the *Kaplan-Meier estimator* as a non-parametric estimator for *S(t),* defined as

$$\hat{S}(t) := \begin{cases} 1, & \text{if } t \le t_1 \\ \prod\limits_{t_i \le t} \left(1 - \frac{d_i}{Y_i}\right), & \text{if } t \ge t_1. \end{cases}$$

Here, the t_i are an increasing sequence of different life-times of demolished buildings, d_i the number of buildings demolished at age t_i, and Y_i the number of buildings having reached an age of at least t_i. The latter variable is called the number of buildings *at risk* at time t_i. A more detailed treatment can be found e.g. in the monograph by Klein and Moeschberger (1999).

In our example, right censoring is given uttermostly by undemolished buildings. However, the sample buildings with known construction years are not enough for estimating the yearly number of building constructions. So, in order to obtain a finer age distribution, a partition of the age class defining intervals was generated from the sample, and the number of buildings constructed within these smaller intervals was estimated by that sample. Finally, the buildings not contained in the sample are assigned the midpoint of the corresponding interval as year of construction (Figure 3).

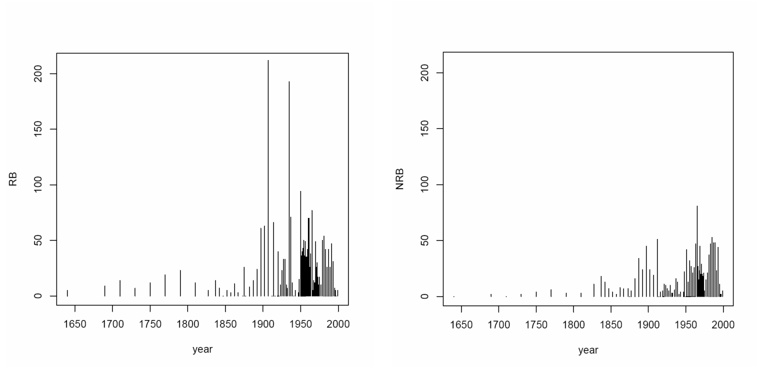

Figure 3 Assigned construction years to unknown RB (left) and NRB (right).

The very left interval was chosen to be [1600,1680) in order to take account that Ettlingen was almost completely destroyed in 1689 when French troops occupied the region during a succession war. There do exist undemolished buildings from before 1698, e.g. St. Martin's Church (built 1479), but the eldest sampled building is of 1670. In any case, we believe the midpoint 1640 is not too far off the true mean construction year for this (refined) age class.

We remark that the distribution calculated from the sample necessarily has to take integer values (number of buildings in each class). Yet, naïvely multiplying the proportions by the total number *N* of buildings does generally not yield natural numbers. Rounding is of no use, as the sum of the roundend estimators for each class will not likely equal *N.* Adopting a remedy from the monograph by Pukelsheim (1993), Ch. 12, let x be a variable *x*, and let a random sample of size *n* from a set of *N* elements give the frequency table for the values of *x*

x_1	$\dots$	x_ℓ
w_1	$\dots$	w_ℓ

where the w_i are assumed positive with sum equal to 1. For estimating the size n_i of class *i* calculate the two numbers

$$d := \left(\sum_{i=1}^{\ell} n_i\right) - n \qquad \text{and} \qquad n_i = \left\lceil \left(n - \frac{\ell}{2}\right) w_i \right\rceil .$$

If *d* = *0,* we are done. Otherwise, if *d* < *0,* then let *j* be such that

$$\frac{n_j}{w_j} = \min_{i \le \ell} \frac{n_i}{w_i}$$

and raise n_j by one. Alternatively, if *d* > *0,* then let *k* be such that

$$\frac{n_k - 1}{w_k} = \max_{i \le \ell} \frac{n_i - 1}{w_i}$$

and lower n_k by one. Reapeating until *d* = 0 yields class sizes $n_1,\dots,n_l$ summing up to *N.*

In any case, from the estimated age distribution and the demolitions we estimate the evolution in time of Ettlingen since 1936, as shown in Figure 4.

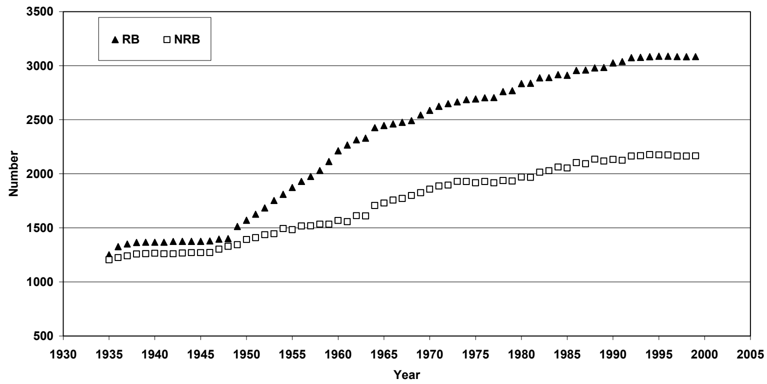

Figure 4 Estimated number of RB and NRB in Ettlingen 1936-2000.

The results obtained yield Kaplan-Meier estimators for *S(t)* (Figure 5). Right censoring itself given by a random sample, we abstain from estimators for the variance or confidence intervals. The seemingly long period of *S(t)* = 1 is due to a *left truncation* of all buildings demolished before the year 1936—these did naturally not enter the study and so did not change the value of *S(t)*.

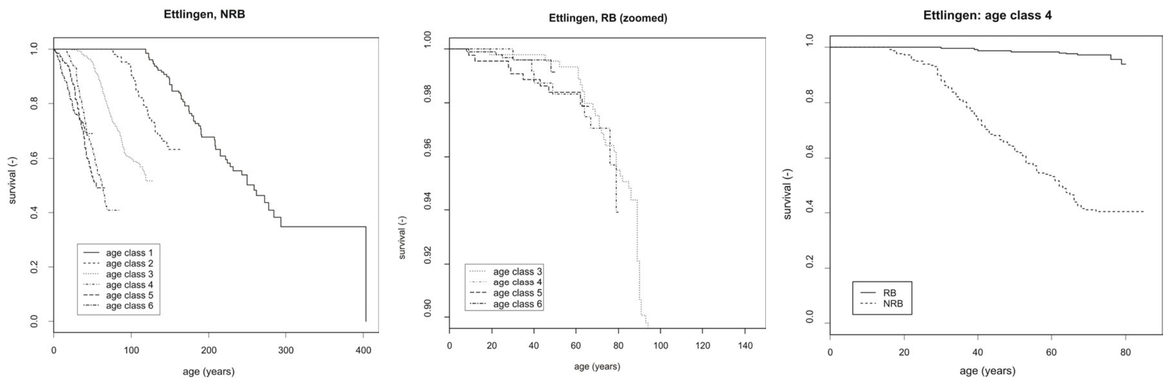

Figure 5 Kaplan-Meier estimators of survival function for Ettlingen.

The survival function of NRB's seems to be lower than that of RB's of the same age class (Figure 5, right). The first 100 years of *S(t)* seem to coincide for RB's of all age classes, but not for NRB's. However, these observations remain conjectural and have yet to be validified statistically.

Anyway, a Weibull fit for *S(t)* allows extending its range to e.g. the next 100 years, and estimating the age class distribution in year 2100. It will not not have decreased much by then (Figure 6).

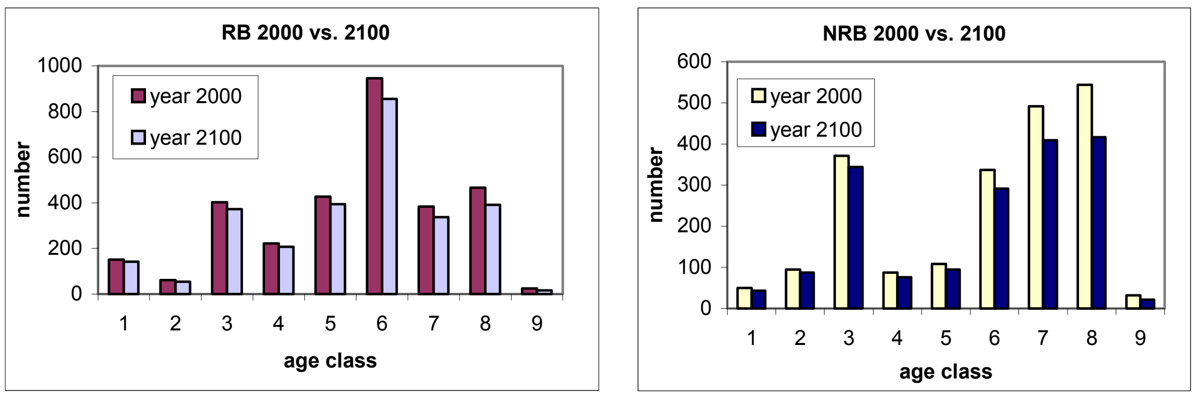

Figure 6 Estimated age class distributions of Ettlingen in years 2000 and 2100.

6. Survival of building components

An empirical investigation leading to the second author's diploma thesis (Buergel-Goodwin (2004)), studies maintenance, repair and replacement operations of three public buildings have reconstructed from the building owner's records over a longer period. Data going down to the component level, a survival analysis of building components is possible.

As the same authority managed the buildings over a long period, information about costs, cleaning etc. was available. Nevertheless, extracting the information from building archives and maintenance records was tedious. Some recent data recorded electronically were already lost whereas older data were well protected against loss. While for the last five years work proof and bills exist, earlier data could be reconstructed only through archived correspondences.

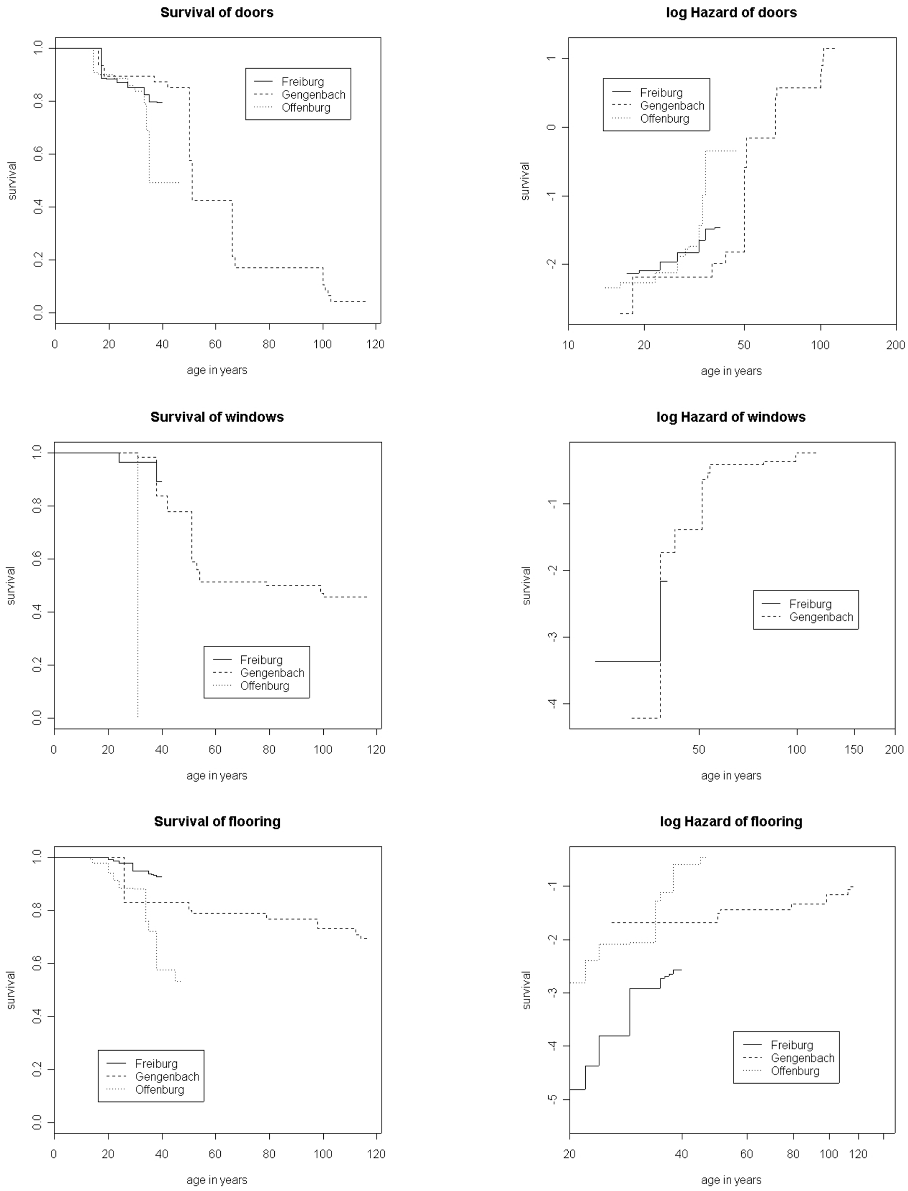

Figure 7 Empirical survival functions (top) and log-log plots of cumulative hazard functions (bottom) of building components.

All original building components for each of the three buildings can be assumed to have been introduced into the building in the same year, so left truncation is excluded, and right censoring can occur only after all uncensored components have been removed. Therefore, the Kaplan-Meier estimator is essentially the survival function of the uncensored components (Figure 7). It seems that the the survival functions might be independent of the buildings. However, as the log-log plot of the cumulative hazard function for the flooring shows, the Gengenbach building's survival function seems to differ from the other two. Of course, we observe the sudden elimination of all windows at the same time in one building, leading to a one-step survival function.

Acknowledgements

The authors thank the BI and the town of Ettlingen for allowing access to data, and also the many helpers for their extraordinary job on the quite tedious and time-consuming data collecting and processing. The Surveyor's Offices at Freiburg and Offenburg put at our disposition the archived documents and the staff was very helpful in answering in detail all questions about maintenance strategies. Wolfgang Bischoff and Bernhard Klar are warmly thanked for fruitful discussions.

References

Bader, G., B. Schwaiger and C. Ferrara, 2001, Methode zur Analyse von Gebäudebeständen, Alterungsmodelle und Prognose von Stoffströmen mit Anwendung auf den Gebäudebestand der Stadt Ettlingen, Research Paper, Institut für Industrielle Bauproduktion, Universität Karlsruhe.

BEVAL, 2004, Validierung eines integrierten, dynamischen Modells des deutschen Gebäudebestandes, http://www.ifib.uni-karlsruhe.de/web/ifib_forschung/forschung_show.php?forschung_id=25.

Bischoff, W.; P.E. C. Bradley, Ferrara and N. Kohler, 2005, Überlebensanalyse von Gebäudebeständen am Beispiel der Stadt Ettlingen, Stadtforschung und Statistik, 1-2005, p. 32-35.

Buergel-Goodwin, E., 2004, Vergleichende Studie zu Erneuerung, Unterhalt und Betrieb von Bestandsgebäuden auf Bauteilebene, diploma thesis, Universität Karlsruhe.

Bradley, P.E., 2004, Validierung eines integrierten, dynamischen Modells des deutschen Gebäudebestandes: Statistischer Teil, Technical Report 2004-01, Institut für Industrielle Bauproduktion, Universität Karlsruhe.

Ferrara, C., 2004, Validierung eines integrierten, dynamischen Modells des deutschen Gebäudebestandes: Allgemeiner Teil, Technical Report 2004-04, Institut für Industrielle Bauproduktion, Universität Karlsruhe.

Meyer, P., 1994, Alterungsverhalten von Bauteilen und Unterhaltskosten: Grundlagen für den Unterhalt und die Erneuerung von Wohnbauten, Bundesamt für Konjunkturfragen, Bern.

Klein, J.P., M.L. Moeschberger, 1999, *Survival Analysis. Techniques for Censored and Truncated Data,* Statistics for Biology and Health, Springer, Corrected third printing.

Kohler, N., U. Hassler and H.,Paschen (eds.), 1999. Stoffströme und Kosten im Bereich Bauen und Wohnen. Studie im Auftrag der Enquete Kommission zum Schutz von Mensch und Umwelt des deutschen Bundestages, Berlin, Springer Verlag.