

# 상수관로에 대한 시간종속형 공변수를 포함한 포괄적 비례위험모형

## The Comprehensive Proportional Hazards Model Incorporating Time-dependent Covariates for Water Pipes

박수완\*

Park, Suwan

### Abstract

In this paper proportional hazards models for the first through seventh break of 150 mm cast iron pipes in a case study area are established. During the modeling process the assumption of the proportional hazards for covariates on the hazards is examined to include the time-dependent covariate terms in the models. As a result, the pipe material/joint type and the number of customers are modeled as time-dependent for the first failure, and for the second failure only the number of customers is modeled as time-dependent. From the analysis on the baseline hazard functions the failure hazards are found to be generally increasing for the first and second failure, while the hazards of the third break and beyond showed a form of a bath-tub. Furthermore, the changes in the baseline hazard rates according to the time and number of break reflect that the general condition of the pipes is deteriorating. The factors causing pipe break and their effects are analyzed based on the estimated regression coefficients and their hazard ratios, and the constructed models are verified using the deviance residuals of the models.

**keywords** : cast iron pipe, hazard rate, pipe break, proportional hazards model, survival function, time-dependent covariate

### 요 지

본 논문에서는 연구대상 지역의 150 mm 주철 상수관로의 첫 번째 파손으로부터 일곱 번째 파손사건에 대한 비례위험모형을 구축하였다. 모형의 구축과정에서 공변수의 위험률에 대한 비례위험 가정을 검사하여 이를 위배할 경우 시간종속형 공변수로 모형화하였다. 그 결과 첫 번째 파손에 대해서는 관로의 제원 및 연결 방식과 급수인구가, 그리고 두 번째 파손 사건에 대해서는 급수인구의 영향이 시간에 따라 변하는 것으로 나타났다. 각 생존시간군의 기저위험률에 대한 분석으로부터 첫 번째와 두 번째 파손에 대해서는 대체적으로 파손 위험률이 시간에 따라 계속해서 증가하는 것으로 나타났으며, 세 번째 파손으로부터 일곱번째 파손사건에 대해서는 파손 위험률이 감소하다가 시간이 지나면 증가하는 욱조 모양으로 추정되었다. 또한 시간과 파손횟수에 따른 기저위험률의 변화 및 각 생존시간군의 중간생존시간으로부터 연구대상 상수관로들은 파손횟수가 증가할수록 전반적인 관로의 상태가 악화되는 것으로 판단된다. 추정된 공변수의 회귀계수와 위험비율을 이용하여 관로파손에 미치는 인자와 그 시간적 영향에 대하여 분석하였으며, 구축된 모형의 이탈잔차를 이용하여 모형의 적합도를 검증하였다.

**핵심용어** : 주철관, 위험률, 관로 파손, 비례위험모형, 생존함수, 시간종속형 공변수

\* 교신저자, 부산대학교 공과대학 사회환경시스템공학부 부교수

Corresponding Author, Associate Prof., Dept. of Civil and Env. Eng., Pusan National Univ., Busan 609-735, Korea  
(e-mail: swanpark@pusan.ac.kr)

## 1. 서론

비례위험모형(Cox, 1972)은 배수관로의 파손율을 위험률의 형태로 모형화하는 통계적 모형으로서, 관로의 내·외부에 작용하는 파손인자의 영향을 정량적으로 파악하고 시간에 따른 개별관로의 매설 후 생존확률을 산정하는데 사용되어 왔다. 비례위험모형의 가장 큰 특징은 파손인자의 영향을 산정하는 함수와 이러한 영향이 없는 상태의 파손경향을 표현하는 기저함수를 분리하여 파손에 영향을 주는 인자들 간의 상대적 효과를 산정하는 것이 가능하다는 것이다. 또한 기저함수의 추정 과정에 있어서 관로의 파손시간에 대한 어떠한 통계적 분포형도 고려하지 않을 수 있으므로 매우 다양한 관로 파손경향을 모형화할 수 있는 장점이 있다. 이러한 비례위험모형의 특징을 이용하여 배수관로의 파손 위험율을 연구한 선행연구는 다음과 같다.

Andreou *et al.*(1987a,b)과 Marks *et al.*(1987)은 상수관로의 파손 경향을 두 단계로 구분하였는데, 비교적 파손 횟수가 적은 관로의 초기 상태를 비례위험모형으로 모형화하였고 파손 횟수가 많은 관로의 후기 상태를 프와송(Poisson) 모형을 이용하여 모형화하였다. Eisenbeis(1994)와 Brémond(1997)은 프랑스 보르도 지방 상수관로의 33년간 파손자료와 와이블(Weibull) 기저위험함수를 이용하여 관로파손횟수를 예측하였다. Lei(1997)는 노르웨이의 상수관로에 대해 비례위험모형을 적용하였으며, 연구대상 관로들은 관중에 따라 시간에 따른 노후화 경향이 다른 것으로 보고하였다. Li and Haims(1992a,b)는 Andreou *et al.*(1987a,b)의 비례위험모형과 프와송 모형을 기초로 관로의 2 단계(two-stage) 유지관리 의사결정 모형을 개발하였는데, 첫 번째 단계에서는 관로의 사용성(availability)을 최대한으로 하면서 관로의 다양한 노후도에 따른 최적 보수/교체 의사결정을 위해 마르코프(Markovian) 모형을 적용하였으며, 두 번째 단계에서는 관망 전체의 사용성을 최대한으로 하면서 유지관리 자금을 최적으로 분배하기 위해 다중수준 분해(multilevel decomposition) 방법을 이용하였다. 상수관로의 파손자료를 이용한 비례위험모형의 적용에 관한 보다 자세한 검토는 Kleiner and Rajani(2001)이 제공하고 있다.

본 논문에서는 박수완 등(2007)의 비례위험모형링을 위한 개별관로 정의 방법을 이용하여 연구대상 지역인 미국 중부 도시의 배수관로 파손 DB 내 관로를 개별관로로 정의하고, 이들 개별관로들의 순차적 파손시간에 대한 비례위험모형을 구축하기 위한 포괄적인 방법론을 설명한다. 본 논문에서 제시된 비례위험모형 구축 방법론은 모형에 사용될 공변수(covariate)라고 불리는 파

손 설명변수(explanatory variable)의 선정 방법, 공변수의 비례위험 가정 검증법, 기저 생존 및 위험함수 산정법, 그리고 구축된 모형의 검증법을 포함한다. 특히, 통계적으로 유의하다고 판단된 공변수의 비례위험 가정에 대한 검사는 비례위험모형링 과정에서 매우 중요한 부분을 차지하는데, 상수관로 파손확률 해석에 비례위험모형을 적용한 이전의 연구에서는 이러한 상수관로에 대한 비례위험모형 공변수의 비례위험 가정에 대한 검증이 이루어지지 않았다. 그러나 본 논문에서는 고려된 모든 공변수에 대한 비례위험 가정을 검증하고 비례위험 가정이 위배된다고 판정된 공변수는 시간 종속형(time-dependent) 공변수로 모형화하였다.

본 논문에서는 박수완 등(2007)에 사용된 자료와 동일한 자료를 이용하여 연구대상 관로들에 대해 첫 번째 파손으로부터 일곱 번째 순차적 파손사건에 대한 비례위험모형을 구축하였다. 그러나 본 논문에서는 박수완 등(2007)에서는 간과된 비례위험모형링에서 매우 중요한 단계인 공변수의 비례위험 가정을 일곱 개의 모형모두에 대하여 검증하였다. 그 결과 첫 번째와 두 번째 파손사건 모형에 대하여 채택된 공변수와 그 회귀계수들이 박수완 등(2007)과 달리 추정되었다. 또한 본 논문에서는 비례위험모형의 기저위험함수를 도출하는 과정에서 추정된 기저누적위험함수의 값을 직접 이용하여 기저위험함수를 도출하므로써 박수완 등(2007)이 기저생존함수를 먼저 추정한 후 기저위험함수를 유도한 오류를 수정하였으며, 이로 인해 본 논문에서는 일곱 개 모형 모두에 대해 박수완 등(2007)과 다른 기저위험함수가 추정되었다. 한편 본 논문에서 세 번째 파손으로부터 일곱 번째 파손사건에 대한 공변수와 그 회귀계수는 박수완 등(2007)과 거의 동일하게 산정되었으므로, 본 논문에서는 첫 번째와 두 번째 파손사건 만에 대해서 공변수의 위험비율을 이용하여 파손 인자의 영향을 분석하였다.

또한 본 논문에서는 모형의 적합도 검증(goodness-of-fit)을 통하여 구축된 모형이 관측된 관로의 파손사건을 적절하게 표현하는가를 검증하였으며, 구축된 모형을 통하여 관로의 파손 경향에 대해 분석하였다. 즉 관로의 파손에 영향을 미치는 인자들의 상대적 영향을 분석하였고, 도출된 기저 생존 및 위험함수로부터 순차적 파손사건의 특성에 대해 분석하였다.

## 2. 비례위험모형의 구축 과정

### 2.1 비례위험모형, 생존함수 및 위험비율의 정의

비례위험모형은 개체  $i$ 가  $t$ 시점까지 생존했을 경우에

다음 미소시간 동안에 사망 혹은 파손될 단위시간당 확률을 나타내는 위험률  $h_i(t)$ 와 이에 영향을 미치는 요인들인  $x_{1i}, x_{2i}, \dots, x_{qi}$ 의 관계를 Eq. (1)과 같이 표현한다.

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_q x_{qi}) \quad (1)$$

여기서  $h_0(t)$ 는 생존시간에 대한 분포형이 결정되지 않은 기저위험함수(baseline hazard function)를 나타내는데, 생존시간은 개체가 사용되기 시작한 시간으로부터 사망 또는 개체에 어떤 특정한 사건이 발생하기까지의 경과 시간을 나타낸다. 생존시간에 영향을 미치는 인자를 나타내는  $x$ 는 흔히 공변수(covariate)라고 불리며,  $\beta$ 는  $x_i$ 에 관련된 회귀계수를 나타낸다.

생존함수는 어떤 개체가 특정 시간을 초과하여 생존하는 확률, 즉 특정 시간까지 어떤 사건이 발생하지 않는 확률을 말하는데, 위험함수와 생존함수의 관계는 Eq. (2)와 같다.

$$S_i(t) = e^{-\int_0^t h_i(\tau) d\tau} = \exp\left(-\int_0^t h_0(\tau) e^{(\beta' x_i)} d\tau\right) \quad (2)$$

$$= [S_0(t)]^{\exp(\beta' x_i)}$$

여기서  $S_0(t)$ 는 생존시간에 대한 분포형이 결정되지 않은 기저생존함수(baseline survival function),  $x$ 는 공변수들의 열벡터,  $\beta$ 는 공변수들의 열벡터를 나타낸다. 생존함수  $S(t)$ 는 기저생존함수  $S_0(t)$ 를 추정된 후 최대우도법(maximum likelihood method)으로 추정된 회귀계수  $\beta$ 를 이용하여 Eq. (2)에 따라 추정될 수 있다.

비례위험모형을 상수관로 문제에 적용시킴으로써 얻을 수 있는 큰 이점은 서로 다른 관로, 즉 서로 다른 공변량을 가지는 관로들 간의 상대적 위험률을 산정할 수 있다는 점이다. 다른 공변량은 두 관로 모두에 대해 일정하고 하나의 공변량의 값만이 단위 크기만큼 증가할 때의 상대적 위험률을 위험비율(hazard ratio)이라고 하며, 위험비율은 일반적으로 Eq. (3)과 같이 산정된다.

$$\widehat{HR} = \frac{\widehat{h}_i(t)}{\widehat{h}_j(t)} = \frac{\widehat{h}_0(t) \exp(\widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i} + \dots + \widehat{\beta}_q x_{qi})}{\widehat{h}_0(t) \exp(\widehat{\beta}_1 x_{1j} + \widehat{\beta}_2 x_{2j} + \dots + \widehat{\beta}_q x_{qj})} \quad (3)$$

$$= \exp\left\{\sum_{m=1}^q \widehat{\beta}_m (x_{mi} - x_{mj})\right\}$$

위험비율을 이용하여 토지개발정도(예를 들어, 도시 및 비도시 지역)와 같은 공변량의 변화가 관로의 파손 위험률에 미치는 상대적 영향을 산정할 수 있게 된다. 따라서, 예를 들면, 도시 지역과 비도시 지역에 각각

위치한 두 관로의 토지개발정도를 나타내는 공변수의 위험비율이 1일 경우 두 관로의 파손 위험률은 도시 지역 또는 비도시 지역으로 구분지어진 매설 환경에는 영향을 받지 않는다는 것을 나타낸다. 비례위험모형의 공변수에 대한 회귀계수의 추정, 기저생존함수 및 기저위험함수의 추정 및 위험비율에 대한 자세한 설명은 Allison(1996) 및 박재민(2006)이 제공하고 있다.

## 2.2 공변수의 정의 및 선정과정

본 연구에서는 연구대상 관로를 Park(2004)의 순차적 관로 파손사건에 대한 비례위험모형 구축 방법론을 이용하여 첫 번째 부터 일곱 번째 파손사건까지의 비례위험모형(Model I ~ Model VII)을 구축하였다. 구축된 각 각의 비례위험모형은 관로의 순차적 파손사건에 따라 분류한 각 생존시간군(Survival Time Group, STG), 즉 STG I에서 STG VII까지의 각 생존시간군에 대한 모형을 나타낸다.

여기서 I부터 VII까지 변하는 STG 번호를 'n' 이라 하면, n은 각 관로군에 속하는 관로에 발생한 파손횟수가 최소 n-1 임을 나타낸다. 따라서, 예를 들면, STG II는 관로에 발생한 총 파손횟수가 최소 '1'인 관로들로 구성된 관로의 집합을 나타내며, STG V는 관로에 발생한 총 파손횟수가 최소 '4'인 관로들로 구성된 관로의 집합을 나타낸다. 이러한 방식으로 연구대상 관로들에 대한 생존시간군을 구축한 이유는 순차적 파손사건에 대한 파손 위험률을 모형화하기 위한 것이다. 따라서, 예를 들면, STG III에 대한 비례위험모형 Model III는 두 번째 파손이 발생한 관로들에 세 번째 파손이 발생할 위험률을 모형화한 결과를 나타낸다. Table 1은 모형의 구축 과정에서 일차적으로 고려된 공변수의 정의와 각 관로에 부여될 수 있는 공변량을 나타낸다.

관로의 길이(L) 및 급수인구(C) 공변수는 Model I에서는 기록된 값을 그대로 사용하는 연속형 변수로 취급되었으나, Model II 이상에서는 공변량의 분포가 정규분포를 따라야 한다는 선형회귀분석의 가정을 만족시키고 통계적으로 유의한 모형의 구축을 위하여 관로의 길이 및 급수인구 공변량들의 자연로그 값들을 구한 다음 평균 이상의 값과 평균 미만의 값으로 구분하여 이분변수(binary variable)로 각 개별관로에 부여하였다. 따라서 모델링에 고려된 공변수는 토지개발정도(DL), 길이(L) 및 GRID 내의 급수인구(C), 관로재질 및 연결방식(TYPE, SR, SF)과 수압유형(PT1, PT2, PT3)이다. Table 1에서 TYPE은 Model I에서는 관로가 Pit-CI(Pit Cast Iron)일 때는 '0', Spun-rigid joint(SR)일 때는 '1', 그리고 Spun-flexible joint(SF)일 때는 '2'의

Table 1. Types and Characteristics of the Covariates

Covariate	Definition	Type	Characteristics
<i>SR</i>	Spun-rigid Joint Cast Iron Pipe	binary	Used only for Model III ~ Model VII; If a pipe is spun-rigid joint, $SR = 1$ , otherwise, $SR = 0$
<i>SF</i>	Spun-flexible Joint Cast Iron Pipe	binary	Used only for Model III ~ Model VII; If a pipe is spun-flexible joint, $SF = 1$ , otherwise, $SF = 0$
<i>TYPE</i>	pipe material	continuous	Used only for Model I and Model II
<i>DL</i>	degree of land development	binary	urban = 1, non-urban = 0
<i>L</i>	pipe length	continuous or binary	For MODEL I continuous variable, otherwise binary variable.
<i>C</i>	customers in GRID	continuous or binary	For MODEL I continuous variable, otherwise binary variable.
<i>PT1</i>	pressure type 1	binary	If a pipe is <i>PT1</i> , $PT1 = 1$ , otherwise $PT1 = 0$
<i>PT2</i>	pressure type 2	binary	If a pipe is <i>PT2</i> , $PT2 = 1$ , otherwise $PT2 = 0$
<i>PT3</i>	pressure type 3	binary	If a pipe is <i>PT3</i> , $PT3 = 1$ , otherwise $PT3 = 0$

값을 가지며, Model II에서는 관로가 Pit-CI일 일 때는 '1', *SR*일 때는 '0', 그리고 *SF*일 때는 '2'의 값을 가지는 연속형 변수(continuous variable)이다. 한편, Model III 이상의 모형에 대해서 관종을 나타내는 변수는 *TYPE*이 아닌 *SR*과 *SF*로서 이들 공변수의 값, 즉 공변량은 Table 1에 나타나 있는 바와 같다.

적절한 모형의 선정방법은 forward, backward 또는 stepwise 변수 선택법에 의한, 오로지 통계적인 기준에만 의존하여 모형을 선정하는 방법을 사용하는 것도 가능하나, 관로의 파손특성을 보다 잘 설명할 수 있는 모형을 구축하기 위해서는 공변수의 가능한 모든 조합 및 공변수 간의 상호작용 그리고 시간종속 공변수(time-dependent covariate)에 대한 검토도 수행하여야 한다. 따라서 각 생존시간군의 최종적인 비례위험모형을 구축하기 위하여, 먼저 모든 공변수에 대한 단일변량분석(univariate analysis)을 통하여 각 변수가 생존율과 어떠한 연관이 있는지를 살펴보고, 생존율에 영향을 준다고 판단되는 변수를 일차적으로 선정하였다. 각 생존시간군에 대한 모형의 최종적인 공변수는 공변수의 가능한 모든 조합 및 공변수 간의 상호작용에 대한 통계적 유의성 검토를 대수우도비 통계량(log-likelihood ratio statistic) (-2Log PL) 및 Akaike Information Criterion (AIC)으로 검사하는 Collet(2003)의 공변수 선택 방법을 수행하여 결정하였다. 모형에 포함될 공변수의 선택과 최대우도 추정법을 이용한 회귀계수의 추정은 통계분석용 소프트웨어인 SAS 프로그램을 이용하였다.

### 2.3 공변수의 비례성 가정 검토 및 회귀계수의 추정

비례위험모형에서 채택하고 있는 가장 중요한 가정은 공변수의 위험률에 대한 비례성 가정이다. 즉, 비례위험모형 내의 하나의 공변수의 값(공변량)이 증가 혹은 감소할 경우 그에 대한 위험률도 공변량의 증감에 비례하여 변화하고, 이러한 경향은 시간에 관계없이 일정하게 유지된다고 하는 것이다. 따라서 만약 모형 구축에 유의한 것으로 판단된 공변수 일지라도 그 공변수에 대한 비례성이 확인되지 않을 경우에는 공변수의 영향이 시간에 따라 변화하는 시간종속형 공변수로 모형에서 다시 고려하여야 한다.

각 생존시간군에 대해 채택된 공변수들의 위험률에 대한 비례성을 검토하기 위하여 Klein and Moeschberger(2003)의 Schoenfeld 잔차를 이용한 스코어 잔차의 변화(score process)를 검토하였다. Fig. 1 ~ Fig. 4는 공변수의 비례성 검토를 위한 Model I의 스코어 과정(score process)을 나타낸다.

시간에 따른 표준 스코어 잔차의 변화는 Brownian Bridge에 근사하고, Brownian Bridge의 값이  $\pm 1.35$ 를 초과할 확률은 0.05로 산정된다(Therneau *et al.*, 1990). 따라서 비례위험모형 공변수의 비례성 위반 판정기준은 Brownian Bridge 즉, 표준 스코어잔차의 최대값이  $\pm 1.35$ 을 초과할 경우 유의수준 5%에 의하여 공변수가 비례성을 가진다는 귀무가설(null hypothesis)을 기각(reject) 할 수 있다. Fig. 1, Fig. 3 및 Fig. 4에 나타난

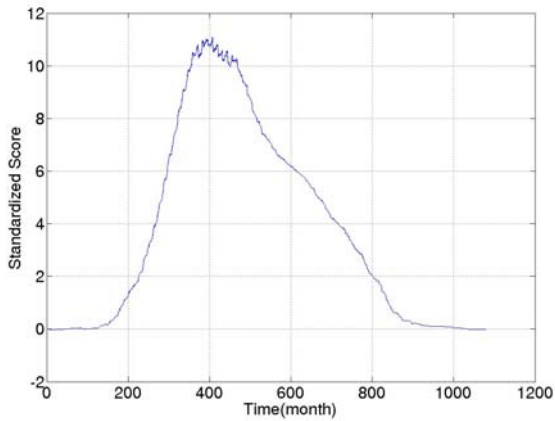


Fig. 1. Standardized Score Residuals of *TYPE* of Model I

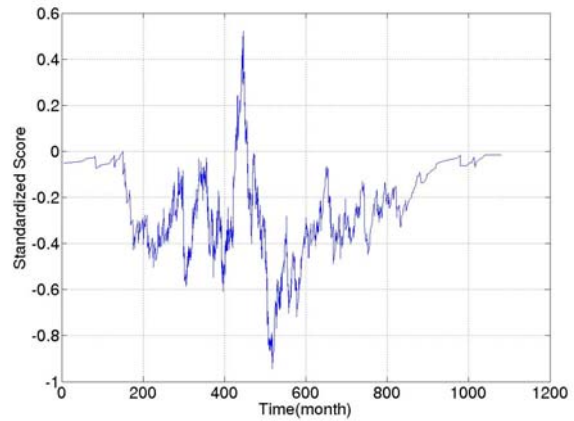


Fig. 2. Standardized Score Residuals of *DL* of Model I

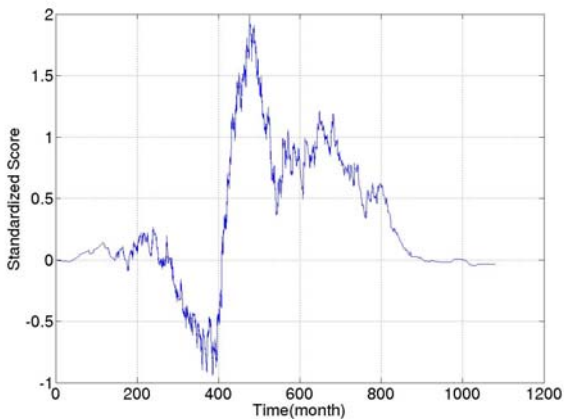


Fig. 3. Standardized Score Residuals of *L* of Model I

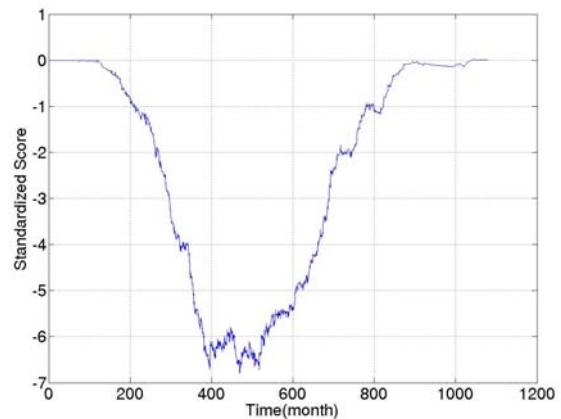


Fig. 4. Standardized Score Residuals of *C* of Model I

바와 같이 Model I에서 표준 스코어 잔차가  $\pm 1.35$ 을 초과하는 공변수는 *TYPE*, *L* 및 *C*이며, 따라서 *TYPE*, *L* 및 *C* 공변수를 시간종속형 변수로 다시 모형화할 필요가 있다.

한편 Model I에서 *L*은 표준 스코어 잔차가  $\pm 1.35$ 을 초과하기는 하나 *TYPE*과 *C*에 비하여 상대적으로 표준 스코어 잔차가 매우 작은 것으로 계산되었다. 따라서 *L*은 *TYPE*과 *C*에 비하여 비례성의 위반 정도가 약한 것으로 분석된다. 실제로 *L*을 시간종속형 변수로 모형화할 경우 모형 자체의 통계적 유의성은 높은 것으로 판정되나, 그 회귀계수가 비현실적으로 크게 산정되어 결과적으로 구축된 모형은 관로 길이의 파손에 대한 영향을 비현실적으로 과대하게 산정하는 것으로 판정되었다. 이러한 경향은 표준 스코어 잔차의 최대값과 최소값이  $\pm 2$  정도로 계산되는 Model II의 *TYPE*과 *L*, Model III의 *SR* 및 Model V의 *SR*과 *SF*에서도 발견되었다. 따라서 본 연구에서는 공변수의 위험률에 대한 비례성 위반 판정기준을 표준 스코어 잔차가  $\pm 3$ 를 초과

하는 경우로 설정하였으며, 이러한 기준에 따라 선택된 공변수를 시간종속형 공변수로 고려하여 그 회귀계수를 구한 결과는 Table 2와 같다. Table 2에서 공변수들 간의 상관관계 그리고 공변수와 시간 사이의 상관관계 즉, 상호작용이 모형에 포함된 것은 통계적으로 더욱 유의한 모형을 구축하기 위한 것으로써 이들 변수들 간의 상호작용을 모형에 포함하였을 경우가 포함하지 않았을 경우 보다 모형의 통계적 유의성이 더욱 증가하였기 때문이다.

## 2.4 기저함수의 추정

각 생존시간군에 대한 비례위험모형의 기저생존함수는 SAS 프로그램의 'baseline' 명령문을 이용하여 산정되는 LLS라고 하는 기록된 각 파손시점에 대한 기저생존확률의 추정치( $\hat{S}_0(t)$ )의 로그-로그 변환값( $\ln(-\ln \hat{S}_0(t))$ )을 적절한 함수형으로 근사시켜서 구할 수 있다. 따라서 각 생존시간군에 따라 다른 형태의 기저생존함수가 도출될 수 있다.

한편, 비례위험모형의 기저위험함수는 기록된 각 파손시간에 대해 추정된 기저누적위험율(baseline cumulative hazards)의 연속된 파손시간 사이의 차이값을 산정한 다음 이를 임의의 함수형으로 추정하여 구할 수 있다(Collet, 2003). 여기서 추정된 기저누적위험율은 SAS 프로그램의 'baseline' 명령문을 이용하여 산정할 수 있다. 그러나 박수완 등(2007)은 기저위험함수를 추정함에 있어서 이러한 방법을 취하지 않고 기록된 각 파손시점에 대한 *LLS*를 이용하여 기저생존함수를 먼저 추정한 다음 이를 다시 기저위험함수로 변환시키므로써 비례위험모델링의 중요한 특성인 기저위험함수에 대한 어떠한 함수형도 미리 가정하지 않는다는 기본 요구 조건을 위배하는 오류를 범하였다. 본 논문에서는 이러한 박수완 등(2007)의 오류를 수정하여 연구대상 관로들에 대해 비례위험모형의 기본적 모형화 법칙을 만족하는

기저위험함수를 추정하였다. 비례위험모형의 기저생존함수와 기저위험함수를 추정하는 자세한 방법은 Collet (2003)이 제공한다.

Fig. 5는 STG I의 *LLS*를 그래프로 나타낸 결과이며, 이를 로지스틱 함수형으로 나타내면 Eq. (4)와 같다.

$$LLS(t) = \frac{28.87}{1 + \exp(-2.613 \cdot \ln t + 14.87)} - 23.17 \quad (4)$$

Eq. (4)를 STG I의 기저생존함수로 변환하면 Eq. (5)와 같다.

$$\begin{aligned} \hat{S}_0(t) &= \exp\left\{-\exp\left(\frac{28.87}{1 + \exp(-2.613 \cdot \ln t + 14.87)} - 23.17\right)\right\} \\ & \quad (5) \end{aligned}$$

Table 2. Estimated Parameters with Time-dependent Covariates

Model	Covariate	Parameter	p-value (Pr < Ch-square)	Hazard ratio	95% Hazard ratio confidence limits	
					Lower limit	Upper limit
I	<i>TYPE</i>	1.34242	<.0001	3.828	3.443	4.257
	<i>DL</i>	0.54472	<.0001	1.724	1.543	1.926
	<i>L</i>	0.000497	<.0001	1.000	1.000	1.001
	<i>C</i>	0.01968	<.0001	1.020	1.019	1.021
	<i>DL · L</i>	0.000259	0.0093	1.000	1.000	1.000
	<i>TYPE · C</i>	-0.01284	<.0001	0.987	0.987	0.988
	<i>TYPE · time</i>	-0.02302	<.0001	0.977	0.977	0.978
	<i>C · time</i>	-0.00003	<.0001	1.000	1.000	1.000
II	<i>TYPE</i>	0.17846	<.0001	1.195	1.117	1.279
	<i>DL</i>	0.34444	0.0183	1.411	1.060	1.879
	<i>L</i>	1.70421	<.0001	5.497	4.609	6.556
	<i>C</i>	1.25702	<.0001	3.515	2.309	5.351
	<i>DL · C</i>	-0.531	0.0153	0.588	0.382	0.903
	<i>C · time</i>	-1.6262	<.0001	0.197	0.177	0.218
III	<i>SR</i>	0.25651	0.3334	1.292	0.769	2.173
	<i>SF</i>	-0.06004	0.8321	0.942	0.541	1.640
	<i>L</i>	0.85493	<.0001	2.351	1.934	2.859
	<i>C</i>	-0.16995	0.0679	0.844	0.703	1.013
IV	<i>L</i>	0.72912	<.0001	2.073	1.612	2.666
V	<i>SR</i>	-0.87835	0.0606	0.415	0.166	1.040
	<i>SF</i>	-1.01847	0.0408	0.361	0.136	0.958
	<i>DL</i>	0.64732	0.0572	1.910	0.980	3.723
	<i>L</i>	0.32466	0.0457	1.384	1.006	1.902
VI	<i>L</i>	0.44147	0.0340	1.555	1.034	2.339
VII	<i>L</i>	0.66856	0.0097	1.951	1.176	3.239

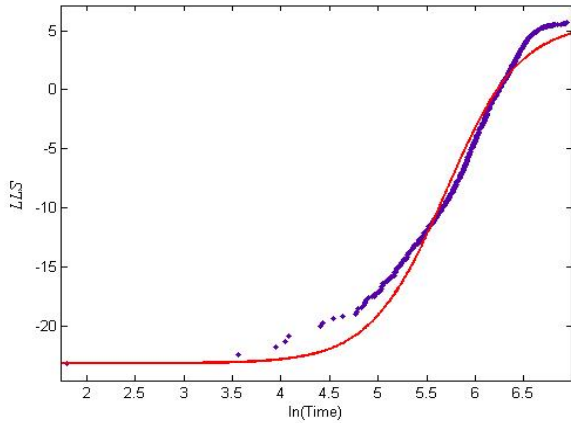


Fig. 5. Plot of  $\ln$  Time and LLS for STG I

Model I의 각 파손시간에 대한 누적기저위험률을 추정 후 연속되는 파손사건 사이의 누적기저위험률의 차이로 산정한 Model I의 기저위험률을 LOESS(locally weighted scatterplot smoothing) 회귀모형을 이용하여 추정한 결과는 Fig. 6과 같다. LOESS 회귀모형은 자료의 변동 경향을 모형화하기 위하여 자료 전체에 어떤 함수형을 적용하지 않고 개별 자료들에 대하여 선형과 같은 낮은 차수의 회귀모형을 개별자료에 대한 근접성을 고려한 가중치를 이용하여 적용하고 이러한 개별적 회귀모형을 연결하는 비모수적 회귀모형의 하나이다.(Cleveland, 1979; Cleveland and Devlin, 1988).

Model I의 생존함수는 Eq. (5)로 나타나는 Model I의 기저생존함수와 Table 1에서 구한 Model I의 공변수 및 그 회귀계수를 이용하면 아래 Eq. (6)으로 표현된다.

한편 STG II의 로그변환한 시간과 기저생존확률의 로그-로그변환값은 Eq. (7)과 같이 선형적 관계를 보이는 것으로 나타났다.

$$LLS(t) = 1.2722 \ln t - 7.1873 \quad (7)$$

파손시간이 Weibull 분포를 따를 경우 로그변환한 시간과 기저생존확률의 로그-로그변환값은 선형적 관계를 가지므로 STG II의 관로의 파손시간은 Weibull 분포를 따른다고 볼 수 있으며, 이로부터 Model II의 기저생존함수는 Eq. (8)과 같이 표현된다.

$$\hat{S}_i(t) = \exp \left\{ - \exp \left( \frac{28.87}{1 + \exp(-2.163 \cdot \ln t + 14.87)} - 23.17 \right) \right\} \quad (6)$$

$$\cdot \exp \left( \frac{1.34 TYPE + 0.54 DL + 0.0005 L + 0.02 C + 0.0027 DL \cdot L}{-0.013 TYPE \cdot C - 0.02 TYPE \cdot t - 0.00003 C \cdot t} \right)$$

$$\hat{S}_i(t) = [\exp(-e^{-7.1873} t^{1.2722})] \exp(0.603 SR_i + 1.834 SF_i + 0.582 DL_i + 0.963 L_i + 0.218 C_i) \quad (9)$$

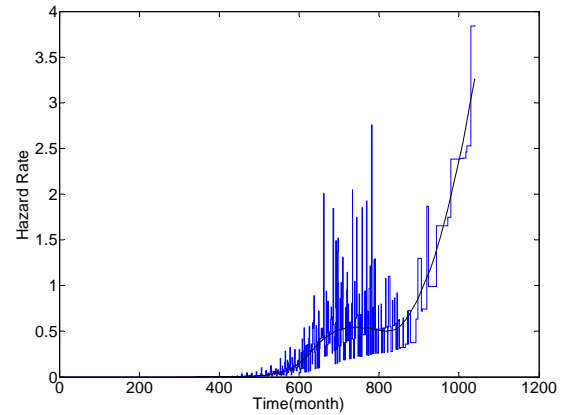


Fig. 6. Graph of the Estimated Baseline Hazard Function for Model I

$$\hat{S}_0(t) = \exp(-\lambda t^\gamma) = \exp(-e^{-7.1873} t^{1.2722}) \quad (8)$$

Model II의 기저위험함수는 Model II의 각 파손시간에 대한 누적기저위험률을 추정 후 연속되는 파손사건 사이의 누적기저위험률의 차이로 산정되며, Model II의 기저위험률을 LOESS 회귀모형을 이용하여 추정한 결과는 Fig. 7과 같다.

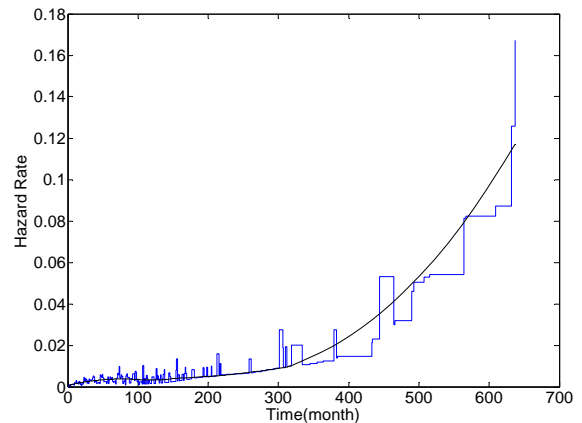


Fig. 7. Graph of the Estimated Baseline Hazard Function for Model II

Model II의 생존함수는 Table 2에서 구한 Model II의 공변수 및 그 회귀계수를 이용하면 Eq. (9)로 표현된다.

STG II와 같이 STG III ~ STG VII의 *LLS* 추정치는 모두 선형으로 산정되었고, 따라서 STG II ~ VII의 파손시간은 Weibull 분포를 따르는 것으로 나타났다. 한편 STG III ~ STG VII에 대해 기록된 각 파손시간에 대해 추정된 기저누적위험률의 연속된 파손시간 사이의 차이 값을 산정한 다음 이를 임의의 함수형으로 추정한 결과, STG III ~ STG VII의 기저위험률은 LOESS 회귀모형보다 모수적 곡선식 접합(parametric curve fitting)이 보다 적합한 것으로 판단되어 STG III ~ STG VII의 기저위험함수는 이차함수형으로 추정되었다.

모든 생존시간군의 기저생존함수 및 기저위험함수를 추정한 결과는 Table 3과 같으며, STG I ~ STG VII의 기저생존함수를 추정한 결과를 그래프로 나타내면 Fig. 8과 같다. Table 3에서 *d*는 LOESS 모형의 차수(degree of local polynomials), 그리고 *q*는 LOESS 모형의 평활 매개변수(smoothing parameter)를 나타낸다. 또한 STG III ~ STG VII의 기저위험함수를 추정한 결과를 그래프로 나타내면 Fig. 9와 같다.

Fig. 9로부터 STG III ~ STG VII의 기저위험률은 파손에 따른 보수 직후의 위험률이 비교적 높으나 시간

이 지날수록 위험률이 감소하다가 다시 증가하는 육조모양과 비슷한 경향을 보인다. 또한 STG III ~ STG VII의 기저위험률은 같은 시간에 대해 일반적으로 파손회수가 많은 관로일수록 높으며, 기저위험률이 감소하다가 증가하는 시점은 파손회수가 많은 관로일수록 짧아진다는 것을 알 수 있다.

첫 번째와 두 번째 파손사건에 대한 기저위험률이 시간에 따라 증가하는 것은 시간에 따른 관로의 노후화로 인해 관로의 파손위험률이 증가하는 것을 나타내며, 파손 횟수가 3회 이상으로 증가할 경우에는 파손 후 초기시간에 관로의 불완전한 보수로 인해 발생하는 부가적인 파손사건에 의해 파손위험률이 높으며, 그 후 일정기간의 안정화단계를 거친 후 다시 관로노후화로 인해 위험률은 높아지는 것으로 분석된다. 한편 Fig. 8로부터 각 생존시간군의 기저생존함수의 생존확률 '0.5'에 해당하는 기저중간생존시간(baseline median survival time)은 STG I으로부터 STG VII까지 약 505, 212, 147, 66, 42, 38, 22 개월인 것으로 추정되어 파손횟수가 많아질수록 순차적 파손사건 사이의 경과시간은 감소하는 것으로 분석된다.

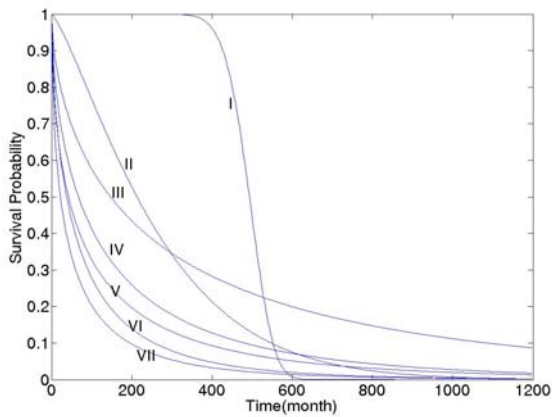


Fig. 8. Graphs of the Estimated Baseline Survival Functions

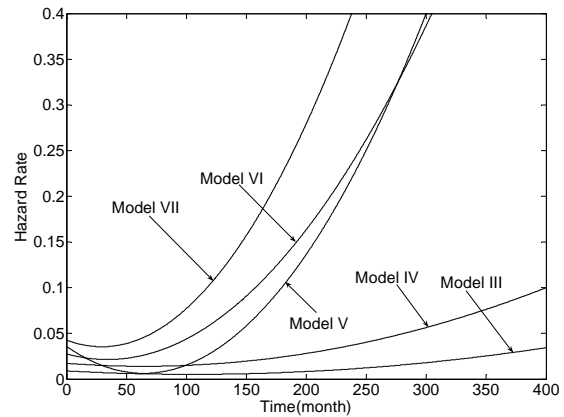


Fig. 9. Graphs of the Estimated Baseline Hazard Function for Model III ~ VII

Table 3. Estimated Baseline Survival and Hazard Functions

STG	Baseline survival Function, $\hat{S}_0(t)$	Baseline Hazard Function, $\hat{h}_0(t)$	R <sup>2</sup>
I	$\exp\left(-e^{\frac{28.87}{1+e^{-2.613 \cdot \ln t + 14.87}} - 23.17}\right)$	LOESS model ( <i>d</i> =2, <i>q</i> =0.25)	N/A
II	$\exp(-e^{-7.1873t^{1.2722}})$	LOESS model ( <i>d</i> =2, <i>q</i> =0.5)	N/A
III	$\exp(-e^{-3.361t^{0.600}})$	$3.336 \times 10^{-7}t^2 - 6.990 \times 10^{-5}t + 0.0089$	0.665
IV	$\exp(-e^{-2.927t^{0.609}})$	$7.632 \times 10^{-7}t^2 - 9.865 \times 10^{-5}t + 0.0173$	0.699
V	$\exp(-e^{-2.417t^{0.548}})$	$7.099 \times 10^{-6}t^2 - 9.161 \times 10^{-4}t + 0.0359$	0.941
VI	$\exp(-e^{-2.642t^{0.627}})$	$5.175 \times 10^{-6}t^2 - 3.546 \times 10^{-4}t + 0.0276$	0.366
VII	$\exp(-e^{-2.072t^{0.551}})$	$8.408 \times 10^{-6}t^2 - 4.972 \times 10^{-4}t + 0.0428$	0.207



## 2.5 모형의 잔차 분석

각 생존시간군에 대해 추정된 비례위험모형의 적합도를 검증하기 위하여 Therneau *et al.*(1990)이 제안한 각 모형의 이탈 잔차(deviance residual)를 산정하였다. 이탈 잔차는 기록된 파손시간과 모형을 이용하여 추정되는 파손시간의 기댓값 사이의 차이를 나타내는 것으로서 Eq. (10)을 이용하여 산정한다.

$$r_{Di} = \text{sgn}(r_{Mi}) [-2\{r_{Mi} + \delta_i \log(\delta_i - r_{Mi})\}]^{1/2} \quad (10)$$

여기서  $\text{sgn}(r_{Mi})$ 는  $r_{Mi}$ 의 값이 음일 경우는 -1, 양일 경우에는 +1의 값을 취하는 함수이며,  $\delta_i$ 는 관측된 시간이 절단된 생존시간일 경우에는 0, 절단된 시간이 아닌 파손시간일 경우에는 1의 값을 가진다. 한편  $r_{Mi}$ 는 관로  $i$ 에 대해 관측된 파손횟수와 구축된 비례위험모형에 따른 파손횟수의 기댓값 사이의 차이를 나타내는  $i$  번째 관로의 마팅게일 잔차(martingale residual)이다 (Collet, 2003).

모형으로 추정되는 개별관로의 이탈 잔차의 절대값이 클수록 기록된 파손시간이 모형으로 추정되는 파손시간의 기댓값과 차이가 많다는 것을 나타낸다. 따라서 이탈 잔차는 모형의 이상치(outlier)를 확인하기 위하여 사용하며, 이탈 잔차의 절대값이 3을 초과할 경우를 모형의 이상치로 정의된다(Allison, 1995). Fig. 10은 모형의 이탈잔차의 예로서 Model I의 이탈잔차를 나타낸다.

이탈 잔차의 절대값이 3을 초과하는 모형은 Model I에서 35 개, 그리고 Model II에서는 3 개가 발견되었으며 다른 모형에서는 발견되지 않았다. 이러한 이상치의 개수는 STG I 및 STG II의 관로 개수에 비해 매우 적고 다른 모든 모형에서는 절대값이 3을 초과하는 이탈 잔차가 발견되지 않았으므로 각 생존시간군에 대해 구축된 비례위험모형들은 기록된 자료에 대체적으로 적합한 것으로 판단된다.

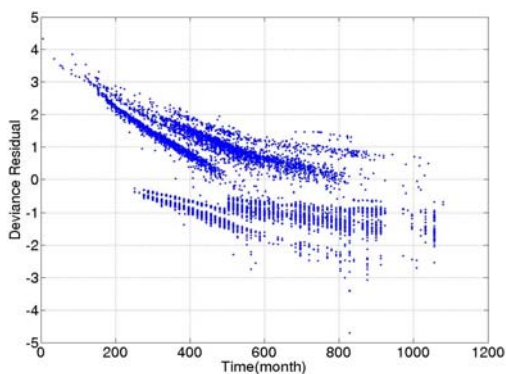


Fig. 10. Deviance Residuals of Model I

## 3. 추정된 회귀계수와 위험비율을 이용한 모형의 분석

Table 2에 보인 바와 같이 본 연구에서 구축된 비례위험모형 Model I 과 Model II 에 대해서는 공변수의 시간종속 효과가 발견되었으나, Model III ~ Model VII 들에 대해서는 그러한 효과가 발견되지 않았다. 본 논문에서 세 번째 파손으로부터 일곱 번째 파손사건에 대한 공변수와 그 회귀계수는 박수완 등(2007)과 거의 동일하게 산정되었으므로, 본 논문에서는 시간종속 공변수가 포함된 첫 번째와 두 번째 파손사건 만에 대해서 공변수의 위험비율을 이용하여 파손 인자의 영향을 분석하였다.

Model I에서의 관로재질(TYPE)만의 전반적인 영향은 TYPE의 위험비율의 95% 신뢰구간이 1을 포함하지 않으므로 각 관로재질 간의 상대적인 위험률은 SR이 Pit-CI에 비해 3.8배 높으며, SF가 SR에 비해 3.8배 높은 것으로 분석된다. 한편 관로재질의 실제적인 위험비율은 TYPE의 시간종속 효과로 인해 시간에 따라 변하게 되는데, 매설 초기의 위험비율은 Pit-CI의 위험률이 가장 낮으나, 매설 후 시간이 경과함에 따라 SR과 SF의 위험률이 Pit-CI에 비해 상대적으로 감소하는 것으로 나타나, 매설 후 약 5년을 기점으로 위험률은  $SF < SR < \text{Pit-CI}$ 의 순서로 변화하는 것으로 추정된다. 이와 같은 현상은 매설초기에는 SR과 SF의 제작결함 및 시공불량으로 인해 SR과 SF의 파손 위험률이 Pit-CI에 비해 높으나, 이러한 매설초기의 문제점이 발현되고 난 이후에는 내구성이 보다 약한 Pit-CI 관로의 파손 위험률이 상대적으로 더 높아지는 것으로 추정된다.

Model I에서 토지개발정도(DL)의 위험비율의 95% 신뢰구간은 1을 포함하지 않으므로 관로 파손 위험률은 도시지역이 비도시지역에 비해 약 1.7배 높으며, GRID 내 급수인구(C)의 위험비율의 95% 신뢰구간도 1을 포함하지 않으므로 급수인구가 100 명 증가할 때마다 관로파손 위험률은 약 7배 증가하는 것으로 분석된다.

Model I의 관로길이(L)의 위험비율의 95% 신뢰구간이 '1'을 포함하지 않으므로 관로길이가 변함에 따라 위험률이 변화하는 것으로 볼 수 있으며 관로길이가 1,000 m 증가할 때마다 위험률은 약 1.6배 증가하는 것으로 산정되어, L의 회귀계수 추정값이 거의 '0'에 가까우나 관로의 길이가 매우 장대할 경우 관로길이의 파손에 대한 영향은 무시할 수 없는 것으로 분석된다. 한편 토지개발정도와 관로길이의 상호작용 공변수의 회귀계수도 정확한 '0'이 아니므로 위험비율의 95% 신뢰구간

도 '1'을 포함하지 않는 것으로 볼 수 있다. 따라서 도시 지역에서 관로길이가 1,000 m 증가할 때마다, 그리고 도시지역 내의 관로길이가 1,000 m 인 관로는 같은 길이의 관로가 비도시지역에 있을 경우에 비해 위험률은 약 1.3배 증가한다.

Model I에서 상호작용을 고려한 관로길이의 총체적인 위험비율은 비도시지역에서는 관로길이가 토지개발 정도의 상호작용 공변수를 제외한 L만의 위험비율인 '1.6'과 같으며, 도시지역에서는 관로길이가 1000 m 씩 증가할 때 위험비율이 약 2.1 배 증가한다. 한편 관로재질과 급수인구의 상호작용 효과로 인하여 같은 관로재질일 경우 급수인구가 100명씩 감소할수록 파손 위험률은 약 3.6배 증가하며, 급수인구가 같을 경우 SF는 SR에 비해 위험률이 0.3배이며, Pit-CI에 비해서는 0.08배이다. Model I에서 급수인구는 그 시간중속 효과에 의해 매설초기에는 급수인구가 많은 관로의 위험률이 높으나, 시간이 지남에 따라 급수인구에 의한 위험률은 감소하는 것으로 분석된다.

Model II에서는 TYPE이 0, 1 및 2일 때 각각 SR, Pit-CI 및 SF로 정의되었으며 TYPE의 위험비율은 약 1.2로 산정되어 관종별 위험률은 SR, Pit-CI, SF의 순서로 커지는 것으로 나타났다. 토지개발 정도의 위험비율은 Model I과 비슷한 값으로 산정되었으나, 관로길이가 급수인구의 위험비율은 각각 '5.5' 및 '3.5'로 Model I에 비해 높게 산정되어 두 번째 파손에 대한 위험률은 첫 번째 파손에 비해 관로길이가 급수인구의 변화에 상대적으로 민감하게 변하는 것으로 분석된다. 한편 토지개발 정도와 급수인구 사이의 상호작용 효과는 급수인구의 영향을 다소 감쇄시키는 작용을 하는 것으로 분석된다. 또한 Model I과 같이 C의 시간중속 영향이 Model II에서도 존재하는 것으로 분석되나, 그 영향은 두 번째 파손에서 더 크게 나타났다.

#### 4. 결 론

본 논문에서는 연구대상 상수관로에 파손에 영향을 미치는 인자들에 대한 시간중속성을 고려하여 순차적 파손사건에 대한 비례위험모형을 구축하였다. 그 결과 첫 번째 및 두 번째 파손 사건에 대한 비례위험모형에서 시간중속형 공변수가 적용되었다. 즉 STG I에 속하는 관로들은 파손에 대한 관로의 제원 및 연결 방식과 급수인구가, 그리고 STG II에 속하는 관로들은 파손에 대한 급수인구의 영향이 시간에 따라 변하는 것으로 나타났다. 본 논문에서 도출된 내용을 정리하면 다음과 같다.

- 1) 구축된 비례위험모형으로 부터 도출되는 각 생존 시간군의 기저위험률에 대한 분석으로부터 첫 번째와 두 번째 파손에 대해서는 대체적으로 파손 위험률이 시간에 따라 계속해서 증가하는 것으로 나타났으며, 세 번째 파손으로부터 일곱번째 파손 사건에 대해서는 파손 위험률이 파손에 따른 보수 이후 계속해서 감소하다가 시간이 경과하면 위험률이 증가하는 것으로 보인다.
- 2) 시간과 파손횟수에 따른 기저위험률의 변화 및 각 생존시간군의 중간생존시간으로부터 연구대상 상수관로들은 파손횟수가 증가할수록 전반적인 관로의 상태가 악화되는 것으로 판단된다.
- 3) 파손 횟수가 증가될수록 모형에 포함되는 공변수의 개수는 감소하며, 파손에 영향을 미치는 관로길이의 영향도 감소하여 파손횟수가 증가할수록 관로의 사용기간에 따른 노후화의 영향이 관로의 파손에 큰 영향을 끼치는 것으로 사료된다.

본 논문에서 제시된 비례위험모형은 연구대상 지역에 국한하여 사용될 수 있을 것이나, 본 논문에서 설명된 공변수의 시간중속성을 고려한 모형의 구축과정과 기저생존 및 기저위험함수를 도출하고 모형의 적합도를 검증하는 방법은 다른 연구대상 지역에도 응용될 수 있을 것으로 사료된다. 이를 통하여 상수관망의 관리자는 관로의 파손에 영향을 미치는 인자를 파악할 수 있으며, 그러한 인자의 파손에 대한 영향을 정량화할 수 있다. 또한 개별관로의 생존함수를 이용하여 관로의 파손 시간을 예측할 수도 있을 것으로 사료된다.

#### 참 고 문 헌

- 박수완, 김정욱, 전환돈 (2007). "상수도 배수관로의 특성에 따른 개별관로 정의 방법을 이용한 파손사건 사이의 비례위험모델링." **수질보전**, 한국물환경학회, 제23권, 제1호, pp. 87-96.
- 박재빈 (2006). 생존분석 이론과 실제, 신광출판사, pp. 182-216.
- Andreou, S.A., Marks, D.H. and Clark, R.M. (1987a). "A New Methodology for Modeling Break Failure Patterns in Deteriorating Water Distribution Systems: Theory." *Advances in Water Resources*, Vol. 10, pp. 2-10.
- Andreou, S.A., Marks, D.H. and Clark, R.M. (1987b). "A New Methodology for Modeling Break Failure Patterns in Deteriorating Water Distribution Systems: Applications." *Advances in Water*

- Resources*, Vol. 10, pp. 11-20.
- Allison, P.D. (1996). *Survival Analysis using SAS: A Practical Guide*, SAS Institute Inc., Cary, NC, pp. 111-183, pp. 173-174.
- Brémond, B. (1997). *Statistical Modelling As Help in Network Renewal Decision*. European commission co-operation on science and technology (COST). Committee C3 - diagnostics of urban infrastructure, Paris, France.
- Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association*, Vol. 74, pp. 829 - 836.
- Cleveland, W.S. and Devlin, S.J. (1988). "Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association*, Vol. 83, pp. 596 - 610.
- Collet, D. (2003). *Modelling Survival Data in Medical Research*, Chapman & Hall/CRC, Boca Raton, pp. 80-87, pp. 97-102, pp. 115- 116.
- Cox, D.R. (1972). "Regression Models and Life Tables." *Journal of Royal Statistic Society*, Vol. 34(B), pp. 187-220.
- Eisenbeis, P. (1994). *Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable*. Ph.D. thesis, University Louis Pasteur of Strasbourg, collection Etudes Cemagref No. 17, France.
- Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York, pp. 374-379.
- Kleiner, Y. and Rajani, B. (2001). "Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models." *Urban Water*, Vol. 3, pp. 131-150.
- Lei, J. (1997). *Statistical approach for describing lifetimes of water mains - Case Trondheim Municipality*. SINTEF Civil and Environmental Engineering, Report No. 22F007.28, Trondheim, Norway.
- Li, D. and Haims, Y.Y. (1992a). "Optimal Maintenance-related Decision Making for Deteriorating Water Distribution Systems 1. Semi-Markovian model for a water main." *Water Resources Research*, Vol. 28, No. 4, pp. 1053-1061.
- Li, D. and Haims, Y.Y. (1992b). "Optimal Maintenance-related Decision Making for Deteriorating Water Distribution Systems 1. Multilevel Decomposition Approach." *Water Resources Research*, Vol. 28, No. 4, pp. 1063-1070.
- Marks, H.D., Andreou, S., Jeffrey, L., Park, C. and Zaslavski, A. (1987). *Statistical models for water main failures*. US Environmental Protection Agency (Cooperative Agreement CR8 1 0558) M.I.T. Office of Sponsored Projects No. 94211. Boston, MA.
- Park, S. (2004). "Identifying the Hazard Characteristics of Pipes in Water Distribution Systems by using the Proportional Hazards Model: 1. Theory." *Journal of Civil Engineering, KSCE*, Vol. 8, No. 6, pp. 663-668.
- Therneau, T.M., Grambsch, P.M. and Fleming, T.R. (1990). "Martingale-based residuals for survival models." *Biometrika*, Vol. 77, pp. 147-160.

(논문번호:09-021/접수:2009.02.17/심사완료:2009.05.11)