통계기법을 활용한 포장파손자료의 잠재오차 보정

Correction of Latent Errors in Pavement Deterioration Data using Statistical Methods

한대석* · 도명식**

Han, Daeseok · Do, Myungsik

Abstract

Successful implementation of infrastructure asset management system can be started with rich and reliable data. However, measurement errors in the data have always existed in the real world caused for many unknown reasons. It disturbs maintenance activities of agencies, and makes negative effects to reliability of research results on forecasting deterioration process and life cycle cost. Above all, it makes a contradiction that road agencies cannot believe their inspection data surveyed by their hands. It is particularly serious in the road pavement management field. Although road agencies are well recognized the fact, inspecting without measurement error would be a great challenge. Considering the facts, this paper aimed to suggest statistical error processing methods to correct latent error included in pavement surface inspection data. As alternatives, this paper suggested two methods based on probability distribution to consider structure of error and reliability of the data. The suggested methods were empirically tested by using pavement inspection data from Korean National Highway. As the result, this paper confirmed that conventional error processing that just removes only visible errors is not enough to cover uncertainty in pavement deterioration process. The suggested methods would be useful for improving reliability of analysis results required for road infrastructure asset management.

Keywords : road infrastructure asset management, pavement inspection data, inversed condition, latent error, korean national highway

요 지

성공적인 자산관리시스템의 도입은 풍부하고도 신뢰할 수 있는 데이터와 함께 시작된다. 그러나 현실에서는 알려지지 않은 다양한 원인에서 비롯되는 관측 오차들은 항상 존재하기 마련이다. 이는 관리자의 유지보수활동을 방해하며 파손과정이나 생애주기비용 예측결과의 신뢰성에도 악영향을 미친다. 무엇보다도 관리자가 자신의 손으로 직접 조사한 자료를 믿지 못하는 모순된 결과를 낳는다. 이런 현상은 도로포장관리분야에서 특히나 심각하다. 도로관리자들은 이러한 사실에 대해 충분히 인식하고 있으나, 현실적으로 오차 없는 조사를 수행한다는 것은 하나의 큰 도전이라고도 할 수 있다. 이러한 점들을 고려하여 본 연구에서는 포장표면 조사자료에 포함된 오류 및 잠재오차를 보정하기 위한 통계적 방법론을 제시하고자 하였다. 방법론으로는 오차의 구조 및 신뢰도를 고려할 수 있도록 확률분포이론에 근거한 두 가지 기법(샘플기준법과 분포기준법)을 제안하였다. 제안된 방법론은 국도에서 수집된 포장상태자료들을 활용하여 적용 가능성을 살펴보았으며, 그 결과 눈에 보이는 오류들만을 제거하는 일반적인 방법론으로는 포장자료에 포함된 불확실성을 충분히 고려할 수 없음을 확인하였다. 제안된 방법론은 도로자산관리에 필요한 다양한 분석의 신뢰성을 개선하는데 유용하게 활용될 수 있을 것이다.

핵심용어 : 도로자산관리, 포장조사자료, 상태역전, 잠재오차, 국도

1. 서 론

우리나라에서도 점점 도로 인프라의 건설수요가 한계에 다다름에 따라 자산관리(infrastructure asset management)에 대한 필요성이 대두하고 있다. 인프라 자산관리는 시설물의 물리적 파손과정과 증장기 생애주기비용에 대한 분석을 통해 효율적인 유지관리 전략을 모색한다는 데 일차적인 목적이 있다. 나아가 지속적인 피드백과정을 통해 최초건설부터

유지보수정책, 예산편성기준에 이르기까지 관리자가 처해있는 유지보수환경의 이질성이 고려하여 자기화(customized)된 자산관리기준을 수립하는 것이 자산관리의 최종 목적이라 할 수 있다. 이는 예산을 편성하는 의사결정자는 물론, 한정된 예산을 집행하는 관리자의 객관적인 의사결정을 돕는 의사결정시스템의 핵심이라고 할 수 있으며, 다양한 이해관계자(국민)에게 예산집행에 대한 투명성을 증명할 수 있는 중요한 수단이라고 할 수 있다(Fwa, 2006).

*정회원 · Osaka 대학 토목공학전공, 박사후연구원 · 공학박사 (E-mail : hands@civil.eng.osaka-u.ac.jp)

**정회원 · 교신저자 · 한밭대학교 도시공학과 교수 · 공학박사 (E-mail : msdo@hanbat.ac.kr)

인프라 자산관리시스템의 도입 및 운영은 장기간에 걸쳐 체계적으로 누적된 풍부한 자료가 뒷받침되어야 한다. 그러나 여기에서 “풍부한 자료”란 단지 양적인 측면만이 아닌 질적인 측면도 포함한다. 질적인 측면에서의 자료수집정책의 실패는 관리자의 유지보수 행위를 혼란에 빠뜨리며, 관리자가 자신이 자신의 손으로 조사한 자료들을 신뢰할 수 없게 만든다. 즉, 막대한 시간, 자본, 노력이 증발하게 되는 것이다.

도로포장관리분야는 자료의 오차/오류가 가장 심각한 분야 중 하나이다. 이는 단순히 도로조사 장비의 계측 오차나 측정 당시의 조사환경에서 비롯되는 측정오차에서만 비롯된다고 보기 어렵다. 여기에는 측정자의 숙련도나 개인적 판단 기준에 따라 발생하는 오차(예: 균열물의 시각조사)가 포함되기도 하며, 조사구간의 불일치, 자료가공 시에 발생하는 오류(예: 유지보수기록의 누락) 등도 포함된다. 즉, 포장분야의 오차는 알려진 “계통 오차”로 보기보다는, 다양하고 불확실한 원인이 혼재하여 발생하는 “우연 오차(accidental error)”로 인식되어야 할 것이다.

포장의 파손과정은 그에 미치는 물리적, 환경적 변수가 다양하여 불확실성이 매우 심각한 것으로 인식되고 있다. 그 불확실성 속에는 자료 내에 포함된 오류/오차도 크게 기인하고 있다. 이미 여러 선행연구에서 포장상태 자료의 불확실성, 오차에 대해 지적하고 있으며(Han 등, 2012a, 2012c; Kobayashi 등, 2012b; Madanat 등, 1994), 이러한 특성 때문에 결정론적인 방법보다는 확률통계이론에 근거한 신뢰성 분석(reliability analysis)이나 마르코프 전이확률과정(Markov transition probability process), 최대우도법(maximum likelihood estimation) 등 샘플 수에 민감한 고급확률통계기법에 주로 의존하고 있다(한대석 등, 2012; 도명식, 2010; Tsuda 등, 2006; Kobayashi 등, 2010). 한편, 이러한 샘플수 문제를 극복하기 위해 MCMC(Markov Chain Monte-Carlo)를 활용한 베이즈안통계(Bayesian statistics)나 국소혼합위험률(local mixture hazard)모형에 대한 선행연구도 찾아볼 수 있다(Han 등, 2012b; Obama 등, 2008; Train, 2009; Lancaster, 1990). 그러나 이 모든 분석에는 “조사된 자료를 신뢰한다”는 중요한 가정이 전제되어 있다. 이러한 선행연구들을 살펴보면 보통 간단한 오류보정 단계를 거쳐게 되는데, 이때의 오류는 유지보수과정 없이 상태가 개선된 것으로 조사되는 “상태역전현상”에 국한되며, 해당 자료를 샘플자료에서 제외하는 수준에서 오차를 고려한다. 그러나 이 상태역전현상의 심각성을 고려했을 때, 정상자료로 여겨지고 있는 자료의 신뢰성 또한 의심하지 않을 수 없다. 사실 현장조사원부터 자료를 활용하는 관리자, 분석가에 이르기까지 이에 관해 잘 인식하고 있으나, 원인을 찾으려는 노력이나 해결방안에 관한 연구는 그리 활발하게 이루어지지 않고 있다. 가장 최근의 연구로써 Kobayashi 등(2012b)은 잠재마르코프과정(hidden Markov process)을 통해 조사자료의 잠재오차에 대해 접근한 바가 있다. 여기에서는 실제 마르코프 전이과정과는 별도로 잠재적 전이과정에 대한 이산분포(discrete distribution)를 도입하여, “참값”에 대한 확률적 접근을 시도하고 있다. 이는 통계적, 이론적으로 완성도가 높은 연구성과라 할 수 있으나, 이산형 자료(예: 포장상태를 “건전도 1, 2, 3...” 혹은 “매우 좋음, 좋음, 보통, 나쁨, 보수요망” 등을 기준으로

구분한 자료)를 기준으로 한 보정방안에서 발생하는 단점들이 존재하며, 개개구간이 아닌 네트워크 적 측면에서의 접근만이 가능하다는 점 때문에 현실에서 발생하는 다양한 수요에 대응하는데 한계가 있다. 또한, 복잡한 이론의 이해와 프로그래밍과정 또한 쉽지 않다.

이에 본 연구에서는 포장표면조사자료에 포함된 오류와 잠재오차를 동시에 고려할 수 있고, 네트워크 뿐만 아니라 개별 샘플에 대한 오차 보정도 가능한 보정방안을 제시하고자 하였다. 본 연구에서 제시하는 보정방안은 확률분포이론에 기초하고 있으며, 오차구조에 따라 두 가지 방법으로 나누어 제시하였다. 제시된 방법론은 우리나라 국토에서 2007년부터 2010년에 걸쳐 수집된 시계열 포장상태 조사자료(총 6,420샘플)를 활용하여 적용 가능성을 확인해 보았다. 사례분석에서는 네트워크 수준에서의 보정은 물론, 마르코프 과정에서 활용하는 이산상태기준을 도입하여 포장의 파손상태에 따른 오차의 구조 및 파손특성을 고찰해 보았다. 참고로, 본 연구에서는 방법론의 제시를 주목적으로 하므로 여러 상태지표 중 균열(crack)만을 대상으로 사례분석을 진행하였다.

2. 포장표면 조사데이터의 오류/오차

2.1 오류/오차의 정의

본 연구에서 적용하고자 하는 자료는 두 시점 간 포장상태변화의 차이인 “파손량”이며, 수식으로 정의하면 다음과 같다.

$$\bar{D} = \bar{h}(\tau_B) - \bar{h}(\tau_A) \quad (1)$$

여기서, $\bar{D}$: 두 조사시점 간의 포장상태의 차이, 여기서 [-]는 조사치(확정치)를 의미

τ_A, τ_B : 조사시점 A, B(본 연구에서 두 시점의 차이는 1년)

h : 포장상태

먼저 포장파손량 자료 $\bar{D}$ 에서 발생 가능한 오류/오차에 대해 정의할 필요가 있다. 여기서 “오류”란 자료처리 과정을 통해 명확하게 판별 가능한 오류로 1장에서 언급한 상태역전현상을 말한다. 이런 경우 파손량이 음수(-)로 나타나기 때문에 눈에 보이는 오류라 할 수 있다. 이제까지의 일반적인 분석과정에서는 해당 자료들만을 삭제하여 분석자료를 구성해왔다(Han 등, 2012b, 2012c; Kobayashi 등, 2010). 그러나 현실에서의 참값은 반드시 존재하기 마련이므로, 이러한 자료들에는 한 방향(-)으로만 발생한 잠재오차를 포함하고 있다고 할 수 있다. 한편, 파손량이 (+)로 나타나는 경우 정상데이터로 여겨지게 되는데, 과연 이 자료들을 얼마나 신뢰할 수 있는가에 대한 의문을 가질 수 있다. 만약 샘플데이터 중에 오류를 나타내는 데이터의 비율이 무시할 수준, 혹은 통계적 의미에서 유의하지 않다면 나머지 정상자료들에 대한 신뢰도가 높아질 것이며, 그 반대인 경우 자료 전체에 대한 품질을 의심해봐야 할 것이다. 즉, $\bar{D}$ 가 양의 값을 가지고 있을 때에도 보이지 않는 잠재오차(+,-)를 의심해 볼 필요가 있는 것이다. 여기서 파손증가량 “0”을 기준으로 음수를 가지는 오류자료의 분포를 Q , 그의 비율은(즉, 누적분

포함수) Q 로 정의한다. 한편, 정상자료의 분포는 P 그의 비율을 P 나타내도록 하자. 즉, P 는 전체자료의 신뢰도와 관계가 깊으며, $P=1-Q$, 그리고 $P+Q=1$ 의 관계가 성립되므로 확률로서의 기본가정을 만족하게 된다(고승곤 등, 1998). 이는 파손량을 변수 $x=D(k)$ 로 한 함수 $f(x)$ 로도 표현할 수 있다. 상기 내용을 그림 1, 2와 식 (2)~(5)로 표현하였다.

$$\begin{cases} k \in P, \bar{D}(k) \geq 0, \\ k \in Q, \bar{D}(k) < 0, \end{cases} \quad (2)$$

$$P = \int_0^{\infty} f(x) dx \quad (3)$$

$$Q = \int_{-\infty}^0 f(x) dx \quad (4)$$

$$P+Q=1 \quad (5)$$

여기서, k : 샘플자료 $k(k=1, \dots, K)$

x : 파손량 $\bar{D}(k)$

P : 정상자료의 비율(분포는 P 로 칭함)

Q : 오류자료의 비율(분포는 Q 로 칭함)

$f(x)$: 분포의 확률밀도 함수

즉, P 는 좌 절단형, Q 는 우 절단형 분포의 누적분포함수(CDF: Cumulative Density Function)이며, $f(x)dx$ 는 각각의 확률밀도 함수(PDF: Probability Density Function)를 나타낸다. 이때 포장상태는 시간의 흐름에 따라 연속적으로 변화하기 때문에 연속형 분포(continuous distribution)를 따르게 되며, 우연 오차는 양 방향 (+, -)으로 항상 발생하는 것으로 가정되기 때문에 다음 조건이 성립한다.

$$\text{Prob}[\bar{D}(k)=0]=0 \quad (6)$$

그림 1. 포장파손자료의 오류/오차

그림 2. 포장파손자료의 확률구조

그러므로 조사시점 τ_B 에서의 예측치 $\hat{h}_k(\tau_B)$ 는 조건은 다음으로 정리될 수 있다.

$$\hat{h}_k(\tau_B) = \begin{cases} \{\bar{h}_k(\tau_A) + \bar{D}(k) + \tilde{e}(k) | \tilde{e}(k) > \bar{D}(k) | \tilde{e}(k) > 0\}, & \bar{D}(k) < 0 \\ \{\bar{h}_k(\tau_A) + \bar{D}(k) + \tilde{e}(k) | 0 < \tilde{h}_k(\tau_B) - \bar{h}_k(\tau_A) < \tilde{e}(k) < 0\}, & \bar{D}(k) > 0 \end{cases} \quad (7)$$

여기서,

$\hat{h}_k(\tau_B)$: 샘플의 B시점에서의 상태 예측치, 여기서 $[\hat{\cdot}]$ 는 예측치를 말함

$\tilde{e}(k)$: 샘플의 임시 보정 오차, 여기서 $[\tilde{\cdot}]$ 는 잠정치를 말함

$\tilde{e}(k)$: 샘플의 확정 보정 오차, 여기서 $[\tilde{\cdot}]$ 는 확정치를 말함

$\tilde{h}_k(\tau_B)$: 샘플의 임시 보정결과, 즉 $\tilde{h}_k(\tau_B) = \bar{h}_k(\tau_A) + \bar{D}(k) + \tilde{e}(k)$

결론적으로 $\bar{h}_k(\tau_A) < \tilde{h}_k(\tau_B)$ 를 만족하게 하는 $\hat{D}(k)$ 의 결정이 본 연구의 핵심이라 할 수 있다.

2.2 연속형 자료의 이산화에 따른 자료의 특성변화

최근 포장분야에서는 마르코프과정을 활용한 분석기법이 많이 활용/응용되고 있다(Tsuda 등, 2006; Kobayashi 등, 2010; 도명식 등, 2007; Madanat 등, 1995; Yang 등, 2006). 마르코프과정은 분석과정에서 연속상태의 원시자료를 이산형 자료로 바꾸는 과정을 필요로 한다. 우리는 이 과정에서 발생할 수 있는 다양한 변수들에 관해 오차의 관점에서 고찰해 볼 필요가 있다.

사실 마르코프 전이과정을 활용하는 경우 분포의 어느 부분을 사용하는지를 명확히 정의하기 어렵다. 연속상태에서 상태역전현상을 나타내는 자료들을 제거한 후 분석자료를 구성하면 보다 신뢰성이 확보되었지만, 일반적으로 샘플수의 최대화를 위해 이산형 자료 $i(i=1, \dots, J)$ 로 변환 후 제거하는 경우가 많다. 이런 경우 $\Delta(i, i+1)$ 사이에서 발생하는 상태역전현상은 사라지게 된다. 이는 0을 기준으로 분포 좌측의 일부분을 우측으로 강제 이동시키는 개념이기 때문에 참값분포를 활용하는 경우보다 기대수명이 짧게 나타날 가능성이 높다. 또 한 가지 중요한 특성은 이산형 자료의 변환 기준에 따라 분석결과가 달라질 수 있다는 점이다. 특히, 이산형의 흡수상태인(absorbing state) J 가 증가한계량에 대한 제한을 만들고, 그 이상의 증가치를 나타내는 자료들을 모두 흡수한다는 점이다. 한 예로 본 연구에서 발견된 샘플들 중에는 1년간 균열증가량이 60%에 달하는 샘플들도 있다. 연속형 자료를 사용 시 이런 자료들은 분석결과에 매우 큰 영향을 미치게 된다. 반면, 마르코프과정에서 이런 자료들의 가능 파손량은 샘플 k 의 τ_A 에서의 이산상태 i 의 기준값에 따라 결정된다(즉, $\Delta(i(k), J)$). 만약 τ_A 에서의 조사치가 $J-1$ 등급에 해당한다면 60%의 증가량도 단지 수 %의 증가량 $\Delta(h(J-1), h(J))$ 으로 간주된다. 즉, 이전시점 상태를 기준으로 최대증가량의 제한량이 동적으로 적용되는 동적 우절단 분포를 사용하는 것이다. 분포의 우측 절단은 분포의 좌 평균 이동을 의미하기 때문에 이는 실측분포보다는 적은 파손량, 즉 긴 기대수명을 도출할 가능성이 존재한다. 이는 분석기법의 특성이 포장상태에 따라 동적으로 이상치를 제거해주는 유익한 의미로도 해석될 수 있으나, 그것이 이상치 인지 가능치 인지는 판단하기 어렵다.

한편, 이산상태의 정의 기준에 따라 1%가 증가한 경우와

10%가 증가한 경우가 같은 파손량으로 여겨지는 경우, 반대로 10%가 증가했으므로 $h(i+1)$ 의 기준을 넘지 못하는 경우 0%로 여겨지는 경우도 발생한다. 이렇듯 샘플 자체의 오차를 제외하고도 적용상에 있어 여러가지 불확실성이 추가적으로 발생한다. 물론 샘플수가 커짐에 따라 오차를 서로 상쇄해가며 평균적인 의미에서의 결과를 도출하겠지만, 이를 위해서는 통계적으로 보다 많은 샘플이 요구된다는 단점을 갖게 된다.

3. 오류/오차 보정방안의 제시

보통 오차란 “참값”의 개념이 존재하고, 그와의 편차를 오차로 정의하곤 한다. 그러나 본 연구에서 다루고자 하는 포장의 파손량은 참값이 되어야 할 측정치에 오차가 포함되어 있어 참값의 개념이 없다. 그러므로 본 연구에서는 포장의 파손량(혹은 오차)이 확률분포를 따른다는 가정하에 해결책을 찾고자 하였다. 이에 따른 방법론으로 본 연구에서는 데이터의 오차구조를 고려한 두 가지 기법을 제안하기로 한다. 여기서 데이터의 오류구조란 분포의 침두가 0을 기준으로 어느 쪽에서 형성되는가를 의미한다. 분석의 흐름을 그림 3에 표현하였다.

그림 3에서 분석과정이 실선으로 표시된 과정을 “분포기준법”으로, 점선과정은 “샘플기준법”으로 정의하기로 하자. 분포기준법은 보정된 분포에서 샘플을 도출하는 반면, 샘플기준법은 보정된 샘플을 이용하여 새로운 분포를 도출하는 차이가 있다. 두 방법론의 선택기준은 확률구조 $P > Q$ 의 상태에 따라 결정되나, 샘플기준법의 경우 어떠한 확률구조에도 적용 가능하기 때문에 그림 3의 흐름도에 “Yes or No”로 표기하였다. 각각의 분석절차를 그림 3의 순서에 따라 설명하

기로 한다.

3.1 데이터의 생성 및 이상치 제거

먼저 두 시점차이의 상태변화량 $\bar{D}(k)$ 를 식 (1)의 정의에 따라 확보한다. 보통 포장파손량에 대해 확률분포를 도출하면 분포의 침두는 분명하지만, 그 분산이 크고 분포 양 끝의 절대값이 비현실적인 경우가 대부분이다. 이는 이상치 제거 과정의 필요성을 의미한다. 이에 대한 기준으로 통계적 의미에서 신뢰 수준을 적용할 수도 있지만, 현장에서 얻어진 경험적 지식이나 연구결과를 활용하거나 이론적 근사분포와 실제 조사치의 분포형태가 유사하도록 그 수준을 정하는 것이 바람직하다. 근사분포가 결정되면 파손량 “0”을 기준으로 P 분포와 Q 분포를 정의한다.

3.2 보정기법의 결정

본 연구에서 제시한 두 가지 방법론 중 어떤 기법을 활용할 것인지에 대한 결정과정이 필요하다. 우선 $P > Q$ 조건이 만족하면 (즉, 데이터의 신뢰성이 비교적 높으면) 분포기준법의 적용이 가능하다. 이는 적용과정이 매우 간단하여 손쉽게 활용할 수 있는 장점이 있다. 반면에, 네트워크 수준에서의 보정이기 때문에 그 활용이 탄력적이지 못하다. 반면 샘플기준법은 제약조건 없이 활용할 수 있는 장점이 있고, 개개 샘플의 조사치를 근거로 하여 보정이 이루어지므로 프로젝트 수준에서의 접근이 가능하나 보정과정이 상대적으로 복잡하다.

3.3 분포기준법

파손량의 오차분포가 정규분포에 근사하면서, $P > Q$ 경우 오류/오차의 보정에 다음의 성질을 활용할 수 있다.

그림 3. 오류/오차 보정방안의 제시

- 오류분포 Q (우절단 분포)는 정규분포에 근사하므로 좌우대칭이다.
- 잠재오차분포 P (좌절단 분포)에서는 파손량이 0에 가까울수록 오차확률이 커지게 된다.
- 잠재오차분포 P 에서 오차의 양은 0과의 거리에 비례한다.

상기 3가지 특성을 활용하면, Q 분포의 확률밀도를 우측으로 회전시켜 P 분포에서 그만큼을 상쇄해주면 참값분포 T 의 확률밀도를 얻을 수 있다. 우 회전된 (right-rotated) Q 분포를 Q' 로 표기하기로 하자. 여기서 T 분포의 PDF의 합이 1이 되도록 $T(f(x)/dx)^{-1}$ 배로 확장시켜주면 보정된 참값분포 T^* 를 얻을 수 있다. 이에 대한 개념을 그림 4와 식 (8), (9)에 표현하였다.

$$T(g(x)dx) = P - Q' = P(f(x+i)dx) - Q(f(x-i)dx) \quad (8)$$

$$T^* = \left[\int_0^\infty T(g(x)dx) \right]^{-1} = \int_0^\infty \frac{T(g(x)dx)}{\int_0^\infty T(g(x)dx)} = 1 \quad (9)$$

여기서, T : 참값분포

T^* : 보정된 참값분포

$[*]$: 분포의 PDF의 합이 1이 되도록 확장

Q' : 회전된 Q 분포

i : 급간기준($i=0, \dots, \infty$)

이렇게 보정된 분포 T^* 는 식 (6)의 조건 $\text{Prob}[\bar{D}(k)=0]=0$ 를 만족시키면서 $(0, \infty)$ 의 범위를 갖게 되고, 일반적으로 좌편향 정규분포의 형태를 띄게 된다. 분포기준법을 활용하게 되면 Q 분포의 오류는 삭제되며, P 분포내의 잠재오차는 Q 분포 만큼 보정된다. 쉽게 말해 자료의 불신도 Q 만큼 보정하는 것이다. 이는 정규분포에서는 오류의 빈도와 그 양이 대칭형태로 나타난다는 간단한 원리를 활용한 것이다. 이제까지의 과정을 통해 얻어진 것은 파손량에 대한 수정된 분포, 즉 정보일 뿐 오차가 보정된 샘플개체가 아니다. 그러나 우리는 T^* 분포의 확률밀도를 알고 있기 때문에 몬테카를로(Monte-Carlo) 기법을 통해 원하는 만큼의 파손량 $\hat{D}(k)$ 를 발생시켜 $\hat{h}_k(\tau_B)$ 를 도출할 수 있다. 이를 식으로 요약하면 다음과 같다.

$$\hat{h}_k(\tau_B) = \bar{h}_k(\tau_A) + \hat{D}(k) \quad (10)$$

$$\hat{D}(k) \sim T^*(g(x)) \quad (11)$$

$$\varepsilon(k) = \hat{D}(k) - \bar{D}(k) \quad (12)$$

여기서,

그림 4. 수정된 참값분포의 형태도출

$\hat{h}_k(\tau_B)$: 시점 B에서의 오차가 보정된 포장상태의 예측값
 $[\cdot]$: 예측값을 의미

$\hat{D}(k)$: 오차가 보정된 포장상태의 파손 증가량

$T^*(g(x))$: 수정된 참값분포 T^* 의 확률밀도 함수

$\sim$: 분포의 확률밀도를 기준으로 추출됨을 의미

여기서 얻어진 $\hat{h}_k(\tau_B)$ 를 오류와 오차가 보정된 참값으로 간주하여 분석에 활용할 수 있다. 참고로 정규분포를 따르는 샘플링 기법은 균등분포(uniform distribution)에서 추출된 두 난수 $U_1, U_2 \sim U[0,1]$ 를 박스물러 알고리즘(Box-Muller algorithm) (Box and Muller, 1958) 혹은 폴라추출법(polar method) (Bell, 1968; Knopp, 1969)을 활용하여 추출할 수 있다. 두 방법은 모두 정규분포에서 도출된 변수의 선형화가 가능하다는 점, 즉, $X \sim N(\mu, \sigma^2) \rightarrow Y = \mu + \sigma X \sim N(\mu, \sigma^2)$ 를 활용한 방법이다(한대석 등, 2012).

3.4 샘플기준법

분포기준법에서는 실측 파손량 $\bar{D}(k)$ 의 상태와는 독립으로 참값 $\hat{D}(k)$ 예측하여 오차량 $\varepsilon(k)$ 를 결정하였다. 즉, 보정은 개별샘플기준으로 이루어 지지만, 보정량은 네트워크 수준에서 분석된 결과를 참조하므로 엄밀히 말하면 네트워크 기준 보정방안이라 할 수 있다. 반면 샘플기준법은 개개 샘플의 $\bar{D}(k)$ 를 고려하여 오류/오차의 보정을 수행하게 된다는 점에서 프로젝트 레벨의 적용이 가능하다. 또한, τ_A 에서의 포장 상태 $\bar{h}_k(\tau_A)$ 기준으로 다수의 분포를 구성하여 접근하면 파손정도에 따른 오차특성 및 파손속도의 이질성에 대해 고려할 수 있다.

샘플기준법은 크게 1) Q 분포의 오류보정, 2) P 분포의 잠재오차 보정으로 나누어 진행되며, 분포 내 모든 샘플의 이동방향과 거리를 개별적으로 결정하는 과정이다. 이에 대한 개념은 그림 1에서 간단히 설명된 바 있다. 분석의 첫 단계는 분포 P, Q 의 PDF의 합이 각각 1이 되도록 확장시켜 독립된 분포로서의 조건을 만족하게 한다. $P+Q=1$ 의 조건이 만족하기 때문에 분포의 확장은 다음식을 통해 쉽게 도출된다.

$$P^*(g(x)dx) = \frac{P(f(x)dx)}{Q}, \int_0^\infty P^*(g(x)dx) = 1 \quad (13)$$

$$Q^*(h(x)dx) = \frac{Q(f(x)dx)}{P}, \int_0^\infty Q^*(h(x)dx) = 1 \quad (14)$$

다음으로 오류분포인 Q 분포의 샘플들을 하나하나 P 분포로 이동시킨다. 이때 Q 분포의 샘플들은 $\bar{x}(k)$ 를 기준으로 한 CDF, 즉 $Q^*(\bar{x}(k))$ 값을 갖게 되는데, 이 정보를 활용하여 이동할 위치를 결정한다. 여기서 주의할 점은 $Q^*(\bar{x}(k))$ 값을 P^* 분포에 대입시켜 그에 해당하는 $P^*(\bar{x}(k))$ 의 위치(점)를 찾아내는 것이 아닌, $Q^*(\bar{x}(k))$ 의 크기 만큼 P^* 분포 내에서의 위치 가능구간 $(0, \bar{x}(k))$ 을 정의한다는 개념으로 양측절단 분포 $P^*(0, \bar{x}(k))$ 를 찾아내는 것이다. 당연히 분포함수의 누적은 각 분포의 좌 극단을 기준으로 한다. 이에 대한 개념을 식 (15)~(17)과 그림 5에 정리하였다.

$$\hat{D}(k) \sim P^*(0, \bar{x}(k)) \quad (15)$$

$$P^*(\bar{x}(k)) \propto Q^*(\bar{x}(k)) \quad (16)$$

그림 5. 오류샘플의 이동가능구간 결정

$$P^*(\bar{x}(k)) = \int_0^{\bar{x}(k)} P^*(g(x))dx \propto Q^*(\bar{x}(k)) = \int_{-\infty}^{\bar{x}(k)} Q^*(h(x))dx \quad (17)$$

여기서 $\propto$ 는 비례함을 의미한다.

이 과정을 적용하게 되면 Q^* 분포의 좌 극단에 위치한 샘플(음수의 값이 가장 큰 샘플)은 P^* 분포의 좌 극단에 위치할 확률이 커지게 되며, 가장 작은 이동가능범위를 갖게 된다. 반대로 Q^* 분포의 가장 우측에 위치한 (즉, 0에 가장 가까운 음수) 샘플들은 P^* 분포의 확률밀도의 특성에 따라 어디든 위치할 수 있게 된다.

다음으로 P^* 분포 내의 샘플들의 잠재오차를 보정해야 한다. Q^* 분포 내의 샘플들이 한 방향 (+)로만 보정되는 반면, P^* 분포내의 샘플들은 P^* 분포 내에서 양 방향 (+,-)으로 이동할 가능성이 존재한다. 즉, 어느 쪽으로 이동할 것인지에 대해 판단하는 과정이 필요하다. 이를 위해 우선 $\bar{x}$ 를 기준으로 좌(-)이동할 확률 $p^*(\bar{x}(k))$ 와 우(+)이동할 확률 $q^*(\bar{x}(k))$ 를 계산한다. 이후 기준확률변수 $r(k) \sim U[0,1]$ 을 발생시켜 방향을 결정한다. 방향이 결정되면, (-)방향의 경우 $p^*(0, \bar{x}(k))$ 에서, (+)방향의 경우 $q^*(\bar{x}(k), 1)$ 구간에서 샘플을 발생시켜 이동거리를 확정한다. 수식과 그림으로 표현하면 그림 6과 수식

그림 6. 잠재오차샘플의 보정

(18)~(23)과 같다.

$$\hat{D}(k) \sim \{p^*(0, \bar{x}(k)), q^*(\bar{x}(k), 1) | r(k)\} \quad (18)$$

$$p^*(\bar{x}(k)) = \int_0^{\bar{x}(k)} p^*(f(x))dx \quad (19)$$

$$q^*(\bar{x}(k)) = \int_{\bar{x}(k)}^1 q^*(g(x))dx = 1 - p^*(\bar{x}(k)) \quad (20)$$

$$\hat{D}(k) = \begin{cases} \hat{D}(k) \sim p^*(0, \bar{x}(k)), & r(k) < p^*(\bar{x}(k)) \\ \hat{D}(k) \sim q^*(\bar{x}(k), 1), & \text{otherwise} \end{cases} \quad (21)$$

$$r(k) \sim U[0, 1] \quad (22)$$

$$\hat{\varepsilon}(k) = \hat{D}(k) - \bar{D}(k) \quad (23)$$

여기서, $r(k)$: 기준확률변수

$U[0,1]$: 표준균등분포(standard uniform distribution)

오류분포 Q 에서 이동한 샘플들과 잠재분포에서 보정된 샘플들은 확률적으로 추출된 자료이므로, 불확실성을 포함하고 있다. 이 불확실성을 제거하기 위해 상기 과정에 대해 몬테 카를로 시뮬레이션을 n 회 실시한다. 즉 샘플 k 별로 n 개의 확률적 가능 값을 갖게 되며, 각 샘플들은 고유의 증가량 분포를 갖게 된다. 이 분포의 기댓값을 $\hat{D}(k)$ 로 간주하여 자료를 보정한다. 이 샘플들을 그대로 사용하거나, 이 샘플들의 평균과 분산을 이용하여 보정된 참값분포 T 를 도출할 수도 있다. 만약 참값분포 T 의 이론치가 음수 부분을 포함한다면 이 부분을 절단한 후 T^* 로 확장한다.

4. 사례분석

사례분석을 위해 국토 전체 도로망에서 2007~2010까지 수집된 포장표면 상태자료를 활용하였다. 총 2,376개의 구간 중(2010년 자료기준) 자료결측 및 유지보수효과가 포함된 구간자료를 제외하고 총 6,420개의 샘플이 확보되었다. 3장에서 제시한 두 가지 보정방안의 적용 가능성을 모두 확인함은 물론 포장의 파손정도에 따른 오차구조 및 파손특성을 확인하기 위해 1) 전체 샘플기준 2) 1~6단계의 이산상태를 기준으로한 샘플그룹을 구성하여 분석을 진행하였다. 3장의 서술 순서에 따라 분석과정을 살펴보기로 하자

4.1 데이터 가공 및 이상치 제거

확보된 자료에는 균열률과 소성변형, 종단평탄성(IRI: International Roughness Index) 자료가 포함되어 있으나, 이중 균열률만을 분석대상으로 하였다. 이는 최근 도로관리 분야 관심이 균열에 집중되고 있는 경향을 반영하기 위함이기도 하며, 종단평탄성의 경우 포장상태의 종합지표로써 다양한 도로파손지표들의 영향을 동시에 받음은 물론이고(오차

표 1. 연간 균열률 증가량의 기초통계

항목	전체	2007~2008년	2008~2009년	2009~2010년
샘플수	6,420	2,140	2,140	2,140
샘플 중 최소증가량 (%)	-61.67	-49.29	-61.67	-38.15
샘플 중 최대증가량 (%)	60.51	22.85	45.30	60.51
평균 (%)	0.29	-0.33	0.19	1.02
표준편차 (%)	5.52	3.77	6.04	6.30

그림 7. 균열증가량의 확률분포(전체자료활용)

도 혼재), 지표간에도 상호영향을 줄 수 있기 때문에 오차의 근원을 찾아내기가 무척 까다롭다. 이에 대한 접근은 향후 연구과제로 남겨두기로 한다. 먼저 연도별 균열률 증가량에 대한 특성을 표 1과 그림 7을 통해 간단히 살펴보자.

표 1과 그림 7에서 다음 세 가지 특징을 찾아볼 수 있다. 첫째, 증가량의 확률밀도가 정규분포의 형태를 따르고 있다는 점. 둘째, 분포의 형태는 상당히 작은 구간에서 형성되고 있으나 데이터의 분산이 상당히 커 이론적인 정규분포와 그 형태가 잘 맞지 않는다는 점. 마지막으로 본 논문에서 오류로 간주하고 있는 상태역전현상이 매우 심각하다는 점이다. 이런 점들은 본 논문에서 문제 시 하고 있는 자료의 불확실성, 오류/오차를 아주 잘 보여주고 있다고 하겠다.

이런 데이터의 신뢰도 확보를 위해 가장 먼저 시행해야 할 사항은 이상치 제거작업이다. 일반적으로 이상치 제거 기준으로 3-시그마 규칙(3-Sigma rule)을 기준을 활용하는 경우가 많으나, 본 연구에서는 보다 현실적인 연구를 위해 그간 국도를 대상으로 수행된 연구결과를 참고함은 물론, 이상치 제거 후의 분포형태가 정규분포의 이론치와 유사할 수 있도록 그 기준을 정하기로 한다. 우선 동일 자료를 사용하여 포장파손속도를 예측한 기존 연구성과를 살펴보면, 균열률 15%까지의 기대수명은 약 9년(8.99년)으로 연도별 평균으로 환산하면 1.67%수준이다(Han 등, 2012b). 이 결과를 통해 우리는 연간 균열증가량의 분산이 그리 크지 않음을 유추할 수 있다. 균열증가량이 정규분포를 따른다고 가정할 때 증가량은 0 이하로 나타날 수 없는 점을 참작하면, 분포를 형성할 수 있는 최대구간, 즉 분포의 우 극단의 값은

그림 8. 균열증가량의 순차적 나열과 누적비율(전체자료활용)

3.34%가 된다. 이는 이상치 제거된 샘플의 최대-최소값의 차이가 이와 유사한 값을 가질 필요가 있다. 또 다른 방법으로는 샘플을 증가량의 크기에 따라 순차적으로 나열한 후 증가치가 안정권에 들어서는 수준을 참조할 수도 있다. 이와 관련하여 그림 8을 살펴보자.

그림 8에서는 대략 15%~85% 구간, 즉 신뢰 수준 70% 수준에서 파손량의 증가치가 안정적으로 형성됨을 볼 수 있다. 이는 1σ(68%)와 유사한 수준으로 볼 수 있다. 마지막으로, 3σ를 기준으로 실제 데이터의 분포형태와 근사정규분포의 유사성을 살펴보았다(표 2, 그림 9 참조).

먼저 표 2에서 데이터의 분포구간을 의미하는 “최대값-최소값”을 살펴보자. 이는 신뢰 수준에 따라 각각 68%, 22%, 4%를 나타내고 있으며, 기존 연구결과(Han 등, 2012b)를 참고한 분포 가능 값 3.34%를 참조할 때, 1σ 수준에서 가장 근접하고 있다. 참고로, 평균은 오류/오차가 포함된 값이므로 아무 의미가 없다. 그림 9(a)~9(c)를 비교해보면, 역시

표 2. 신뢰 수준별 이상치 제거결과(%)

항목	3σ (99.7%)	2σ (95%)	1σ (68%)
평균	0.275	0.262	0.211
표준편차	4.973	2.944	0.837
최소값	-32.680	-10.741	-1.751
최대값	34.781	11.327	2.309
최대값-최소값	67.461	22.069	4.060

그림 9(a). 신뢰 수준별 실제분포와 이론분포의 비교(전체자료활용, 신뢰 수준: 3σ)

그림 9(b). 신뢰 수준별 실제분포와 이론분포의 비교(전체자료활용, 신뢰 수준: 2σ)

그림 9(c). 신뢰 수준별 실제분포와 이론분포의 비교(전체자료 활용, 신뢰 수준: 1σ)

1σ 수준에서 실측치와 정규분포의 이론치가 거의 일치하고 있음을 확인할 수 있다. 물론 1σ 이하 수준으로 이상치를 제거하면 더욱 근사한 형태를 볼 수 있으나, 통계적 의미를 참작하여 1σ 신뢰 수준을 이상치 제거 기준으로 활용하였다.

4.2 분포기준법의 활용

1σ 수준에서 이상치가 제거된 실측 샘플들의 이론적 정규 분포를 D 로 정의하자. 정규분포의 첨두는 분포의 기대값인 $\pi(D)$ 에서 형성되며, 이 값을 기준으로 다시 P 분포와 Q 분포가 정의된다. 본 연구에서 이들의 CDF값들인 Q 와 P 는 각각 0.4003, 0.5997로 계산되었다. 이는 표면적인 자료의 신뢰 수준이 60%라 해석해도 무방하다. 식 (8), (9)의 절차에 따라 P 분포의 PDF에서 오류분포인 Q 분포의 PDF들을 빼주면, 참값분포의 확률밀도 $T(g(x)dx)$ 들이 구해진다. 이를 PDF의 합이 1이 되도록 확장시키는 절차를 거치게 된다. 이 과정을 통해 도출된 D , P , Q , T , T^* 분포를 그림 10과 11에서 비교하였다.

그림 10에서는 그림 4에서 설명한 순서에 따라 도출되는 분포들을 나타내고 있다. 예상과 같이 수정된 참값분포 T^* 는 좌 편향 분포의 형태를 띠고 있다. 또한, 기존연구결과를 기준으로 참조한 분포의 우 극단 값의 기대치가 3.34% 임을 고려하였을 때, 이에 해당하는 T^* 분포의 PDF는 0.0007 수준으로 거의 완벽하게 일치함을 확인하였다. 그림 11을 통해 분포기준법의 최종결과인 T^* 분포와 수정 전 분포인 D 분포를 비교해보면, 두 분포의 분산, 분포의 형태, 첨두의 위치, 기대치에서의 차이가 확인된다. 이는 왜 보정 절차가 요

그림 10. P , Q , T , T^* 분포의 비교

그림 11. D 분포(보정전)와 T^* (보정후)분포의 비교

그림 12. T^* 분포를 활용한 $\hat{D}(k)$ 의 샘플링

구되는지를 잘 보여주고 있다.

한편 T^* 분포에서 10,000개의 샘플을 발생시켜 분석자료를 구성해보았다. 그림 12는 이 샘플링 된 자료의 상대도수가 모 분포인 T^* 의 형태와 유사함을 보여준다. 이 분포의 기대값은 1.192%로 첨두의 약간 오른쪽에 위치하고 있다. 이를 균열률 15%까지의 도달시간으로 환산하면 12.58년으로, 기존연구결과인 8.99년(Han 등, 2012b)과 비교하면 약 3.59년의 차이가 발생한다. 이런 차이는 다양한 원인이 복합적으로 관여하고 있으나, 가장 큰 이유는 기존연구(Han 등, 2012b)에서 동적좌절단 분포에 해당하는 자료들만 사용하였기 때문이다.

4.3 샘플기준법의 활용

$P>Q$ 의 조건이 충족되지 않는 경우 샘플기준법을 활용하게 된다. 샘플기준법은 $P>Q$ 의 조건과 관계없이 적용 가능하기 때문에 분포기준법에서 사용했던 자료를 그대로 사용할 수도 있으나, 보다 의미 있는 정보제공을 위해 샘플들을 6단계의 이산상태기준으로 재구분한 자료를 활용하였다. 즉, 포장파손정도를 기준으로 한 6개의 분포에 대한 접근을 시도한다. 이를 통해 샘플기준법에 대한 적용 가능성 확인은 물론, 포장파손정도에 따른 파손속도의 이질성을 확인할 수 있다. 또한, 일반적인 오류만을 제거하는 과정을 거쳤던 기존연구결과(Han 등, 2012b)와의 비교도 가능하게 된다. 먼저 포장파손정도를 이산형 기준인 “건전도 $i(i=1, \dots, 6)$ ”로 정의하기로 한다. 자료의 분류결과를 표 3과 그림 13~15를 통해 살펴보자.

자료의 특성치를 살펴보면 각 값은 상태기준의 정의에 따

표 3. 건전도별 자료특성(균열률, %)

항목	건전도 1	건전도 2	건전도 3	건전도 4	건전도 5	건전도 6
건전도 분류기준	<0.5%	0.5%~5%	5%~10%	10%~15%	15%~20%	>20%
샘플수	1,213	2,223	654	223	89	104
최소값	-0.062	-1.257	-5.653	-10.384	-15.593	-26.848
최대값	2.312	2.962	2.960	2.314	1.602	1.336
평균	0.504	0.368	-1.093	-3.347	-6.512	-11.901
표준편차	0.595	0.989	2.140	3.826	5.621	9.604
P	0.801	0.645	0.305	0.191	0.123	0.108
Q	0.199	0.355	0.695	0.809	0.877	0.892

그림 13. 건전도별 신뢰도 P , Q 의 변화

그림 14. 건전도별 평균, 표준편차의 변화

른 제약을 받으면서 비례-반비례 관계 형성하고 있다. 먼저 표 3에서 눈에 띄는 점은 이산형 자료의 활용 시 나타나는 샘플 드롭핑(sample dropping) 현상, 즉 상태가 나빠질수록 샘플 수가 줄어드는 현상이다. 이는 분석결과와 신뢰성은 포장의 건전도에 비례한다는 의미로도 해석 가능하다. 참고로 건전도 6의 경우 가능 값의 범위가 넓어서 5단계보다 샘플 수가 약간 증가한 경향을 가졌다.

그림 13에서는 건전도에 따른 $P(i)$, $Q(i)$ 의 변화과정을 비교하고 있다. 건전도 2까지는 $P > Q$ 로 어느 정도 자료의 신뢰성이 확보되고 있으나 신뢰도 3 이후부터 급격히 감소하기 시작하여 마지막 단계의 $P(J=6)$ 는 0.108까지 그 신뢰성이 떨어지고 있다. 또한, 그림 14, 15에서는 3단계부터 평균도 양수에서 음수로 변화하였으며, 자료의 분산 또한 기하급수적으로 커지고 있음을 보여준다. 이러한 특징들은 2.2절에서 언급하였던 이산형 기준 자료분류에 따른 것으로 해석할 수 있다. 한 예로, 1단계의 경우 시점 τ_A 에서 0%~0.5%의

그림 15. 건전도별 연간 균열증가량의 분포형태 변화

균열률을 가진 자료들만을 대상으로 하므로 상태역전현상도 0.5%이내에서 만 발생한다. 즉, 1~5단계까지는 이산상태기준(0.5%, 5%, 10%, 15%, 20%)을 반영한 좌 절단 분포형태임을 의미하며, 마지막 6단계의 경우 제약이 없다(최대 100%).

샘플기준법의 순서를 다시 간략히 요약하면 1) $P(i)$, $Q(i)$ 로 독립분포구성, 2) $P^*(i)$, $Q^*(i)$ 로 분포확장, 3) $k(Q^*(i))$ 를 $Q^*(i)$ 분포로 이동시켜 오류수정, 4) $k(P^*(i))$ 를 $P^*(i)$ 분포내에서 잠재오차수정, 5) 3단계와 4단계에 대한 n 회 몬테카를로기법 적용한 샘플링과정, 6) n 샘플들의 기대치를 활용한 최종 $\hat{D}(k)$ 의 확정, 7) $\hat{D}(k)$ 를 활용한 참값분포 $T(i)$ 도출로 정리할 수 있다. 즉, 우리는 중간결과로 $Q(i)$ 의 오류보정결과와 $P(i)$ 의 잠재오차 보정결과를 각각 얻을 수 있으며, 최종적으로 참값분포 $T(i)$ 와 비교할 수 있다. 결과를 표 4에 요약하였다.

표 4가 내포하고 있는 내용은 다음으로 요약할 수 있다. 1) $Q(i)$ 의 평균증가량은 $P(i)$ 보다 항상 작으며, 그 차이는 건전도가 악화될 수록 커진다. 이는 분포 $Q(i)$ 가 항상 좌측에 위치하기 때문에 당연한 결과이다. 2) 분포의 분산(표준편차)도 건전도가 악화될 수록 커지며, $P(i)$ 가 $Q(i)$ 보다 항상 작다. 3) $T(i)$ 분포의 기대값을 참조하면 포장의 파손속도는 균열률 5% 미만(건전도 1,2)까지는 연도별 파손속도가 0.7% 수준이나 이후 빨라지기 시작하여 균열률 20% 이상에서는 이의 3배 수준인 2.2% 수준까지 증가한다. 그림 16에서는 각 건전도 별 $T(i)$ 분포의 형태를 비교하고 있다.

4.4 오차 보정방안 및 기존연구결과 간의 비교

먼저, 두 오차 보정방안에서 도출된 최종결과(T 분포)들을

표 4. 샘플기준법을 활용한 건전도별/분포별 보정결과

항목	건전도 1			건전도 2			건전도 3		
	Q(1)	P(1)	T(1)	Q(2)	P(2)	T(2)	Q(3)	P(3)	T(3)
샘플수	124	1089	1213	809	1414	2223	353	301	654
평균	0.646	0.710	0.704	0.442	0.940	0.719	0.460	1.464	0.872
표준편차	0.036	0.015	0.027	0.220	0.021	0.274	0.389	0.138	0.532
항목	건전도 4			건전도 5			건전도 6		
	Q(4)	P(4)	T(4)	Q(5)	P(5)	T(5)	Q(6)	P(6)	T(6)
샘플수	154	301	455	71	18	89	87	17	104
평균	0.764	2.108	1.901	1.023	2.785	1.943	1.768	4.688	2.246
표준편차	0.626	0.059	0.734	0.826	0.048	1.024	1.450	0.150	1.714

그림 16. 건전도별 T분포의 비교

그림 17. 샘플기준법과 분포기준법의 보정결과 (T*분포)비교

비교해 볼 필요가 있다. 그림 17을 살펴보면, 두 가지 보정 방법 모두 다 좌 편향 정규분포의 형태를 띠고 있으며, 침 두위치나 분산 등 상당히 유사함을 확인할 수 있다. 다만,

비고: 분포기준법은 선형증가를 가정함

그림 18. 기존연구결과와 제안 모델 간의 기대수명 비교

샘플기준법의 침두부분의 PDF가 좀 더 높고, 분포 우측 부분의 분산이 약간 작음을 감안할 때, 파손량이 상대적으로 작게 예측될 가능성이 있다. 이는 좌 절단형 $P(i)$ 분포의 PDF가 0에 가까울수록 높은 형태를 갖기 때문이다.

표 5와 그림 18에서는 1) 기존연구결과(Han 등, 2012b), 2) 샘플기준법활용결과, 3) 분포기준법 활용결과를 파손속도와 기대수명을 기준으로 비교하고 있다. 예상한 바와 같이, 본 연구에서 제시한 두 방법론을 활용한 결과가 기존연구에서 도출된 기대수명보다 길게 예측되었다. 이는 조사오류로 간주되어 그동안 활용되지 않았던 상태역전현상에 대한 오차 보정에 기인한 것이다. 또한, 건전도 1,2단계에서는 오류/오차보정에 따른 예측결과의 차이가 심각하지 않으나, 상태가 악화될 수록 그 영향이 커지는 것이 나타났다. 한편, 샘플기준법과 분포기준법에 따른 기대수명은 각각 15.34년,

표 5. 보정기법에 따른 건전도별 파손속도 및 기대수명비교

건전도	상태기준 (균열률)	파손속도 (%)			기대수명 (년)		
		샘플기준법	분포기준법	기존연구	샘플기준법	분포기준법	기존연구*
1	<0.5%	0.70	1.19	0.33	0.71	0.42	1.53
2	0.5~5%	0.72		1.19	6.26	3.77	4.19
3	5~10%	0.87		2.20	5.73	4.19	2.27
4	10~15%	1.90		5.00	2.63	4.19	1.00
5	15~20%	1.94		-	2.57	4.19	-
6	>20%	2.25		-	2.23	4.19	-

비고: *베이저안 마르코프 위험률(Bayesian Markov Hazard)모형 (Han 등, 2012b)

12.58년으로, 샘플기준법에 따르는 경우가 약간 길게(2.76년) 예측되었다. 이 차이는 그림 17에서 비교한 두 참값분포의 형태를 참조하면 쉽게 이해할 수 있다.

본 연구에서는 P , Q 의 조건에 따라 두 가지 안을 제시했으나, 비교 가능한 참값이 없어서 어느 방안을 더 신뢰할 수 있는가는 판단하기 어렵다. 다만, 두 가지 방법론이 다른 개념으로 오차보정에 접근하고 있음에도 상당히 유사한 결과를 보여주고 있는 점은 상호비교적 측면에서 신뢰할 수 있다고 판단된다.

5. 결 론

본 연구에서는 도로자산관리에 있어 근간이 되는 상태변화 자료의 신뢰성을 높이기 위한 통계적 보정방안을 제시하고자 하였다. 본 논문에서는 포장상태 조사자료가 눈에 보이는 오류인 상태역전현상, 그리고 보이지 않는 잠재오차를 포함하고 있다고 가정하고 있으며, 이에 대한 보정방안으로 1) 분포기준법과 2) 샘플기준법을 제시하였다. 두 방법론은 우연 오차는 확률분포에 근거한다는 가정하에 고안되었으며, 자료의 신뢰 수준 및 활용목적(네트워크기준, 구간기준)에 맞추어 적용할 수 있도록 설계되었다. 한편, 포장데이터의 특성에 적합한 이상치 제거방안을 제시하였으며, 최근 마르코프 이론의 적용사례가 점점 확대되고 있는 점을 감안하여 연속형 자료의 이산화 과정에서 발생하는 오류/오차의 가능성을 자료구성의 관점에서 고찰하였다. 제시된 방법론은 한국국도에서 2007년부터 2010년까지 수집된 균열률의 시계열 자료를 통해 그 적용 가능성을 살펴보았다. 사례분석을 통해 얻어진 주요결과를 요약하면 다음과 같다.

먼저, 파손량이 음수를 띄는 자료들이 전체의 40%를 차지할 정도로 상태역전현상이 매우 심각함을 확인하였다. 이는 분석에 앞서 반드시 오류에 대한 제거과정이 필요하고, 나아가 정상자료로 인식되는 자료들까지 그 신뢰성을 의심해봐야 할 필요성을 의미한다. 다음으로 파손량 자료의 확률분포는 상대적으로 작은 구간에서 형성되나, 그 분산이 매우 크다는 점을 확인하였다. 이에 본 연구에서는 기존연구결과를 활용한 분포의 위치정보(평균과 분산), 증가량의 변화추세특성, 마지막으로 실측자료의 이론적 분포와의 유사성을 고려하여 이상치 제거를 시행하였다. 그 결과 정규분포의 1시그마 수준인 68% 수준에서 이상치를 제거하는 것이 가장 적합한 것으로 분석되었다.

수집된 자료에 샘플기준법을 활용한 결과 참값분포는 좌편향 정규분포의 형태로 변형되었으며, 이 분포의 기대값을 활용한 기대수명은 12.58년으로 같은 자료를 사용한 기존연구결과인 9년보다 약간 길게 예측되었다. 이러한 차이는 기존연구에서 상태역전 현상이 포함된 자료만을 제거하면서 파손량이 적을 가능성이 높은 자료들이 분석에서 배제되었기 때문이다.

한편 이산상태를 기준으로 구분된 6개의 분포($P > Q$ 와 $P < Q$ 가 혼재)를 구성하여 각각의 분포형태의 특징을 살펴본 결과 포장의 파손 정도에 따라 파손속도 및 오차구조에 이질성이 존재함을 확인하였다. 각 분포에 샘플기준법을 적용하여 오류 및 잠재오차를 보정한 결과 포장의 파손속도는

균열률 5% 미만(건전도 1, 2)까지는 연도별 파손량이 0.7% 수준이나 이후 빨라지기 시작하여 균열률 20% 이상에서는 이의 3배 수준인 2.2%까지 증가함을 알 수 있었다.

마지막으로 두 보정기법의 최종 보정결과를 비교해 본 결과 접근방법이 확연히 다름에도 분포의 특성(침투와 분산)이 거의 일치함을 확인하였으며, 여기서 도출된 기대수명들은 보이는 오류만을 제거한 후 수행되었던 기존연구보다 상대적으로 긴 기대수명을 도출하였다. 본 논문에 나열된 결론들은 포장의 공용성 분석에 앞서 자료의 신뢰성에 대한 접근이 반드시 필요함을 말해주고 있다.

본 연구에서는 도로관리자 또는 분석가들 사이에서 항상 논란의 대상이었던 “상태역전현상”, “관측자료의 신뢰성”에 대한 문제에 대한 사후대책을 논의하였다. 그러나 도로관리자들은 이러한 보정방법을 적용하기 이전에 왜 자료에 오류/오차가 발생했는지에 관한 원인을 반드시 찾아내야 할 필요가 있다. 데이터의 오차구조, 현장경험, 그리고 포장상태조사 장비의 첨단화 등을 감안할 때, 포장데이터의 오류/오차는 조사구간의 불일치에서 오는 경우가 가장 크다고 사료된다. 이에 대한 해결책으로 GIS/GPS를 활용한 도로정보의 시각화/디지털(digital)화가 시급하다.

감사의 글

본 논문에서 사용된 포장상태 자료는 국도 포장관리를 담당하는 한국건설기술연구원에서 제공받았음을 밝히며, 연구수행을 위해 도움을 주신 한국건설기술연구원 관계자 분들께 감사의 말씀을 전합니다.

참고문헌

- 고승곤, 양완영(1998) **일반통계학**, 교우사.
- 도명식(2010) 신뢰성 개념을 이용한 포장의 평균수명 및 신뢰도 예측, **대한토목학회 논문집**, 대한토목학회, 제30권 제5D호, pp. 497-504.
- 도명식, 이종달, 김주현(2007) **SOC계획과 시스템 분석**, 대가.
- 한대석, 도명식(2012) 도로포장 유지보수 전략에 따른 기대수명과 보수비용산정, **대한토목학회 논문집**, 대한토목학회, 제32권 제4D호, pp. 345-356.
- Bell, J. R. (1968) Algorithm 334: normal random deviates, *Communications of the Association for Computing Machinery (CACM)*, Vol. 11, No. 7, pp. 498-499.
- Box, G.E.P. and Muller, M.E. (1958) A note on the generation of random normal deviates, *The Annals of Mathematical Statistics*, Vol. 29, No. 2, pp. 610-611.
- Fwa, T. F. (2006) *The handbook of highway engineering*, Taylor & Francis Group, LLC, London.
- Han, D. and Kobayashi, K. (2012a) Criteria for the development and improvement of PMS models, *KSCE J. of Civil Engineering* (In review).
- Han, D., Kaito, K., and Kobayashi, K. (2012b) Application of Bayesian estimation method with Markov hazard model to improve deterioration forecasts for infrastructure asset management, *KSCE J. of Civil Engineering* (In review).
- Han, D., Kobayashi, K., and Do, M. (2012c) Section-based calibration method for pavement deterioration forecasting model, *KSCE J. of Civil Engineering* (Forthcoming).
- Knopp, R. (1969) Remark on algorithm 334 [G5]: normal random

- deviates, *Communications of the Association for Computing Machinery (CACM)*, Vol. 12, No. 5, pp. 275-281.
- Kobayashi, K., Kaito, K., and Nam, L.T. (2012a) A Bayesian estimation method to improve deterioration prediction for infrastructure system with Markov chain model, *International Journal of Architecture, Engineering and Construction*, Vol. 1, No. 1, pp. 1-13.
- Kobayashi, K., Kaito, K., and Nam, L.T. (2012b) A statistical deterioration forecasting method using hidden Markov model for infrastructure management, *Transportation Research Part B, Elsevier*, Vol. 46, pp. 544-561.
- Kobayashi, K., Do, M., and Han, D. (2010) Estimation of Markovian transition probabilities for pavement deterioration forecasting, *KSCE J. of Civil Engineering*, Vol. 14, No. 3, pp. 341-351.
- Lancaster, T. (1990) *The Econometric Analysis of Transition Data*, Cambridge University Press, New York.
- Madanat, S. and Ben-Akiva, M. (1994) Optimal inspection and repair policies for infrastructure facilities, *Transportation Science*, Vol. 28, No. 1, pp. 55-62.
- Madanat, S., Mishalani, R., and Ibrahim, W. (1995) Estimation of infrastructure transition probabilities from condition rating data, *J. of Infrastructure Systems, ASCE*, Vol. 1, No.2, pp. 120-125.
- Obama, K., Okada, K., Kaito, K., and Kobayashi K. (2008) Disaggregated hazard rates evaluation and bench-marking, *J. of Civil Engineering, JSCE*, Vol. 64, No. 4, pp.857-874 (In Japanese).
- Train, K.E. (2009) *Discrete Choice Methods with Simulation (Second edition)*, Cambridge University Press, New York, USA.
- Tsuda, Y., Kaito, K., Aoki, K., and Kobayashi, K. (2006) Estimating Markovian transition probabilities for bridge deterioration forecasting, *J. of Structural Engineering and Earthquake Engineering, JSCE*, Vol. 23, No. 2, pp. 241-256.
- Yang, J., Lu, J.J., Gunaratne, M., and Dietrich, B. (2006) Modeling crack deterioration of flexible pavements: Comparison of recurrent Markov chains and artificial neural networks, *J. of Transportation Research Board, ASCE*, No. 1974, pp. 18-25.

(접수일: 2012.6.26/심사일: 2012.8.20/심사완료일: 2012.9.10)