

다변량 로지스틱 회귀분석을 이용한 증기발생기 전열관 ODSCC의 POD곡면 분석

이재봉*, 박재학[†], 김홍덕**, 정한섭**

Evaluation of the Probability of Detection Surface for ODSCC in Steam Generator Tubes Using Multivariate Logistic Regression

Jae Bong Lee, Jai Hak Park, Hong-Deok Kim and Han-Sub Chung

Key Words: Probability of Detection(검출확률), Steam Generator Tube(증기발생기 전열관), Multivariate Logistic Regression(다변수 로지스틱 회귀분석)

Abstract

Steam generator tubes play an important role in safety because they constitute one of the primary barriers between the radioactive and non-radioactive sides of the nuclear power plant. For this reason, the integrity of the tubes is essential in minimizing the leakage possibility of radioactive water. The integrity of the tubes is evaluated based on NDE (non-destructive evaluation) inspection results. Especially ECT (eddy current test) method is usually used for detecting the flaws in steam generator tubes. However, detection capacity of the NDE is not perfect and all of the "real flaws" which actually existing in steam generator tubes is not known by NDE results. Therefore reliability of NDE system is one of the essential parts in assessing the integrity of steam generators.

In this study POD (probability of detection) of ECT system for ODSCC in steam generator tubes is evaluated using multivariate logistic regression. The cracked tube specimens are made using the withdrawn steam generator tubes. Therefore the cracks are not artificial but real. Using the multivariate logistic regression method, continuous POD surfaces are evaluated from hit (detection) and miss (no detection) binary data obtained from destructive and non-destructive evaluation of the cracked tubes. Length and depth of cracks are considered in multivariate logistic regression and their effects on detection capacity are evaluated.

1. 서 론

원자력발전소 증기발생기 전열관은 방사능물질이 존재할 수 있는 1차측 배관과 방사능물질을 엄격히 통제하는 2차측 배관의 경계를 이루고 있다. 열악한 운전환경에 의해 이들 전열관에는 다

양한 형태의 결함들이 발생하고 있다. 그러므로 이 결함들의 검출 및 관리는 방사능 물질의 누출에 대한 안전관리에 있어서 필수적인 사항이 된다⁽¹⁻⁴⁾.

증기발생기 전열관에 존재하는 결함은 가동중 검사(in service inspection)에서 수행되는 비파괴검사에 의해 검출된다. 그러나 이때 수행되는 비파괴검사 결과는 결함의 검출여부와 결함의 크기에 대한 정보에 있어서 불확실성을 내포하게 된다. 즉 비파괴검사에 의한 결과만으로는 실지 전열관에 존재하는 모든 결함들을 파악하지 못한다는 것을 의미한다^(5,6). 그러므로 비파괴검사장비에 대한 신뢰도 평가는 전열관에 존재하는 결함들에

* 회원, 충북대학교 안전공학과
E-mail : jhpark@chungbuk.ac.kr
TEL : (043)261-2460 FAX :(043)264-2460

** 충북대학교 안전공학과 대학원
*** 한전전력연구원

대한 보다 정확한 정보를 도출하기 위해 필수적이다⁽⁶⁾.

본 연구에서는 구 증기발생기 전열관의 자연결함에 대한 와전류검사(ECT, eddy current test) 회송시험(round robin test) 결과와 파괴분석결과를 이용하여 국내 증기발생기 가동중검사에 사용되는 비파괴검사 시스템의 결함검출에 대한 신뢰성을 분석하였다. 와전류검사 회송시험에는 국내 3개의 검사업체에서 9개의 검사팀이 구성되어 참여하였다. 1개의 검사팀에는 3명의 조사원이 포함되어 있다. 파괴분석은 한국원자력연구소에 본 분석을 위해 새로 설치된 semi hot lab에서 수행되었다. 구 증기발생기 전열관의 인출 및 회송시험 그리고 파괴분석의 총괄은 한전전력연구원에서 수행하였다.

분석에 이용된 결함의 형태는 2차축 응력부식균열(ODSCC)이다. 로그 로지스틱 회귀분석법을 이용하여 결함깊이와 길이에 대하여 결함의 검출 확률(POD, probability of detection)을 계산하였다.

2. 결함 검출확률(POD)의 계산

일반적으로 결함의 검출에 대한 신뢰성은 POD를 이용하여 나타내며, 결함 크기의 함수로서 식 (1)과 같이 정의된다. 식 (1)에서 a는 결함의 크기를 의미한다. 실지 존재하는 결함과 검출된 결함의 수를 비율로 나타낸 것이다^(5,6).

$$POD(a) = \frac{\text{The number of detected flaw}}{\text{The number of physical flaw}} \quad (1)$$

식 (1)과 같이 정의된 POD를 연속된 함수의 형태로 나타내기 위해 많이 사용되는 것이 로지스틱함수 또는 로그 로지스틱 함수이다⁽¹⁾. 로그 로지스틱 함수의 형태를 식 (2)에 나타내었다. 식 (2)에서 β_0 와 β_1 는 회귀계수이다. 식 (2)에서 $\log(a)$ 에 log를 취하지 않고 결함의 크기, a를 그대로 사용하면 로지스틱 함수가 된다.

$$POD(a) = \frac{\exp(\beta_0 + \beta_1 \log(a))}{1 + \exp(\beta_0 + \beta_1 \log(a))} \quad (2)$$

구 증기발생기 전열관의 자연결함에 대한 ECT 회송시험결과에서는 각각의 결함크기가 설명변수

(독립변수)가 되며, 해당결함이 검출(hit)될 때 “1” 그리고 검출되지 않을 때(miss) “0”의 값이 반응변수(종속변수)가 된다. 이와 같은 이진자료(binary data)의 형태를 이용하여 식 (2)에서의 회귀계수를 구하면 POD 곡선을 도출할 수 있다. 이때 식 (2)에서의 회귀계수를 구하는 방법으로 최대우도법(maximum likelihood method)이 많이 사용된다.

2.1 로지스틱 회귀모형의 성질

일반적인 식 (3)과 같은 선형회귀모형에서 종속변수(Y/x)는 연속적인 값(continuous value)으로 평균이 $X\beta$ 이고 분산이 $\sigma^2 I$ 인 정규분포를 따른다고 가정된다. 여기서 X는 설명변수, β 는 회귀계수, σ 는 분산 그리고 I 는 항등행렬이다.

$$\begin{aligned} Y_i &= E(Y_i|x_i) + \epsilon_i \\ &= (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) + \epsilon_i \\ &= x_i^T \beta + \epsilon_i \end{aligned} \quad (3)$$

그러나 때때로 종속변수가 이진자료일 경우가 있다. 이때 종속변수는 어떤 사상이 발생할 경우 “1”, 그리고 사상이 발생하지 않을 경우 “0”으로 나타내며 연속적인 값이 아닌 이산형변수(discrete variable)가 된다. 이러한 경우 사용할 수 있는 회귀모델들 중 하나가 로지스틱회귀모형(logistic regression model)이다⁽⁷⁻⁹⁾.

식 (3)에서 종속변수 Y_i 는 두 가지 값만 가지는 Bernoulli 확률변수(random variable)로 표 1과 같은 확률분포를 가지게 된다. 표 1에서 $\pi(x_i)$ 는 독립변수들의 값이 x_i 일 때 종속변수의 값이 1일 될 확률, 즉 어떤 사상이 발생될 확률을 의미한다.

Table 1 Bernoulli random variable.

종속변수(Y_i)	확률
1	$P(Y_i = 1 x_i) = \pi(x_i)$
0	$P(Y_i = 0 x_i) = 1 - \pi(x_i)$

그리고 독립변수들의 값 x_i 에서 종속변수의 조

전부 기대값(conditional expectation)은 식 (4)와 같다.

$$\begin{aligned} E(Y_i|x_i) &= \sum_{y_i=0,1} y_i P(Y_i=y_i|x_i) \\ &= x_i^T \beta \end{aligned} \quad (4)$$

그러나 식 (4)의 모형에서는 독립변수 $-\infty$ 와 ∞ 의 영역에서 종속변수의 기대값이 0과 1사이의 값으로 나타낼 수 없다. 그러므로 $\pi(x_i)$ 를 적절히 변환할 필요가 있다. 우선 식 (5)와 같이 승산(odds)의 형태로 변환하여 종속변수의 기대값의 영역에 변화를 준다.

$$g(x_i) = \frac{P(Y_i=1|x_i)}{P(Y_i=0|x_i)} = \frac{\pi(x_i)}{1-\pi(x_i)} \quad (5)$$

여기서 승산 $g(x_i)$ 은 어떤 사건이 발생되지 않을 확률에 대한 발생될 확률의 비율을 의미한다. 승산 $g(x_i)$ 의 범위는 $0 < g(x_i) < \infty$ 이다. 그리고 식 (5)의 승산 $g(x_i)$ 에 자연대수(natural logarithm)를 취하면 종속변수의 범위는 $-\infty$ 와 ∞ 로 변하게 되며, 이러한 변환을 로짓변환(logit transformation)이라고 한다. 이때 모형은 로짓에 대하여 선형이 되고 식 (6)과 같이 나타낼 수 있다.

$$\eta(x_i) = \ln \left[\frac{\pi(x_i)}{1-\pi(x_i)} \right] = x_i^T \beta \quad (6)$$

식 (6)을 다시 $\pi(x_i)$ 에 대해 나타내면 식 (7)과 같이 되고, 이를 로지스틱 함수(logistic function)^{o)}라고 한다.

$$\pi(x_i) = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} = \frac{\exp(\eta(x_i))}{1 + \exp(\eta(x_i))} \quad (7)$$

그리고 독립변수에 자연로그 또는 상용로그를 취하여 나타낸 함수를 로그 로지스틱 함수(log-logistic function)라고 한다.

2.2 로지스틱회귀모형의 모수추정

로지스틱회귀모형에서 모수인 회귀계수를 추정

하는 데 가장 널리 널리 사용되는 방법은 최대우도방법이다.

최대우도방법에서 식 (8)의 로그우도함수를 각 회귀계수에 대하여 편미분한 함수를 0으로 하는 정규방정식을 풀면 로그우도함수 및 우도함수를 최대로 하는 회귀계수를 추정할 수 있다.

$$\begin{aligned} &\log [L(\beta)] \\ &= \sum_{i=1}^n [y_i \log(\pi(x_i)) + (1-y_i) \log(1-\pi(x_i))] \end{aligned} \quad (8)$$

식 (7)의 모델에서의 k개의 회귀계수를 n개의 관측치로부터 최대우도법을 이용하여 구하는 알고리즘은 다음과 같다. 식 (8)의 정규방정식은 비선형이기 때문에 회귀계수들을 직접구할 수 없고 반복법에 의해 근사해를 추정할 수 있다.

$$\frac{\partial \log [L(\beta)]}{\partial \beta} \text{ 을 } u(\hat{\beta}) \text{로, } \frac{\partial^2 \log [L(\beta)]}{\partial \beta_j \partial \beta_k} \text{ 을 } H(\beta)$$

로 나타내면 식 (8)의 정규방정식을 만족하는 해를 Newton-Raphson 방법으로 구할 수 있다. 이때 $H(\beta)$ 을 Hessian행렬이라고 한다.

식 (8)의 로그우도함수 β_k 에 대한 일차편미분, $u(\hat{\beta})$ 은 연쇄법칙(chain rule)에 의해 식 (9)와 같이 나타낼 수 있다.

$$\begin{aligned} &\frac{\partial \log [L(\beta)]}{\partial \beta_k} \\ &= \frac{\partial \log [L(\beta)]}{\partial \pi(x_i)} \cdot \frac{\partial \pi(x_i)}{\partial \eta(x_i)} \cdot \frac{\partial \eta(x_i)}{\partial \beta_k} \end{aligned} \quad (9)$$

식 (9)에서 $\eta(x_i)$ 은 식 (6)에서 설명한 승산의 로짓변환된 값이다. 이후 $\eta(x_i)$ 은 η_i , $\pi(x_i)$ 은 π_i 로 표기한다. 식 (9)를 다시 풀어쓰면 식 (10)과 같이 된다.

$$\begin{aligned} &\frac{\partial \log [L(\beta)]}{\partial \beta_k} \\ &= \sum \left[\frac{y_i}{\pi_i} - \frac{1-y_i}{1-\pi_i} \right] \cdot \pi_i(1-\pi_i) \cdot x_{ik} \\ &= \sum x_{ik}(y_i - \pi_i) \\ &= X^T(Y - \pi(X)) \end{aligned} \quad (10)$$

또한 로그우도함수 β_k 에 대한 이차편미분, $H(\beta)$ 은 식 (11)과 같이 구할 수 있다. 식 (11)에서 W는 가중값 $\pi_i(1-\pi_i)$ 를 대각원소로 갖는 $n \times n$ 대각행렬이다.

$$\begin{aligned} \frac{\partial^2 \log[L(\beta)]}{\partial \beta^2} &= \frac{\partial[x_{ik}(y_i - \pi_i)]}{\partial \beta} \\ &= -\frac{\partial \pi_i}{\partial \beta} \cdot x_{ik} \\ &= -x_{ik} \cdot \pi_i(1-\pi_i) \cdot x_{ik} \\ &= -X^T W X \end{aligned} \quad (11)$$

만약 $\hat{\beta}$ 을 구하고자 하는 정규방정식의 해라고 하고, β^* 를 $\hat{\beta}$ 근방의 벡터라고 할 때 $u(\hat{\beta})$ 은 다음과 같이 나타낼 수 있다.

$$\hat{\beta}_{r+1} \approx \hat{\beta}_r - H^{-1}(\hat{\beta}_r)u(\hat{\beta}_r) \quad (12)$$

식 (12)에서 r은 반복횟수이다. $\hat{\beta}_0$ 의 초기값을 식 (12)에 대입하고 반복연산을 한 후 식 (8)의 일차편미분 값이 일정오차 이하가 되면 그때의 $\hat{\beta}$ 이 회귀계수의 벡터가 된다. Hessian 행렬, $H(\beta)$ 에 -1을 곱하게 되면 정보행렬(information matrix)이 되며 $I(\beta)$ 로 표기한다. 식 (12)를 $I(\beta)$ 를 사용하여 다시 나타내면 식 (13)과 같이 된다.

$$\hat{\beta}_{r+1} \approx \hat{\beta}_r + I^{-1}(\hat{\beta}_r)u(\hat{\beta}_r) \quad (13)$$

$I(\beta)$ 의 역행렬(inverse matrix) $I^{-1}(\beta)$ 은 회귀계수의 분산공분산 행렬(asymptotic variance-covariance matrix)라고도 하며, 이를 이용하여 각 회귀계수의 분산 및 공분산을 구할 수 있다.

2.3 신뢰구간추정 및 적합도 분석

$\eta(x)$ 의 추정치 $\hat{\eta}(x) = X^T \hat{\beta}$ 의 신뢰구간을 이용하여 각 독립변수에 대하여 100(1- α)% 신뢰하한에 해당하는 검출확률의 추정치($\hat{\pi}_\alpha(x_i)$)를 구할 수 있다. 식 (14)는 $\hat{\beta}$ 의 분산으로부터 $\hat{\eta}(x)$ 의 100(1- α)% 신뢰하한 값을 계산하는 식이다. $\hat{\beta}$ 의

분산은 정보행렬, $I(\beta)$ 의 역행렬 $I^{-1}(\beta)$ 인 회귀계수의 분산공분산으로부터 알 수 있다.

$$\hat{\eta}_\alpha(x) = \hat{\eta}(x) - z_{1-\alpha} \sqrt{\widehat{Var}[\hat{\eta}(x)]} \quad (14)$$

식 (14)의 $\hat{\eta}_\alpha(x)$ 을 식 (15)에 대입하여 각 독립 변수에 대하여 100(1- α)% 신뢰하한에 해당하는 발생 확률의 추정치($\hat{\pi}_\alpha(x_i)$)를 구할 수 있다.

$$\hat{\pi}_\alpha(x_i) = \frac{\exp(\hat{\eta}_\alpha(x_i))}{1 + \exp(\hat{\eta}_\alpha(x_i))} \quad (15)$$

로지스틱 회귀분석에서 일반회귀분석에서의 SSE(sum of square for error)와 같은 의미로 회귀분석의 적합도를 측정하는 척도로 사용되는 deviance가 있다. deviance는 식 (16)을 이용하여 구할 수 있다.

$$D = -2 \ln [L(\hat{\beta})] \quad (16)$$

식 (16)에서 $L(\hat{\beta})$ 는 우도함수를 의미한다. D는 SSE와 마찬가지로 입력 자료의 수가 증가할 수록 큰 값을 보인다. 그러므로 D값을 이용하여 적합도를 상대비교 할 경우 자료의 수가 같아야 하며, 적어도 그 차이가 작아야 한다⁽⁷⁻⁹⁾.

3. ECT 검사에 의한 ODSCC 검출확률

구 증기발생기 전열판에서 자연적으로 발생된 ODSCC에 대한 와전류검사 회송시험 결과와 파괴분석결과를 이용하여 비파괴검사 시스템의 2차 측 축방향 응력부식균열의 POD 곡선을 도출하였다. 균열크기에 따른 균열검출 확률은 식 (2)를 이용하여 나타내었다. 균열의 크기는 균열의 최대깊이(TW%), 유효깊이(effective or structural depth) 그리고 유효깊이로 계산되는 파열압력(ksi)의 형태로 나타내었다. 여기서 유효깊이는 복잡한 균열의 형상을 weakest link 방법을 이용하여 가장 낮은 파열압력을 도출하는 직사각형 균열로 나타내었을 때의 깊이를 의미한다⁽²⁾. 이 때 파열압력의 계산에 사용되는 식은 EPRI에서 제시하는

ligament tearing equation^{o]} 사용된다²⁾.

Fig. 1은 균열의 크기를 최대깊이, 유효깊이 그리고 파열압력으로 나타내었을 때의 POD 곡선을 식 (2)의 로그-로지스틱 회귀모델로 분석한 결과이다. Fig. 1에 균열 검출확률의 기대치를 나타낸 POD 곡선과 95% 신뢰하한곡선을 나타내었다.

최대깊이를 이용한 POD의 경우 균열 최대깊이 15(TW%)에서 균열이 검출되기 시작하여 100(TW%)부분에서 검출확률이 1이 되고 있음을 볼 수 있으며, 유효깊이의 경우 전반적으로 최대깊이에 의해 계산된 POD보다 높은 값을 보이고 있다. 파열압력에 의한 POD는 최대깊이 또는 유효깊이에 의한 POD와의 비교가 어렵다. 파열압력의 계산에 사용되는 변수로 깊이와 길이가 함께 사용되기 때문이다.

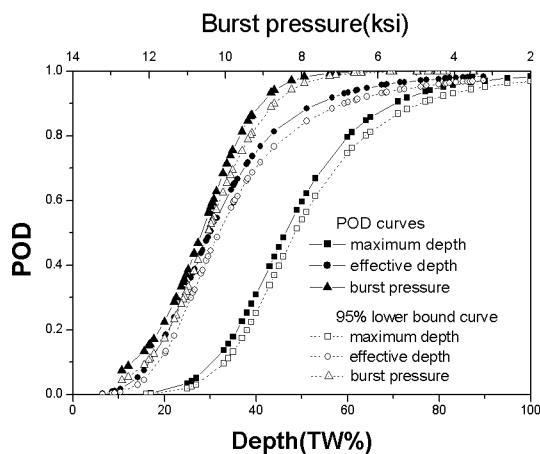


Fig. 1 POD curves and 95% lower confidence bound curves of ECT inspect system for ODSCC maximum depth, effective depth and burst pressure.

최대깊이, 유효깊이 그리고 파열압력의 값을 균열의 크기변수로 하여 계산된 각각의 POD 곡선의 적합도를 비교하기 위해 식 (16)의 D값을 표 2에 나타내었다. 표 2에서 최대깊이를 이용한 POD 곡선이 가장 좋은 적합도를 보이고 있다. 이는 최대깊이를 이용한 POD 곡선이 hit 및 miss 자료와 가장 잘 일치하고 있다는 것을 의미한다.

균열의 형상은 실지 2차원적으로 존재하고 있다. 균열의 크기를 깊이와 길이로 나타낼 때의 POD 곡면을 다변수 로그-로지스틱 회귀분석으로 구하였다.

Table 2 Logistic regression results and deviances.

# of data	Hit	Miss	Variable	Logistic Regression Results	
				E(β)	S.D.(β)
479	316	163	Constant	20.64	1.96
			Max. Depth	-12.38	1.159
			D	320.36	
470	316	154	Constant	13.04	1.34
			Eff. Depth	-8.83	0.87
			D	325.27	
470	316	154	Constant	-33.31	3.97
			Pressure	32.45	3.87
			D	342.63	

표 3에 전체길이와 최대깊이 그리고 유효길이와 유효깊이를 이용하여 다변수 로그-로지스틱 회귀분석을 수행한 결과를 나타내었다. 표 3에서의 D의 값들이 표 2에서의 D의 값들보다 작은 값을 보이고 있다. 이는 하나의 변수를 이용하였을 때보다 2개의 변수를 이용하였을 때 hit 및 miss 자료와 보다 잘 부합하는 POD 확률을 구할 수 있다는 것을 의미한다. 그리고 표 3에서 유효길이와 유효깊이를 이용하였을 때 보다 전체길이와 최대깊이를 이용하였을 때가 D의 값이 작은 것을 확인 할 수 있다.

Table 3 Multivariate logistic regression results and deviances.

# of data	hit	miss	Variable	Logistic Regression Results	
				E(β)	S.D.(β)
470	316	154	Constant	21.05	2.04
			Length	-1.20	0.39
			Max. Depth	-12.10	1.18
			D	301.60	
470	316	154	Constant	13.54	1.43
			Eff. Length	-1.13	0.49
			Eff. Depth	-8.65	0.91
			D	319.95	

전체 균열길이와 최대 균열깊이를 이용한 2-D POD 곡면을 Fig. 2에 나타내었다. Fig. 2에서 균열의 최대깊이에 대해서는 검출확률이 크게 변하는 반면 균열길이에 대해서는 변화가 완만하다는 것을 볼 수 있다. Fig. 2에서 균열의 검출확률이 균열길이 보다 최대깊이에 영향을 많이 받는다는 것을 알 수 있다.

그러나 동일한 최대깊이에서 균열길이가 변함에 따라 균열검출확률이 0.3~0.4까지 차이가 발생하는 것을 발견할 수 있다. 이는 최대균열깊이가

검출확률에 가장 크게 영향을 주고 있지만, 정확한 검출확률의 추정을 위해서는 균열길이까지 포함한 2차원적 검출확률을 고려해야 하는 것을 의미한다.

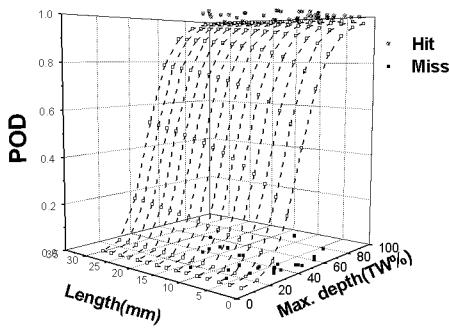


Fig. 2 The 2-D POD surface of ECT inspect system for ODSCC maximum depth(TW%) and length(mm).

4. 결 론

구 증기발생기 전열관에서 자연적으로 발생된 2차측 축방향 응력부식균열에 대한 와전류검사 회송시험결과와 파괴분석결과를 이용하여 검출확률을 분석한 결과 다음과 같은 결론을 얻었다.

1) ECT 탐상 시스템을 이용하여 구 증기발생기 전열관 ODSCC를 검출할 때의 POD 곡선 및 2-D 곡면을 로그 로지스틱 회귀분석으로 도출할 수 있었다.

2) ODSCC 검출확률은 1차원 POD 곡선의 경우 최대깊이, 유효깊이 그리고 파열압력을 변수로 하였을 때 최대깊이의 경우가 가장 적합한 POD 곡선을 도출하였다.

3) ODSCC 검출확률의 보다 정확한 추정을 위해서 1차원적인 POD 곡선보다 2차원적인 POD 곡면을 이용할 수 있음을 보였으며, 유효길이와 유효깊이를 이용하였을 때 보다 전체 균열길이와 최대 균열깊이를 이용하였을 때 적합한 POD 곡면을 도출할 수 있었다.

참고문헌

- (1) Steam Generator Integrity Assessment Guidelines: Revision 2, EPRI, Palo Alto, CA: 2006. 1012987.
- (2) Steam Generator Degradation Specific Management Flaw Handbook, EPRI, Palo Alto, CA: 2001. 1001191.
- (3) Chung, H.S., Kim, G.T., and Kim, H.D., 2000, "A Study on the Integrity Assessment of Detected S/G Tube", KEPRI.
- (4) Jae-Bong Lee, Jai-Hak Park, Hong-Deok Kim, Han-Sub Chung and Tae-Ryong Kim, 2005, "Analysis of Crack Growth Behavior Using In-situ Inspection data in Steam Generator Tubes", *Transaction of the Korean Society of Pressure Vessel and Piping*, Vol. 1, No. 1, pp. 119-126.
- (5) Jae Bong Lee, Jai Hak Park, Sung Ho Lee, Hong-Deok Kim and Han-Sub Chung, 2005, "Statistical Estimation of Crack Growth Behaviors in Steam Generator Tubes Using Monte Carlo Method", *Proceedings of 18th International Conference on Structural Mechanics in Reactor Technology*, pp. 3855-3866.
- (6) Berens, A.P., 1989, "NDE reliability data analysis", *Metals Handbook*, 9th ed., Vol. 17, pp. 689-701.
- (7) Weisberg, S., 2005, *Applied Linear Regression*, 3rd ed., John Wiley & Sons.
- (8) Collett, D., 1994, *Modelling Binary Data*, Chapman & Hall.
- (9) McCullagh, P., Nelder, J.A., 1989, *Generalized Linear Models*, 2nd ed. Chapman & Hall.