

Document downloaded from:

<http://hdl.handle.net/10251/108479>

This paper must be cited as:

Debón Aucejo, AM.; Carrión García, A.; Cabrera Marcet, E.; Solano, H. (2010). Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering & System Safety*. 95(1):43-48. doi:10.1016/j.ress.2009.07.004



The final publication is available at

<https://doi.org/10.1016/j.ress.2009.07.004>

Copyright Elsevier

Additional Information

# Comparing Risk of Failure Models in Water Supply Networks using ROC Curves.<sup>\*</sup>

A. Debón<sup>a,\*</sup> A. Carrión<sup>a</sup>

<sup>a</sup>*Centro de Gestión de la Calidad y del Cambio  
Dpt. Estadística e Investigación Operativa Aplicadas y Calidad.  
Universidad Politécnica de Valencia.  
E-46022. Valencia. Spain.*

E. Cabrera<sup>b</sup>

<sup>b</sup>*Dpto. De Ingeniería Hidráulica Y Medio Ambiente  
Instituto Tecnológico del Agua  
Universidad Politécnica de Valencia.  
E-46022. Valencia. Spain.*

H. Solano<sup>c</sup>

<sup>c</sup>*Universidad Diego Portales  
Santiago. Chile*

---

## Abstract

The problem of predicting the failure of water mains has been considered from different perspectives and using several methodologies in engineering literature. Nowadays, it is important to be able to accurately calculate the failure probabilities of pipes over time, since water company profits and service quality for citizens depend on pipe survival; forecasting pipe failures could have important economic and social implications. Quantitative tools (such as managerial or statistical indicators and reliable databases) are required in order to assess the current and future state of networks. Companies managing these networks are trying to establish models for evaluating the risk of failure in order to develop a proactive approach to the renewal process, instead of using traditional reactive pipe substitution schemes.

The main objective of this paper is to compare models for evaluating the risk of failure in water supply networks. Using real data from a water supply company, this study has identified which network characteristics affect the risk of failure and which models better fit data to predict service breakdown.

The comparison using the Receiver Operating Characteristics (ROC) graph leads us to the conclusion that the best model is a Generalised Linear Model. Also, we propose a procedure that can be applied to a pipe failure database, allowing the most appropriate decision rule to be chosen.

## 1 Introduction

Worldwide, water supply systems (WSS) face the problem of aging infrastructures and increasing maintenance costs. The classic reactive approach used by most companies is obviously not the best way of managing this essential public service, from both quality and availability optics, proactive strategies are required. However, proactive approaches require information and models to evaluate risks, predict the best measures to take and to forecast water supply network performance. The need for proactive models is even greater in developing countries, with tight economic restrictions, than in advanced countries. Authors such as Saegrov [18] conclude by identifying a knowledge gap and research possibilities, mainly relating to data collection and how best to use existing data for the development and calibration of predictive deterioration models, risk assesment methods, etc. In this study we will present an analysis of the reliability data from the water supply network of a medium-sized city on the Spanish Mediterranean coast. Three different models will be used to analyze data, to identify the main factors affecting pipe failure and to predict the risk of failure.

Fenner [11] concludes that to be cost effective, proactive maintenance involving inspection and repair must be focused on those pipes which can be shown to have an early predisposition to failure. In line with that we suggest some statistical tools which can optimise and prioritise any proactive work. The main objective of this study is to promote the use of quantitative tools in the management of water supply systems, to evaluate their present state and to forecast the future deterioration of infrastructures.

The contents of this article are structured as follows. In the second section we present the three survival models we have applied: the Cox Proportional Hazard model, Accelerate Failure models and the Generalized Linear models. We also introduce a Receiver Operating Characteristics (ROC) graph as a

---

\* The authors are indebted to the anonymous referees whose suggestions improved the original manuscript. This study was partially supported by a grant from MEyC (Ministerio de Educación y Ciencia, Spain, project CGL2004-05507). The research of Ana Debón was partially supported by a grant from Generalitat Valenciana (grant No. GVPRE/2008/103).

\* Corresponding author. Tel: +34 963877007 (Ext. 74961); Fax: +34 963877499.  
E-mail: andeau@eio.upv.es

way to evaluate concordance between models and real data. The third section starts describing data from the water supply network studied. In this third section, we apply the three different models to these data, commenting on their advantages and disadvantages, as well as on their suitability in the failure risk analysis in question. The section ends with a comparison of two different fits by means of the ROC curves. Finally, we will come to some conclusions in which we try to assess their validity not only for this case but also, to some extent, for any other WSS.

## 2 Survival Analysis

Statistical models of differing complexity have been suggested in the literature for predicting pipe breaks in water distribution systems, from Proportional Hazard models to Generalised Linear Models (GLM). A review of these statistical models can be found in Kleiner and Rajani [13]. The most recent contributions have also been compiled by Yamijala et al. [21] with a special emphasis on comparing the accuracy and usefulness of these models by means of goodness of fit statistics. These models were designed to show the impact of each predictive variable for the risk of failure of a individual pipe.

### 2.1 Cox model

Cox regression, also called Proportional Hazard model or Duration model, is designed to analyze the time lapse until an event occurs or the time lapse between events. One or more predictor variables, called covariables, are used to predict a status (event). The Proportional Hazard model [6] has been widely used in analyzing survival data.

The model specifies that the survival time  $T$ , given the covariate vector  $x$ , has the hazard function

$$h(t; x) = h_0(t)exp(\beta'x) \tag{1}$$

where  $h_0(t)$  is an unspecified baseline function and  $\beta$  is the regression coefficient vector associated with  $x$ . The common definition of hazard rate in Survival Analysis is the probability that an individual assumed to be alive until instant  $t$  will not survive the following limited time interval,

$$h(t) = \lim_{\Delta t} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \tag{2}$$

The interest is mainly in the proportional factors rather than the baseline hazard. This model is semi-parametric because while the baseline hazard can take any form, the exponent is linear in the covariables. Remarkably, even though the baseline hazard is unspecified, the Cox model can still be estimated by the method of partial likelihood, developed by [6] in the same paper in which he introduced the Cox model.

Similar methods were used by Eisenbeis [8,9] applying Cox's Proportional Hazard model for four water pipe networks in France. To describe the early deterioration behavior, [5,3,2] used a Cox Proportional Hazard Model based on US data.

## 2.2 Accelerated lifetime models

Accelerated lifetime models are designed to analyze what influence the covariables have on the failure time  $T$ , whose expression is,

$$\ln(T) = \mu + x'\beta + \sigma Z. \quad (3)$$

where  $Z$  is the random variable which describes the random behavior of  $\ln(T)$ . The accelerate lifetime model can be viewed as analogous to the Proportional Hazard model. It differs from the Cox model in that the covariates act to the time to failure, whereas failure hazard is affected in the Cox model [13]. There are different types of Accelerate Lifetime models. The simplest parametric model is the exponential one which assumes a constant risk function,

$$h(t) = \lambda, \quad 0 \leq t \leq \infty.$$

However, in the majority of phenomena of interest the hypothesis that the risk function is constant is overly restrictive. The Weibull distribution defines a more general model whose risk function is,

$$h(t) = \lambda^\alpha \alpha t^{\alpha-1}, \quad 0 \leq t \leq \infty,$$

where the parameters  $\lambda$  and  $\alpha$ , denominated scale and shape parameters respectively, are positive values. The Accelerate Lifetime model translates into Proportional Hazard when  $Z$  has a Weibull distribution [7].

It is possible to specify other distribution for the variable  $Y = \ln(T)$ , which have any real value. One possibility is to consider that  $\ln(T)$  has a Normal distribution with average  $\mu$  and variance  $\sigma^2$ , which means that  $T$  follows a Lognormal distribution. In this case  $h(0) = 0$ , grows until reaching a maximum

and subsequently decreases again towards 0; These characteristics hardly make it realistic.

One distribution obtained with the same procedure and which, different to the Lognormal, has relatively simple expressions of the basic functions is the Log-logistic distribution. We can say that  $T$  has a Log-logistic distribution if the variable  $\ln(T)$  has a logistic distribution, which is symmetrical distribution similar to the Normal standard, except at the tails.

Lei [14] applied both the Cox model and the Accelerated Lifetime models to the distribution system of Trodheim, Norway. This study did not say whether the model was validated or comment on the quality of the predictions. Eisenbeis [9] applies the Weibull model to the same data that had applied in the Cox model.

### 2.3 Generalised Linear Models

Generalised Linear Models (GLM) are an extension of linear models for non-normal distributions of the response variable and non-linear transformations. A regression model constitutes a specification for the variable mean,  $m$ , in terms of a small number of unknown parameters  $\beta_0, \beta_1, \dots, \beta_p$  corresponding to the covariables. In the particular case of linear models, we wish to find a linear function such that  $E(Y|X) = m = \beta_0 + \sum_{i=1}^p \beta_i x_i$ , with the variance of  $Y$  supposed to be constant,  $var(Y) = \sigma^2$ .

In a different way, a generalised linear model (GLM) provides a method for estimating a function of the average of the response variable as a linear combination of the set of predictive variables, that is

$$l(E(Y|X)) = l(m) = \eta(x) = \beta_0 + \sum_{i=1}^p \beta_i x_i.$$

The function of the response average,  $l(m)$ , is called function *link*, and is considered to be the same as a linear function of the predictors,  $\eta(x)$ , which is called *linear predictor*. Each component  $y_i$  of  $Y$  has a Binomial, Poisson or Gamma distribution. The GLM comprehensive reference is [16]. GLM with qualitative predictors are well described in [1].

In the field of Reliability these GLM have been used before by [4] and [21].

A Receiver Operating Characteristic (ROC) graph is a technique for visualizing, organizing and selecting classifiers based on their performance [10]. ROC graphs are commonly used in medical decision making, and in recent years have been used increasingly in machine learning and data mining research. The purpose of this paper is to use it in research into pipe failure analysis. We consider problems where the items can only belong to two classes and some classification models (or classifiers) that produce a continuous output (e.g., an estimate of hazard rate) to which different thresholds may be applied to predict class. For each individual we have both the model prediction and the actual class. Given a classifier and an instance, there are four possible outcomes, Table 1 showing the possibilities

Table 1  
Results in table form

		True class	
		A	B
Predicted class	A	True positives	False positives
	B	False negatives	True negatives
		Total positives	Total negatives

The true positive rate (tp rate) of a classifier is estimated as

$$\text{tp rate} = \frac{\text{True positives}}{\text{Total positives}}.$$

The false positive (fp rate) rate of a classifier is estimated as

$$\text{fp rate} = \frac{\text{False positives}}{\text{Total negatives}}.$$

Additional terms associated with ROC curves are

$$\text{sensitivity} = \text{tp rate}$$

and

$$\text{specificity} = 1 - \text{fp rate} = \frac{\text{True negatives}}{\text{Total negatives}}.$$

ROC graphs are two-dimensional graphs in which the tp rate is plotted on the Y axis and the fp rate is plotted on the X axis. A ROC graph depicts

relative tradeoffs between benefits (true positives) and cost (false positives). Several points in ROC curve are important to highlight. The lower point (0,0) represents the strategy of never issuing a positive classification; such a classifier commits no false errors but also gains no true positives. The opposite strategy, of unconditionally issuing positive classifications, is represented by the upper right point (1,1). The point (0,1) represents perfect classification. Informally, classifiers with a ROC curve closer to this point are preferred. The diagonal line  $y=x$  represents the strategy of randomly guessing a class. It is possible for a classifier to perform worse in a specific region than another classifier if its curves are cut.

The model will be good if we can discern the pipes that break and those which do not. The ROC curve provided an overall representation of the accuracy. If the test did not allow discrimination between classes, the ROC curve was the diagonal joining the vertices from lower left to upper right. The accuracy of the test increased as the curve moves from the diagonal towards the upper left corner. To evaluate the discriminative performance of the model and to compare classifiers, we wanted to reduce ROC performance to a single scalar of value representing expected performance. Calculating the area under the ROC curve of the classifier, for short AUC, was a common method. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1, the random guessing procedure having an area of 0.5. The greater the AUC, the better is the classifier.

Moreover, we can obtain the optimal operating point on the ROC curve, which is the point closest to the top-left corner. This gives the threshold value that will provide a good compromise between sensitivity and specificity. This threshold value of pipe failure risk can be selected to detect pipes which have a high risk of failure. There are other tools that allow classification of the broken and non-broken pipes like the discriminatory analysis. However, the ROC curves have the advantage of being able to decide about the breakage point or risk for the classification. They can be used where managers are more risk-averse or risk accepting than that which uses the “optimum” point.

### **3 Application to real data**

#### *3.1 Data*

In the development of the research project, we had access data from the water supply company of a medium-sized Spanish city. The water supply company gave us access to a database containing information on pipe sections making up the network. The database includes 32,387 entries, corresponding to each of

the sections. Among other variables, the entries contain: section identification, section diameter, pressure, installation year, pipe material, date of failure, section length, traffic conditions and type of failure. There were some problems with the quality of data: data corresponding to the oldest sections were not reliable; the failures have only been included in the data-base since 2000 (when the use of the GIS<sup>1</sup> was established), and there was no possibility of recovering older failure data. This means a very high censoring rate, up to 98%. Also, no consideration was given to the fact that a pipe section can fail more than once because the database structure was not prepared to consider this. Some minor errors had to be corrected prior to using the database, frequently meaning the loss of the corresponding failure entries. One of the major problems with the database was the lack of reliability of the data on the oldest pipe sections. Due to that lack of reliability of older data, only those pipes installed after 1940 were considered. According to the database, four different materials have been used: ductile cast iron, gray cast iron, polyethylene and asbestos cement. The database also records the traffic conditions of the installation area of the pipes considering three kinds of traffic: under sidewalk, normal traffic and heavy traffic.

Because failures occurring between 1941 and 1999 are unknown, and those between 2000 and 2006 are known, we consider that failures occurring before 1941 are not important following [15], who found left-censoring is a minimal problem. We need a further condition on the censoring para to apply the previous Cox model, specifically that the censure is independent and uninformative which assumes that the likelihood for censored observations does not depend on  $\beta$  [12].

Moreover, given that we only had the year of the failure, the times have been calculated discretely in number of years, and therefore failures equal to 0 pose a problem when the logarithm is applied to them, the problem being resolved by substituting them for  $1/365$ , on the assumption that the pipes had lasted at least one day.

### 3.2 *Cox model*

As is common in survival techniques, the estimation of the covariable effects (diameter, pressure, material, length) are analyzed on the hazard rate associated with pipe duration. We begin with the Proportional Hazard model, also called Cox Regression [6]. Its mathematical expression is in equation (1). By analogy, in this work, hazard rate (2) is the probability that a pipe, which

---

<sup>1</sup> A geographic information system (GIS), also known as a geospatial information system, is any system for capturing, storing, analyzing and managing data and associated attributes which are spatially referenced to Earth.

has not failed until instant  $t$ , would start breaking within the following time interval. Whereas in our case, the hazard function  $h(t)$  measures the hazard rate that a pipe may break, conditioned by the fact that it had not broken until the instant  $t$ . Moreover,  $x_i$  is the vector of covariates for the  $i$ th case. The main assumption is that the hazard rate in all cases is a multiple of an unspecified baseline hazard rate  $h_0(t)$ . The Cox proportional-hazards regression model was fitted in R with the `coxph` function using library `survival` [17]. Detailed arguments and estimations are available in [20].

The results of the Cox Regression are shown in Table 2, which includes the value of the coefficients for each of the covariables, the exponential of the coefficient values (which express the effect of the corresponding covariable on the hazard rate), and the standard error and significance for each of the coefficients. In Tables 2, 3 and 4 the values listed above the double line are model coefficients including all variables and below the double line are when the non-significant variables are omitted. The reference categories, heavy traffic and asbestos cement, for the qualitative variables, traffic and material, respectively, do not appear in those tables.

The meaningfulness of the positive coefficient  $\beta$  which corresponds to the length of the pipe can be interpreted thus: the failure hazard rate is higher in the longer pipes than in the shorter ones. Similarly, if we consider the pressure effect, this risk value increases with higher pressures. On the contrary, the meaningfulness of the negative coefficient which corresponds to the diameter of the pipe can be interpreted thus: the failure hazard rate is lower in the widest pipes. If we consider the traffic effect in isolation, according to our results the risk that a pipe will break is lower with normal traffic or under a sidewalk than with heavy traffic. If we consider the impact of material on the risk of failure, gray cast iron is not significant and the risk that a pipe will break is higher with ductile cast iron or polyethylene than with asbestos cement material.

From Table 2 each individual regression coefficient value in the second column below the double line can be interpreted in this way: for explanatory variable length, the increase in the risk of failure for an increase of 1 m is 0.4%; in the case of diameter the decrease in the risk of failure for an increase of 1 mm is 0.3%; with an increase of 1 in the preasure the risk of failure increase by 2.2%; under sidewalk and normal traffic decrease normal hazard rate by 50.6% and 42.0% respectively; the failure risk of asbestos cement pipes or gray cast iron pipes is the same but ductile cast iron increases the risk by 42.2% while polyethylene increases it by 471.4% if the rest of covariables remain fixed.

Table 2  
Cox Regression

Covariables	$\beta$	$\exp(\beta)$	$se(\beta)$	$sig.$
length	0.004	1.004	0.000	0.000
diameter	-0.003	0.997	0.001	0.003
pressure	0.023	1.023	0.005	0.000
traffic				
under a sidewalk	-0.726	0.483	0.264	0.006
normal traffic	-0.551	0.576	0.272	0.044
material				
ductile cast iron	0.331	1.392	0.146	0.023
gray cast iron	-0.134	0.875	0.184	0.473
polyethylene	1.728	5.631	0.266	0.000
length	0.003	1.004	0.000	0.000
diameter	-0.003	0.997	0.001	0.001
pressure	0.022	1.022	0.005	0.000
traffic				
under a sidewalk	-0.706	0.494	0.263	0.000
normal traffic	-0.545	0.580	0.272	0.045
material				
ductile cast iron	0.352	1.422	0.143	0.001
polyethylene	1.743	5.714	0.265	0.000

### 3.3 Weibull model

It is important to highlight that in the Accelerate Lifetime models the interpretation of the coefficients  $\beta$  is different from the Proportional Hazard model. The covariable effects (diameter, pressure, material, length) are analyzed on the failure time. We are going to apply the Weibull distribution on account of its good properties. Moreover,  $x_i$  is the vector of covariates for the  $i$ th case. The Weibull regression model was fitted in R with `survreg` function using library `survival` [17]. Detailed arguments and estimations are available in [20].

Results of the Weibull Regression are shown in Table 3. The table includes the value of the coefficients for each of the covariables, the exponential of the coefficient values, which express the effect of the corresponding covariable

on the failure time, and the standard error and significance for each of the coefficients. In Table 3 only gray cast iron material is non-significant as in the Cox model.

Table 3  
Weibull Regression

Covariables	$\beta$	$\exp(\beta)$	$se(\beta)$	$sig.$
intercept	5.241	188.95	0.1887	0.000
length	-0.002	0.998	0.0002	0.000
diameter	0.001	1.001	0.0005	0.002
pressure	-0.013	0.987	0.0026	0.000
traffic				
under a sidewalk	0.504	1.656	0.1432	0.000
normal traffic	0.363	1.437	0.1462	0.001
material				
ductile cast iron	0.227	1.255	0.0759	0.003
gray Cast iron	-0.088	0.916	0.0971	0.365
polyethylene	-0.598	0.550	0.1397	0.000
log(scale)	-0.627	0.534	0.0439	0.000
intercept	5.224	185.67	0.1871	0.000
length	-0.002	0.998	0.0002	0.000
diameter	0.001	1.001	0.0005	0.002
pressure	-0.013	0.987	0.0026	0.000
traffic				
under a sidewalk	0.521	1.684	0.1420	0.000
normal traffic	0.370	1.447	0.1458	0.001
material				
ductile cast iron	0.239	1.270	0.0748	0.003
polyethylene	-0.587	0.556	0.1388	0.000
log(scale)	-0.629	0.533	0.0438	0.000

The meaningfulness of the negative coefficient  $\beta$  which corresponds to the length of the pipe can be interpreted thus: the failure time is smaller in the longer pipes than in the shorter ones. Similarly, if we consider the pressure effect, this failure time value increases with higher pressures. On the contrary, the meaningfulness of the positive coefficient which corresponds to the diame-

ter of the pipe can be interpreted thus: the failure time is higher in the widest pipes. If we consider the traffic effect in isolation, according to our results the time that a pipe breaks is higher with normal traffic or under a sidewalk than with heavy traffic. If we consider the impact of material on failure time, the time when a pipe breaks is higher with ductile cast iron and smaller with polyethylene than with asbestos cement. In Gray cast iron is not significant.

From Table 3 each individual regression coefficient value can be interpreted in this way, for explanatory variable length, the decrease in failure time for an increase of 1 m is 0.2%, in the case of diameter the increase in failure time for an increase of 1 mm is 0.1%, with an increase of 1 in the preassure the failure time decrease by 1.3%, under a sidewalk and normal traffic increase time by 68.4% and 44.7% respectively, the failure time of an asbestos cement pipe or a gray cast iron pipe is the same but ductile cast iron increases the time in 27.0% while polyethylene decreases it by 44.4% if the rest of covariables remain fixed.

### 3.4 *Generalized linear models*

In order to check the consistency of these previous results, we use a different technique: the Generalised Linear Model analysis which allows us to analyze the data on the assumption that the number of pipe failures is Poisson. Our model is formally a Poisson Generalized Linear Model with a logarithmic link function as in the model used in [4], our model was fitted in R with `glm` function. Detailed arguments and estimations are available in [20]. Results of the Generalized Linear Model are shown in Table 4.

We can observe some differences comparing with Cox model, Weibull model and GLM, specifically, in material variable, where results for polyethylene are not significant. The significance of the remaining variables in the three models is the same.

As regards the values and sign of the coefficients, we can observe that they are similar in three models, despite the coefficients corresponding to materials. Three aspects deserve a special mention: the greater increase for two kinds of traffic, different sign in ductile cast iron material and the non influence of polyethylene in the risk. The differences in the models for polyethylene pipe material are especially striking. Polyethylene pipe is often marketed as being more resistant to breaks in certain circumstances. The results are due to the polyethylene material used in more recently installed pipes, as well as it being a less frequently used material in our database. This fact strengthens the decision to prefer GLM over Cox or Weibull.

Table 4  
GLM Regression

Covariables	$\beta$	$\exp(\beta)$	$se(\beta)$	$sig.$
Constant	-3.339	0.000	0.3919	0.000
Age	0.011	1.011	0.0058	0.053
length	0.003	1.003	0.0004	0.000
diameter	-0.003	0.997	0.0009	0.000
pressure	0.027	1.027	0.0045	0.000
traffic				
under a sidewalk	-1.156	0.000	0.2639	0.000
normal traffic	-0.787	0.455	0.2718	0.004
material				
ductile cast iron	-1.795	0.000	0.1805	0.000
gray Cast iron	0.310	1.363	0.1827	0.089
polyethylene	0.168	1.183	0.2808	0.550
Constant	-3.263	0.038	0.378	0.000
Age	0.011	1.011	0.005	0.034
length	0.003	1.003	0.000	0.000
diameter	-0.003	0.997	0.000	0.000
pressure	0.028	1.028	0.004	0.000
traffic				
under a sidewalk	-1.231	0.292	0.260	0.000
normal traffic	-0.827	0.437	0.271	0.002
material				
ductile cast iron	-1.838	0.159	0.171	0.000

### 3.5 ROC curve

The above models are designed to show the impact of each variable in the pipe failure rate. To evaluate the discriminate performance of the models and to compare them we may represent the ROC curve and calculate the AUC. They are typically used for any test in which there are two possible outcomes that are of interest to the researcher. Does the test give a correct positive result (A) or a correct negative result (B). In this case, it corresponds to the fact that either the pipe has actually failed (A) or not (B). A ROC curve is considered

to be a summary measurement in the sense that it uses all possible decision thresholds for the test to create the curve. The ROC Curves serve to compare models with the same response variable, as is the case with the GLM and Cox whose response is the hazard rate, but not that of Weibull whose response is failure time. The ROC curves were obtained in R using the ROCR library [19].

In Table 5 we show the fp rate and tp rate values for three thresholds, the optimal operating point being the central one.

Table 5

ROC curve decisions

Model	fpr (sensivity)	tpr (1-specificity)	risk threshold
Cox	0.4620	0.8680	0.003000
	0.3337	0.8122	0.006209
	0.3081	0.7690	0.007998
GLM	0.2788	0.7919	0.010003
	0.2445	0.7690	0.017952
	0.2267	0.7487	0.022396

When comparing the ROC curves for the two models, tests with larger areas are typically considered to be better tests, it is possible for tests to have the same area and yet not be considered to be equivalent. Thus, is important to compare ROC curves and their areas.

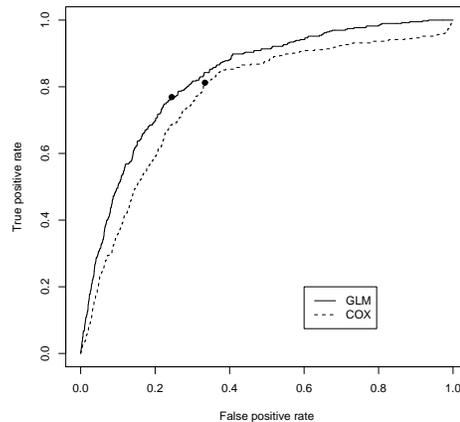


Fig. 1. ROC curves.

Figure 1 shows ROC curves based on the corresponding models whose comparison allows us to say that the hazard rate obtained by GLM discriminates

better between the pipes that break than those that do not, because its curve is always above the corresponding Cox model curve.

In addition, the AUC corresponding to the risk obtained by GLM is 0.8278 greater than the AUC corresponding to the risk obtained by Cox model of 0.7686. This confirms the superiority of the GLM.

Likewise, Figure 1 shows the ROC curve of each model with the optimal operating point marked.

## 4 Conclusions

First of all, what is suggested from here on is that, as far as is possible, the date of failure should be noted in order to have a failure time variable without zeros and to resolve the problem with the data at the same time achieving greater precision.

The study has sought to provide insight into the impact of different variables on the risk of failure in water supply networks. The analysis described above showed that pipes which were less prone to failure had the following characteristics: short lengths, large diameters, low pressure and installed under a sidewalk. As regard to material, there is not the same solution for the three models: pipes made with asbestos material for the Cox model, but pipes made with ductile cast iron are for the Weibull model and GLM.

We have compared two different models by choosing the best fits for each one of them. Specifically, we have compared the Cox model and GLM. The comparison is carried out by applying the ROC curve, from which we can conclude that the Cox model model produces worse fit than GLM. Moreover, this methodology has allowed us to establish a threshold at which a pipe can be considered high-risk, decision making which allows for the renewal of the network.

In relation to the work of other authors, we should highlight one distinctive feature of the methodology presented here, which is the possibility of comparing the different models with a simple and objective criterion. ROC graphs are able to provide a measure of classification performance, to be a complement to scalar measures such as goodness of fit statistics proposed by [21].

In short, we propose a statistical tool which provides a clear framework for decision support in the diagnosis and rehabilitation of water supply systems.

## References

- [1] Agresti, A., 2002. *Categorical Data Analysis*. Wiley, 2<sup>a</sup> Edición.
- [2] Andreou, S., Marks, D., Clark, R., 1987a. A new methodology for modeling break failure patterns in deteriorating water distribution systems: Applications. *Advances in Water Resources* 10.
- [3] Andreou, S., Marks, D., Clark, R., 1987b. A new methodology for modeling break failure patterns in deteriorating water distribution systems: Theory. *Advances in Water Resources* 10.
- [4] Boxall, J., O'Hagan, A., Pooladsaz, S., Saul, A., Unwin, D., 2007. Estimation of burst rates in water distribution. *Water Management* 160.
- [5] Clark, R., Stafford, C., Goodrich, J., 1982. Water distribution systems: a spatial cost evaluation. *Journal of Water Resources Planning and Management Division ASCE* 108.
- [6] Cox, D., 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34.
- [7] Cox, D., Oakes, D., 1984. *Analysis of Survival Data*. Chapman and Hall, London.
- [8] Eisenbeis, P., 1994. Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable. Tech. rep., Série Équipements pour l'eau et l'environnement, Études du CEMAGREF.
- [9] Eisenbeis, P., 1999. Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable. Ph.D. thesis, University Louis Pasteur of Strasbourg.
- [10] Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27.
- [11] Fenner, R., 2000. Approaches to sewer maintenance: a review. *Urban Water* 2.
- [12] Fleming, T., Harriton, D., 2002. *Counting processes and Survival analysis*. John Wiley and Sons, Nueva York.
- [13] Kleiner, Y., Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* 3.
- [14] Lei, J., 1997. Statistical approach for describing lifetimes of water mains- case trondheim municipally. Report no 22f007.28, SINTEF Civil and Environmental Engineering, Trondheim, Norway.
- [15] Marks, D.H., et al., 1985. Predicting urban water distribution maintenance strategies: A case study of New Haven Connecticut. US Environmental Protection Agency (Co-operative Agreement R8 1 0558-01-0).
- [16] McCullagh, P., Nelder, J., 1989. *Generalized Linear Models*. Chapman and Hall, London.

- [17] original by Terry Therneau, S., ported by Thomas Lumley, 2008. survival: Survival analysis, including penalised likelihood. R package version 2.34-1.
- [18] Saegrov, S., Melo-Baptista, J., Conroy, P., Herz, R., LeGauffre, P., Moss, G., Oddevald, J., Rajani, B., Schiatti, M., 1999. Rehabilitation of water networks survey of research needs and ongoing efforts. Urban Water 1.
- [19] Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2007. ROCR: Visualizing the performance of scoring classifiers. R package version 1.0-2.  
URL <http://rocr.bioinf.mpi-sb.mpg.de/>
- [20] Venables, W., Ripley, B. D., 2002. Modern Applied Statistics with S-PLUS. Springer, Nueva York.
- [21] Yamijala, S., Guikema, S., Brumbelow, K., 2009. Statistical models for the analysis of water distribution system pipe break data. Reliability Engineering and System Safety 94.