Reliability Analysis of Water Distribution Networks in Salvador, Bahia by Use of Survival Analysis and Ishikawa Diagram: Study Focusing on Real Losses

Lucas Almeida[*a], Karla Esquerre[a], and Brenner Silva[a]
[a] Federal University of Bahia, Salvador, Brazil

Abstract: The average real loss index of the Water Supply System in Salvador City, Bahia is 37.5%. Understanding the importance, complexity and factors that can influence the behavior of distribution networks in a water supply system, the use of tool such as survival analysis is appropriate for failure investigation and its recurrence. The study presents a survival analysis model and an Ishikawa Diagram of a Water Distribution Network of the city of Salvador, Bahia, according to the database for requesting services for the period from 2008 to 2014. The results of survival analysis in conjunction with the discussion of the Ishikawa Diagram generate a reliability diagnosis of the Water Distribution Network. This approach enables a greater sensitivity about the system in operation, serve as an auxiliary tool for decision making to minimize the occurrence of failures, real losses, and can facilitate the improvement of response times to repair leaks, providing reduction costs for operation and maintenance.

Keywords: Survival Analysis, Ishikawa Diagram, Water Distribution Networks, Real Losses.

1. INTRODUCTION

Equivalent to any large-scale infrastructure, water supply systems require constant maintenance and rehabilitation, allied, when necessary, to the expansion of water distribution networks to avoid multiple failures in case of extreme events and to meet future demands [1]. However, in a system of water supply with severe losses, it is ideal that the responsible authorities act more actively on the maintenance management and plan or make alternative policy to better to characterize the situation and to foment new solutions in order to mitigate the situation, instead of the isolated practice of corrective maintenance.

The deficiency of studies related to the probability of occurrence of events weakens the decision-making processes of managers, since processes such as risk assessment are important, through structuring the process of identification, prioritization and, later, strategy formulation of effective maintenance [2].

Given the importance, complexity, and factors that can influence the behavior of water distribution networks in a water supply system, the ideal management and maintenance strategy must be global and sustainable. The development of strategies should consider the risk analyzed, and be composed of two pillars of equal importance, which are efficient methods to monitor, repair or replace infrastructures, and effective tools to model the deterioration of the network and predict failures [3].

The present paper presents a proposal for a model for survival analysis and an Ishikawa Diagram of a Water Distribution Network of a District Metered Areas (DMA) in the Salvador City, Bahia, based on data and maintenance service request information.

2. CASE STUDY AND DATASET

A significant amount of the population and municipalities of the State of Bahia is inserted in a region with low rainfall rates and few rivers that stand out, among these rivers, the Paraguaçu River [4].

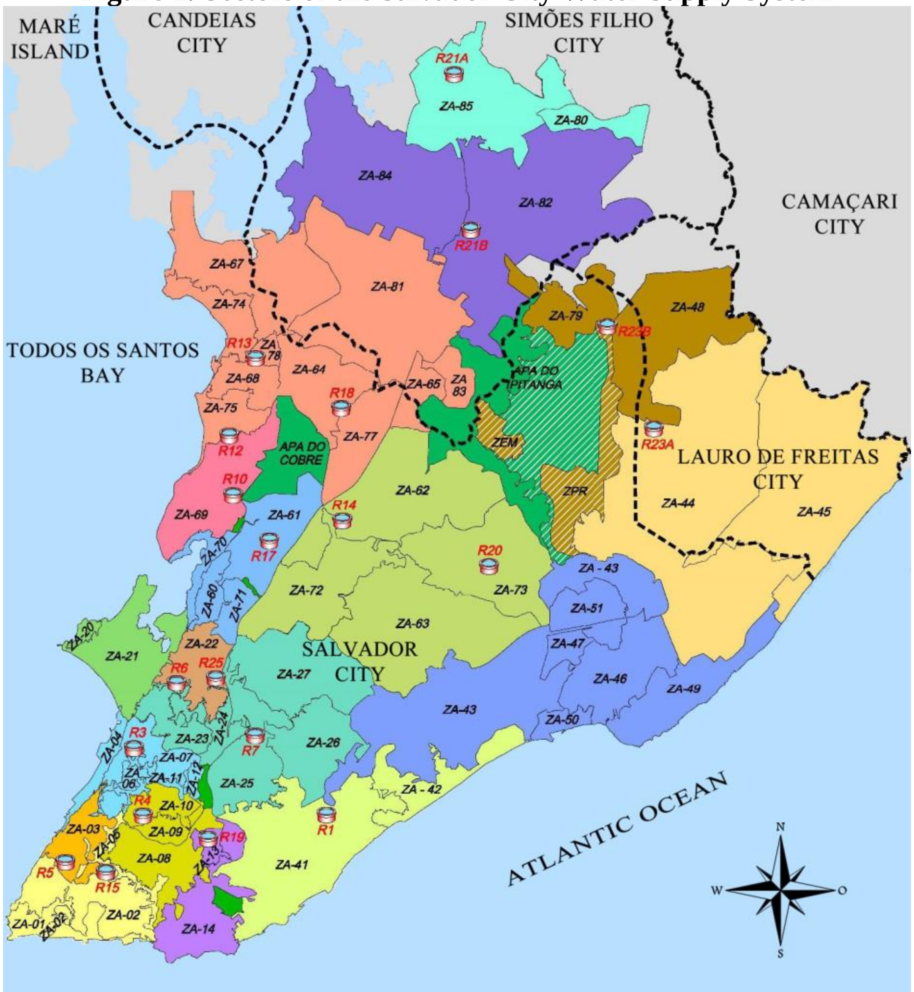
[*] lucasmascalmeida@gmail.com

Located in another hydrographic basin and with a milder climate, the Metropolitan Region of Salvador has an estimated population of about 3.8 million people. However, the Paraguaçu River, through transposition, is responsible for 60% of this water supply [5, 6].

The water is catchment, treated, stored and distributed, and according to the use and billing (billed or unbilled) by the Water Companies it is possible to classify the consumption. In the Salvador water supply system, it is estimated that the average loss index for non-revenue is 50%, with 37.5% referring to real losses (due to leaks in the distribution network and reservoirs) [5]. The metrics elucidate negative aspects as low efficiency, quality and reliability of the system, reduction of water resource availability, waste of energy due to water transportation, economic impact, among other issues [7].

The Water Company responsible for the management and operation of the water supply service of Salvador sectored the city in four big regions and divided into fifty-one DMAs, as shown in Figure 1, according to the presence of flow meter and the pressure/flow control in the distribution network. The Water Supply System of Salvador has the current demand for water supply for human consumption estimated at 10.8 m³/s, however, 4.05 m³/s are classified as real loss [5].

Figure 1: Sectors of the Salvador City Water Supply System



Source: Adapted from Bahia, 2015.

Historical data of service request of one of the DMAs of the water supply system of Salvador, from January 2008 to December 2014, totaling 12247 service calls was analyzed. Of these calls, 7149 events were analyzed because they were classified as a source of actual loss in the water distribution network. Each call is identified with request protocol, service request number, request time, date and time of execution, service discrimination, service time (from service request to completion), address, neighborhood, zone, location leakage and pertinent observations. Despite the existence of a large amount of information, not all information is complete or is presented in a partial way (eg service discrimination and leak location), and the uncertainty associated with the address is large, making it impossible to locate the fault exactly DMA.

It was considered that the object of this study is the DMA, failure is the occurrence of actual water loss in the system; it was not considered a real loss in relation to situations other than the information in the service worksheet.

The computational modeling was made using software R version 3.4.3 coupled in RStudio software version 1.1.423 being developed from the survival and ggplot2 packages.

3. METODOLOGY

Directed at developing a diagnosis and reliability model of the individual of interest, the survival analysis tool and the Ishikawa Diagram to verify cause and effect relationships were used.

3.1 Survival Analysis with Censored Data

Survival analysis methods are generally to investigate survival data for one or more individuals, such as student dropout data for a given undergraduate course or to investigate the failure of equipment such as the pump. The results of a survival analysis allow analysis of failure time patterns, comparison of groups of individuals and their respective survival functions, and diagnostic of some risk factors that trigger an event of interest [8].

Usually system failures are studied element by element or element set. The leaks in pipelines and network accessories provide relevant water leaks, analysis of this set in a water distribution network is performed due to lack of accuracy as to the location of the fault and or state of the pipeline, which are usually old and tend to fail with greater ease [9, 10]. However, to investigate the behavior of the water distribution network in relation to the real water losses, the individual to be analyzed is the network itself in the DMA. The fault event is defined as any real loss that generated that occurred during the assessed period; not mattering of the volume of water, number of economies affected or even the cause. For example, a maintenance service in a situation of unbilled unmetered consumption, such as the substitution of hydrometer, but with high water wastage, was considered a failure in the network.

The main feature for survival data is the long period of data collection and information, and the presence of censorship, which for some reason monitoring the failure of the study individual has generated incomplete or partial information, and can be categorized as censored to the right , censored the left and or censored at intervals of time [11].

In field investigations through technical visits to the Water Company, it was found that on some occasions the night service requests were only counted on the following day during the morning shift, this procedure discharges the exact time of failure and consequently assigns a characteristics of the morning shift events. As a consequence, it was considered that all morning events have interval censorship, the time to the left of the interval is the time of the last failure of the previous day and the time to the right is the time in the database. For the afternoon and evening events, the exact time of failure was considered.

3.1.1 Turnbull's Nonparametric Estimator for Interval-Censored Data

When there are difficulties in interpreting the survival analysis data, non-parametric statistics are conventionally used to estimate the survival function of the individual (Time versus Probability of failure), as the non-parametric maximum likelihood estimators [12], and the example of the Kaplan-Meier’s [13] and Turnbull’s estimators [14].

The method of analysis of survival times using the Kaplan-Meier non-parametric estimator is one of the possible algorithms for estimating the survival function for right censored data [13]. In the presence of interval censorship it is common to use this same estimator considering the mean of the extremes of the censorship interval as observed time, thus reducing the problem for a right censorship case only, and then using only the Kaplan-Meier estimator [15].

The product-bound estimator, Turnbull's maximum likelihood non-parametric estimator for the survival function of a data set in the presence of interval censorship [15]. The mathematical development of the algorithm will not be presented, only a brief description of the interactive procedure to find the solution will be addressed, because the purpose of this work is in the analysis for this case study, not in the development of the method. For more details about this method, it is recommended to read the Turnbull's original article [14].

I. Obtain an initial survival estimate, for example, using the Kaplan-Meier estimator;

II. Calculate the expected probability and with this the proportion of observations of the interval;

III. Update the survival estimates equaling to the ratio of interval observations;

IV. Repeat step II using the updated estimates;

V. Stop when to reach the convergence criterion.

3.1.2 Adjust and Identification of Parametric Model

The non-parametric model is limited, it does not allow a complete survival diagnostic of the individual, the main survival functions such as, failure rate and density are not products of this type of estimation [12]. The non-parametric modeling is done, the next step is to estimate parameters and select the model that best describes the behavior of the system. The selection of the model is do through graphic analysis and the Akaike Information Criterion (AIC) test.

The results of the non-parametric modeling required parameter estimation and model selection to better describe the behavior of the system. The selection of the model took into consideration graphic analyzes and the Akaike Information Criterion (AIC) Test.

The first graphical method applied consisted of comparing the parametric survival function with the Turnbull's Estimator, graphically comparing the estimated survival functions for each proposed model. The most appropriate model is that the estimated survival points are closer to the values obtained by non-parametric estimation. The best points adjusted to the abscissa equal to ordinate, where x is the non-parametric survival and y is the parametric survival.

Another graphical method used to evaluate the models was the comparison of the graphs of the cumulative distribution Functions of each model versus Time, and the Graph of cumulative function non-parametric versus Time, the objective of this second graph is also to evaluate which model better describes the non-parametric curve.

Hypothesis testing procedures are not always adequate procedures for the statistical identification of the model. Based on the classical procedure of maximum likelihood estimation and aiming at a statistical identification of models, a new estimation of the test of theoretical criterion of information is presented. The maximum likelihood estimates of the parameters that give the minimum AIC are defined by [16]:

$$\text{AIC} = (-2)^{*} \log(\text{maximum likelihood}) + 2^{*}(\text{number of parameters}) \tag{1}$$

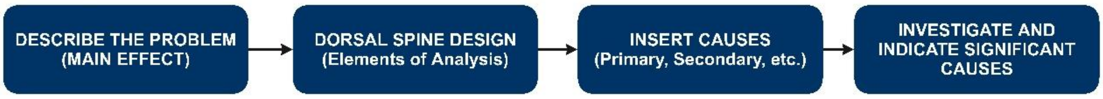
A good model is one that has minimal AIC among all other models. The AIC test was used to evaluate the parameters of the models that were pre-selected in the graphic evaluations.

3.2 Ishikawa Diagram

The Ishikawa Diagram is an easy to apply method of analysis of cause and effect that allows determining the global risk of an event with multiple relevant causes. The application makes it possible to determine the risk of primary and secondary causes of the cause categories and the global risk. It enables the structuring of decision-making in areas of vulnerability, precisely oriented on the causes that determine high-risk values. There are no instruments for risk analysis based exclusively on the Ishikawa Diagram, but there are instruments that include elements of primary or complementary analysis of this type [17].

Based on information from the literature [9, 10], information collected from the system maintainer and the use of the brainstorming technique of the authors of this work, the potential causes of failures are grouped by affinity in six elements, known as the 6Ms: environment, method, measurement, machine, labor and raw material [18]. The diagram is currently used in a very flexible way, it is used presenting elements different from the original form proposed by Ishikawa in terms, groups (being more or less than the 6Ms) and nomenclatures, this characteristic increased its application in the area of analysis of risk. At Figure 2, show the steps for constructing the Ishikawa Diagram.

Figure 2: Steps to Create the Ishikawa Diagram



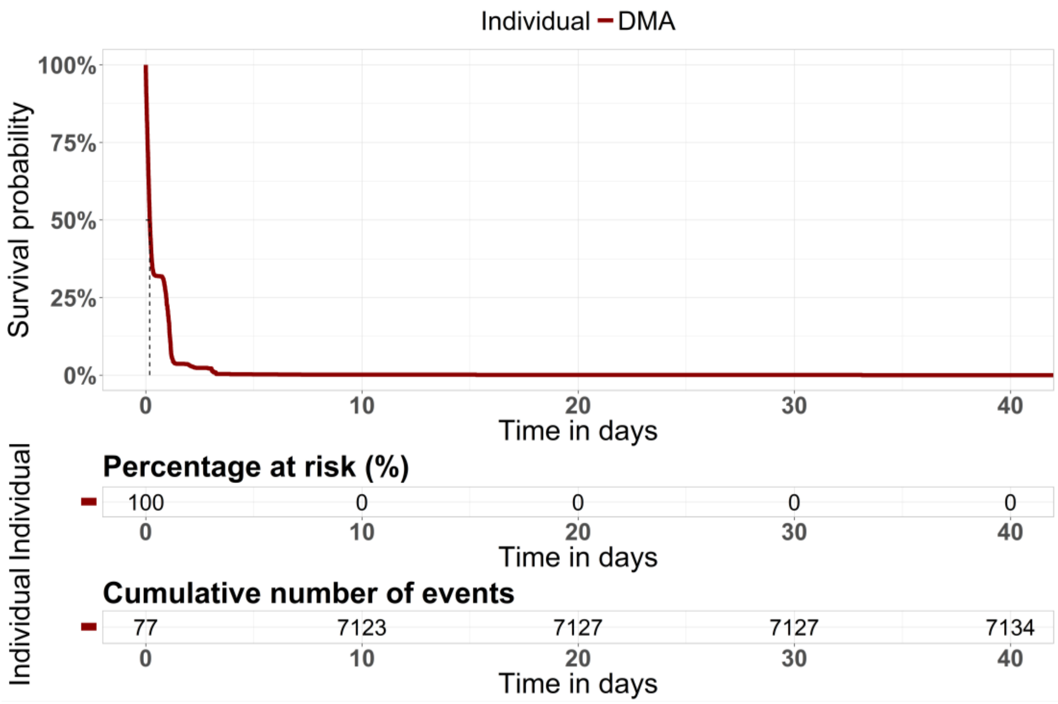
The Ishikawa Diagram was produced in the study presented with the purpose of complementing the diagnosis of the individual of interest. From the results of the diagram, it was possible to base the qualitative behavior of the individual in relation to the main effect.

4. ANLYSIS, RESULTS AND DISCUSSION

4.1 Mathematical Modelling

In the curve shown in Figure 3 is possible to perceive a large concentration of events belonging to the initial interval of failure time. The percent-at-risk and the cumulative number of events complement the behavioral representation of survival data, these numbers represent quantities that illustrate this behavior at the beginning of the curve, and it is possible to extract from these values an indication of the necessity of amplifying the amplitude of the failure time of the study.

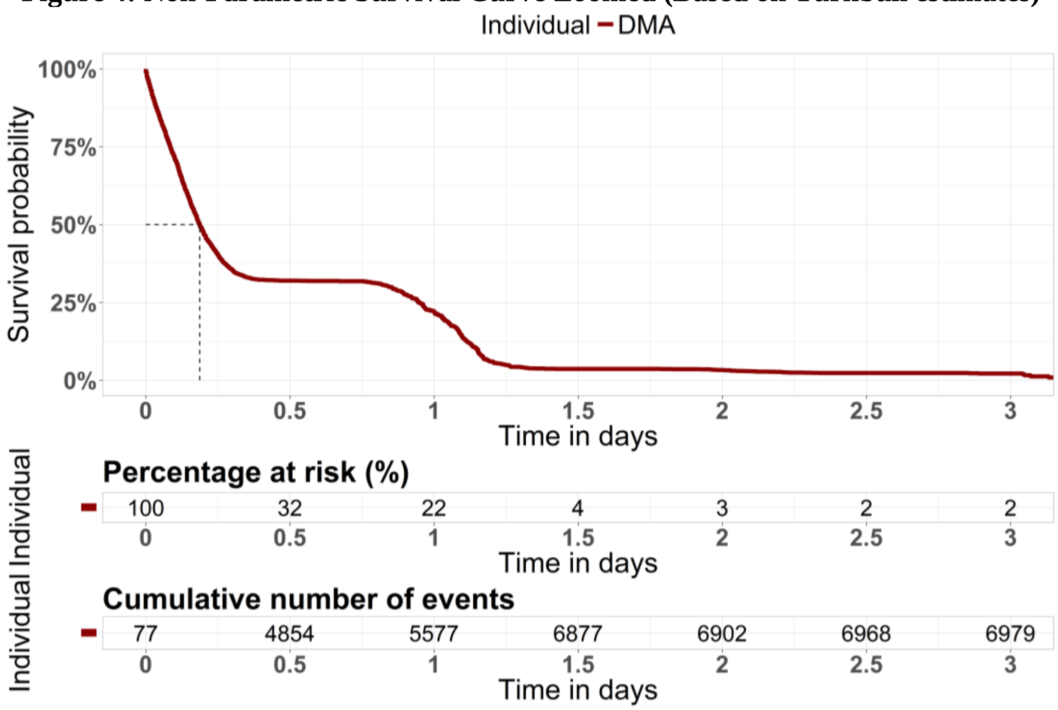
Figure 3: Non-Parametric Survival Curve (Based on Turnbull estimates)



For a better use of the information that the non-parametric estimation offers, The Figure 4 brings a zoomed on the initial failure time of the survival curve, from time 0 to time 3.5 days.

Observing Figure 4 it is possible to observe that more than 50% of the study events occur with a recurrence time of less than 0.5 days. Bringing this interpretation to the object of this study and complemented with the information of 32% of percent-at-risk in time 0.5 days, this non-parametric estimation indicates that 78% of times between failures are concentrated at intervals of amplitudes less than 12 hours.

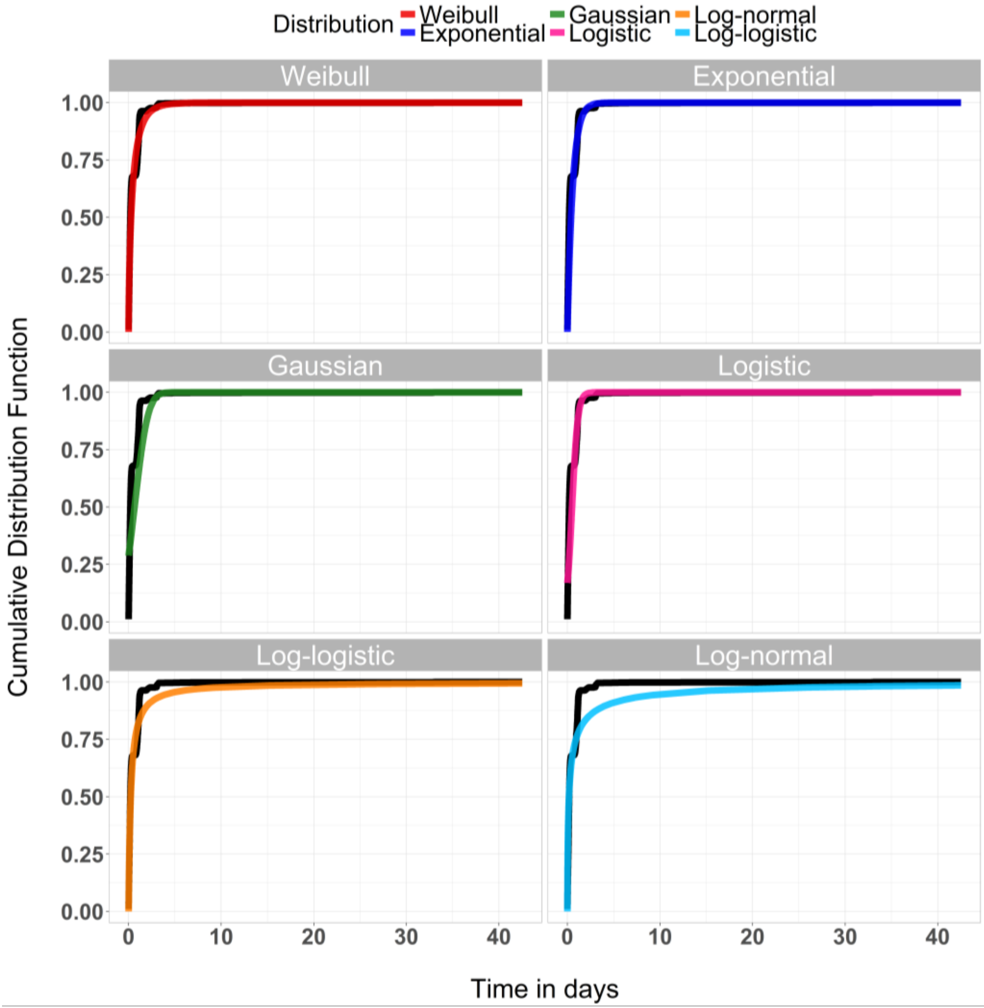
Figure 4: Non-Parametric Survival Curve Zoomed (Based on Turnbull estimates)



Understanding that the district of the case study is part of a larger system in size and complexity and that this can lead to behaviors that do not match the most common probabilistic models, this case study considers as possible probabilistic models the Weibull, Exponential, Gaussian, Logistic, Log-Normal and Log-Logistic.

Trying to evaluate which distribution best describes the data behavior was used graphic methods and later to test the quality of the model was used the AIC test. For the graphical evaluate was choosing the time versus cumulative distribution function curve.

Figure 5: Graphical Evaluate (Cumulative Distribution Function)



The application of this graphical test (Figure 5) for evaluation of fit models present Weibull and Exponential as the best distributions of probability, the results AIC test agrees this graphical evaluation.

Table 1: Parameters Values and AIC Test for the Models

Statistics	Weibull	Exponential
Scale (α)	0.4427051	0.5680257
Shape (γ)	0.7025420	-
AIC Test	6917.1186	8444.8342

The results shown in Table 1 indicate that the Weibull model present the lowest value of the AIC statistic, showing that Weibull is the best model.

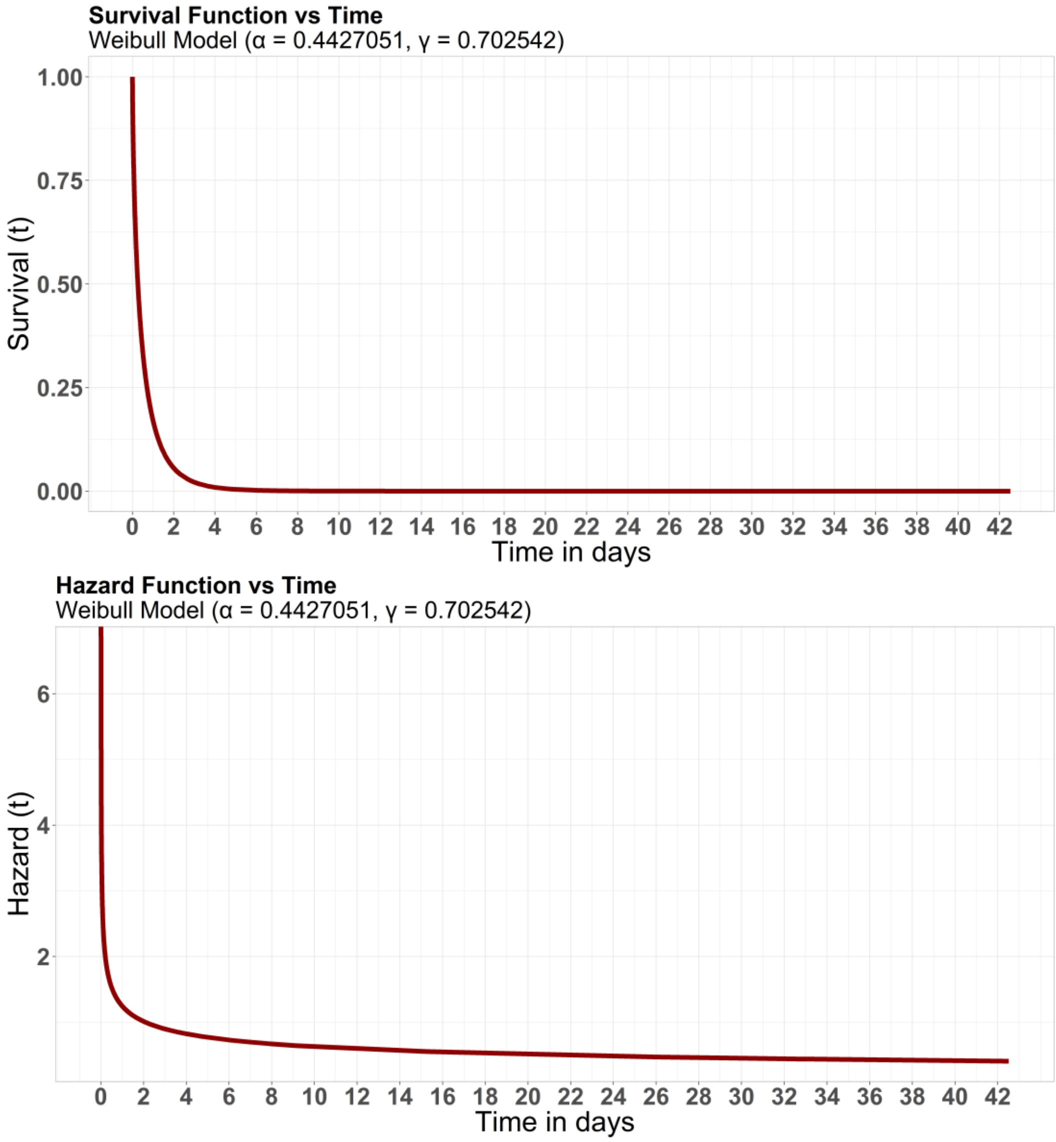
4.2 Survival Analysis Results

Equation 2 shows the mathematical formulation of the hazard function for Weibull distribution.

$$h(t)= \frac{0.7025420}{0.4427051}\left(\frac{t}{0.4427051}\right)^{(0.7025420 - 1)}, t > 0 \tag{2}$$

Figure 6 shows the graphs of the survival functions and the hazard function of the model. Survival and hazard function graphs indicate that most fault recurrences are at the beginning of the curves.

Figure 6: Model's Graphics (Survival Function and Hazard Function)



The survival percentage of the system and the risk decrease strongly at the beginning of the curve. These graphs indicate that there is a high risk when the time between failures is between 0 and 5 days. Considering the response time for fault repair, this peak period represents that the real losses of the system are almost uninterrupted, cases of more than 5 days without losses are rare.

The parametric and non-parametric estimation methods confirm that the risk of failure is very intense for times between low amplitude faults.

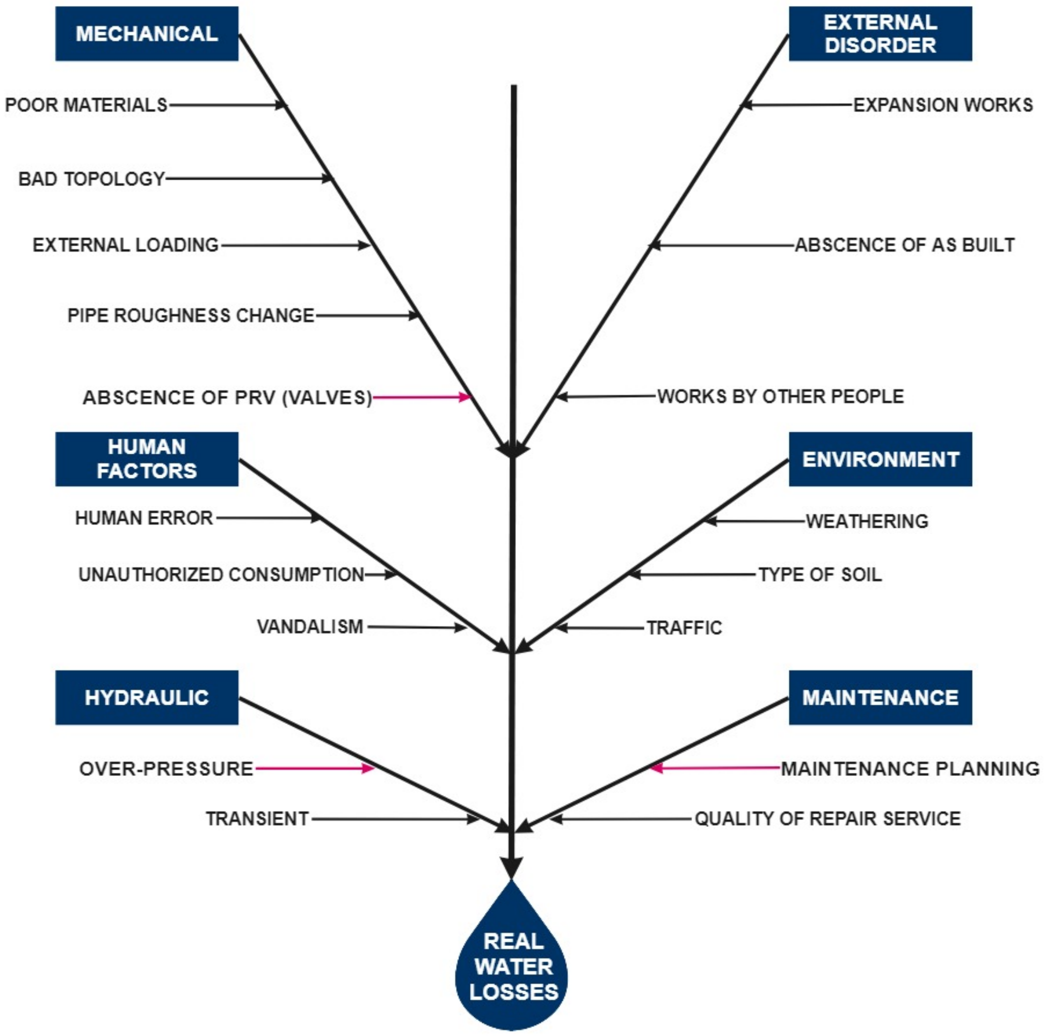
4.3 Ishikawa Diagram Discussion

An Ishikawa Diagram was adding to the system reliability diagnosis, considering that the definition of the main effect is the same as the system failure (Figure 7). The information collected is from three sources, literature, database and technical visit to the Water Company. It presents potential elements that can cause a crash or affect system performance.

Several failure modes result in real water losses from a water supply system. The purpose of this Ishikawa Diagram is to present a visual summary of the principal failure modes and their causes. For the study in question, the diagram shows only one level of potential causes.

The elements of the Ishikawa Diagram illustrated in Figure 7 are mechanical, hydraulic, human factors, environmental, maintenance and external disorder. Among failure modes, it is possible to highlight the high pressure, the insufficiency of pressure regulating valves and the maintenance-planning model.

Figure 7: Ishikawa Driagram for Real Water Losses



Maintenance planning and management, insufficient pressure regulating valves and high pressure are modes of failure that can coexist in various events, and the presence of these failure modes may explain the high hazard for times between low amplitude faults. These failure modes are generally not detected at the failure locate because they are intrinsic characteristics of the management and the supply system in question. For instance, in the moment an event occurs because of the rupture of a tube the repair applied to the fault is strictly a physical repair in the tube, but the characteristic of the system of working with high pressures is not solved, so the repaired tube may fail again for the same cause of rupture shortly after repair.

5. CONCLUSIONS

In this paper, the survival analysis for a water supply network in a Water Supply District of the Salvador city, Bahia, was carrying on the basis period from 2008 to 2014 of historical failure data. The approach of the analysis considering the recurrence of time between failures pointed to a high failure rate for times between low amplitude faults, from 0 to 5 days, this index will can interpreted as a need to review management and maintenance planning procedures of the DMA under study..

The proposed Ishikawa Diagram to complement the quantitative analysis of survival data with a qualitative discussion of the individual presents the need for investigation of failure modes due to high pressure, insufficiency of pressure regulating valves and maintenance planning and indication of investments in the repair of these failure modes. Understanding that this type of approach will minimize the incidence of failures for times between low amplitude faults.

The results of the survival analysis in conjunction with the discussion of the Ishikawa Diagram allow the reliability diagnosis of the Water Distribution Network. This type of approach presents an indicative that can assist in decision making that minimizes the occurrence of the failure, will can improve leakage response times, reducing operation and maintenance costs for water services.

Acknowledgements

The present work was supported by the “Fundação de Baiana de Amparo à Pesquisa” (FAPESB), the “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES), the “Programa de Pós-Graduação em Engenharia Industrial” (PEI) and the GAMMA (Growing with Applied Multivariate Analysis) group of the Federal University of Bahia.

References

[1] D. G. Yoo, D. Kang, H. Jun, and J. H. Kim, “*Rehabilitation priority determination of water pipes based on hydraulic importance*”, Water (Switzerland), vol. 6, pp. 3864–3887, (2014).
[2] P. Chemweno, L. Pintelon, P. N. Muchiri, and A. Van Horenbeek, “*Risk assessment methodologies in maintenance decision making: A review of dependability modelling approaches*”, Reliab. Eng. Syst. Saf., vol. 173, pp. 64–77, (2018).
[3] S. E. Christodoulou, “*Water Network Assessment and Reliability Analysis by Use of Survival Analysis*”, Water Resour. Manag., vol. 25, pp. 1229–1238, (2011).
[4] M. Schaer-Barbosa, M. E. P. dos Santos, and Y. D. P. Medeiros, “*Viabilidade do reúso de água como elemento mitigador dos efeitos da seca no semiárido da Bahia*”, Ambient. Soc., vol. 17, pp. 17–32, (2014).
[5] Bahia. Secretária de Infraestrutura Hídrica e Saneamento (SIHS). “*Relatório Sinopse*”, in Plano de Abastecimento de Água da Região Metropolitana de Salvador, Santo Amaro e Saubara, vol. 1, 2015, Salvador.
[6] M. F. C. de Jesus and J. O. R. de Morais, “*Políticas Públicas Para Revitalização de Bacias Hidrográficas*”, in Semeando Águas no Paraguaçu, I. R. Lamas, L. S. Rita, and R. M. Miranda, Eds. Rio de Janeiro: Conservação Internacional, 2016, Rio de Janeiro.
[7] M. R. N. Vilanova, P. Magalhães Filho, and J. A. P. Balestieri, “*Performance measurement and indicators for water supply management: Review and international cases*”, Renew. Sustain. Energy Rev., vol. 43, pp. 1–12, (2015).
[8] C. Kartsonaki, “*Survival analysis*”, Diagnostic Histopathol., vol. 22, pp. 263–270, (2016).
[9] M. Güngör, U. Yarar, and M. Firat, “*Reduction of water losses by rehabilitation of water distribution network*”, Environ. Monit. Assess, vol. 189, pp. 498, (2017).
[10] A. Gheisi, M. Forsyth, G. Naser, and M. Asce, “*Water Distribution Systems Reliability : A Review of Research Literature*”, J. Water Resour. Plan. Manag., vol. 142, (2016).

[11] E. A. Colosimo and S. R. Giolo, “*Análise de sobrevivência aplicada*”, Blücher, 2006, São Paulo.
[12] K. Bogaerts, A. Komarek, and E. Lesaffre, “*Survival analysis with interval-censored data: A practical approach with examples in R, SAS, and BUGS*”. CRC Press - Taylor & Francis Group, 2017, Boca Raton.
[13] E. L. Kaplan and P. Meier, “*Nonparametric estimation from incomplete observations*”, J. Am. Stat. Assoc., vol. 53, pp. 457–481, (1958).
[14] B. W. Turnbull, “*The empirical distribution fuction with arbitrarily grouped, censored and truncated data*”, J. R. Statical, vol. 38, pp. 290–295, (1976).
[15] J. C. B. COSTA, “*Modelos semiparamétricos de fração de cura para dados com censura intervalar*”, Universidade de São Paulo, USP, Brasil, 2016, São Paulo.
[16] H. Akaike, “*A New Look at the Statistical Model Identification*”, IEEE Trans. Automat. Contr., vol. 19, pp. 716–723, (1974).
[17] I. G. Ciocoiu, G. Ilie, and C. N. Ciocoiu, “*Application of Fishbone Diagram To Determine the Risk of an Event With Multiple Causes*”, Manag. Res. Pract. vol. 2, pp. 1–20, (2010).
[18] A. V. M. De Luna, “*Gestão e Melhoria de Processos em uma Indústria Farmacêutica Pública: Estudo de Caso do Processo de Fabricação de Comprimidos Revestidos*.”, Universidade Federal Fluminense, UFF, Brasil, 2013, Rio de Janeiro.