한국철도학회논문집 제19권 제4호 pp. 547-554 (2016년 8월)
JOURNAL OF THE KOREAN SOCIETY FOR RAILWAY VOL.19, NO.4 pp. 547-554 (August 2016)
ISSN 1738-6225(Print)
ISSN 2288-2235(Online)

궤도틀림 진전율 추정을 위한 베이지안 회귀분석 모형 연구

A Bayesian Regression Model to Estimate the Deterioration Rate of Track Irregularities

박범환*
Bum Hwan Park

Abstract This study considered how to estimate the deterioration rate of the track quality index, which represents track geometric irregularity. Most existing studies have used a simple linear regression and regarded the slope of the regression equation as the progress rate. In this paper, we present a Bayesian approach to estimate the track irregularity progress. This Bayesian approach has many advantages, among which the biggest is that it can formally include the prior distribution of parameters which can be derived from historic data or from expert experiences; then, the rate can be expressed as a probability distribution. We investigated the possibility of applying the Bayesian method to the estimation of the deterioration rate by comparing our bayesian approach to the conventional linear regression approach.

Keywords : Track irregularity, Track quality index, Linear regression, Bayesian method

초 록 본 연구는 궤도 틀림을 관리하기 위한 궤도 품질 지수(TQI)의 진전율 추정에 관한 것이다. 이와 관련한 기존 연구 대부분은 시간에 따른 TQI 값의 선형 회귀분석을 통해 구해진 기울기를 기준으로 상수 진전율을 제시하는 데 그치고 있다. 본 연구는 과거 데이터 혹은 전문가의 식견으로부터 도출되는 파라미터의 사전 분포를 효과적으로 반영할 수 있으며, 파라미터값의 확률 분포를 유도해 낼 수 있는 베이지안 방법론에 기초한 진전율 추정 모델을 제안하고, 기존의 전통적인 회귀분석 모형과의 비교 연구를 통해, 베이지안 방법론의 활용 가능성을 검토해 보았다.

주요어 : 궤도틀림, 궤도품질지수, 회귀분석, 베이지안 방법론

1. 서 론

궤도틀림(track irregularity)은 열차의 주행 안전성 및 승차감에 절대적인 영향을 미치는 요인일 뿐만 아니라[1]. 전체 철도 시스템의 유지보수 비용의 상당부분을 궤도 유지보수 비용이 차지할 정도로, 철도 시스템 유지보수에 있어 최우선 순위의 관리 대상이다. 궤도틀림은 궤간(gauge), 줄(방향, alignment), 면(고저, surface 혹은 longitudinal level), 수평(cross level, cant), 평면성(twist) 틀림 등으로 구분되는데, 유지보수 측면에서 줄틀림, 면틀림이 가장 중요한 관리 대상으로 거론된다. 예를 들어, 유럽 표준 EN 13848-5는 예방보수의 한도(Alert Limit)값으로 면틀림과 줄틀림의 표준편차 값을 추천하고 있다[2].

국내에서 사용하는 고속철도의 궤도틀림 관리 기준으로는 절대틀림과 표준편차를 활용하는데, 궤도의 유지보수는 표준편차보다는 절대틀림 한계값에 기초하여 이루어지고 있다[1]. 그러나 절대틀림에 기초한 궤도 관리는 특정 지점의 틀림량이 한계값을 넘느냐 그렇지 않느냐에 따른 사후보수(corrective maintenance) 방법으로, 궤도 관리를 보다 효과적으로 시행하기 위해서는 궤도 상태를 표현할 수 있는 지표를 개발하고 그것의 진전율을 계산할 수 있어야 한다. 이러한 궤도 상태를 나타내는 지표를 궤도품질지수(Track Quality Index, 이하 TQI)라 한다.

여러 가지 통계량이 TQI로 활용될 수 있는데, 틀림량이 절대 한계값을 넘는 개수의 비율을 의미하는 P값(P value), 틀림량들의 표준편차, 평균편차[1,3] 등이 가장 일반적이며, 이 외에도 미연방 철도국(FRA)에서 활용하는 궤도의 실측길이를 활용한 지표가 있다[3,4]. 그러나 위에서 서술한 모든 TQI들은, 물리적인 요인에 의해 정확히 계산해 낼 수 있는 확정적인 지표가 아니고, 설명할 수 없는 요인에 의한 변동성에 의해, 확률 분포를 가정할 수 밖에 없는 확률 변수로 간주된다는 공통점을 갖고 있다[5].

*Corresponding author. Tel.: +82-70-8855-1638, E-mail: beomi72@hanmail.net.
© 2016 The Korean Society for Railway. All rights reserved.
http://dx.doi.org/10.7782/JKSR.2016.19.4.547

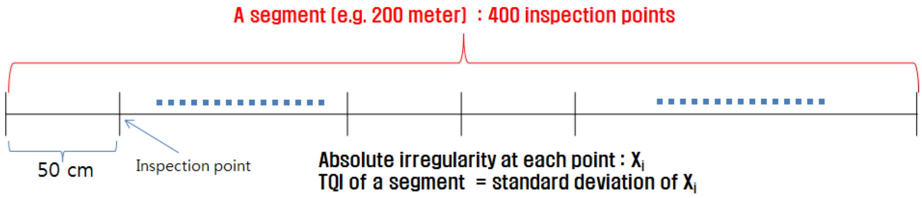

Fig. 1. Segment, inspection points, and TQI.

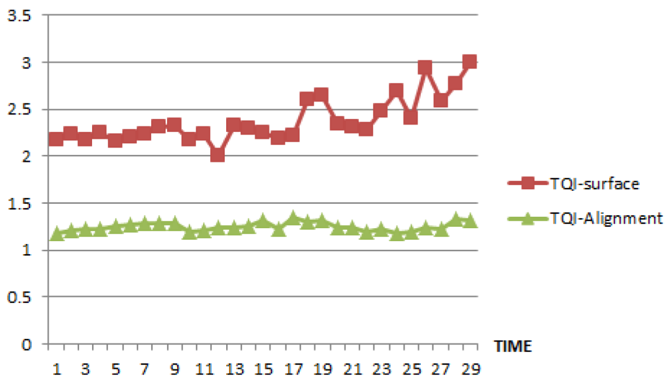

Fig. 2. TQI-surface and TQI-alignment.

TQI를 측정하는 궤도의 단위를 설정하기 위해서는, 구성품의 종류, 구조물의 유무, 도상의 종류 등 궤도의 속성이나 특성 별로 동질성(homogeneity)이 확보될 수 있는 있는 단위로 구분되어야 하는데 이를 세그먼트(segment)라고 한다. 세그먼트를 어떻게 나누어야 하는지에 대한 정형화된 방법은 없지만, [6]에서 제시된 것처럼 시설 구조(토공, 교량, 터널 등), 궤도 장치(체결구, 이음매 유무 등) 등 다양한 기준에 의해 세그먼트를 구분할 수 있으며, 하나의 세그먼트 길이는 짧게는 수 백미터, 길게는 수 키로미터에 이를 수 있다. 아래 그림은 200미터 길이의 세그먼트에 대해, 궤도검측차로부터 50 cm단위로 수집된 궤도틀림량(X_i)을 이용하여 총 400개의 틀림량데이터들의 표준편차를 TQI로 설정하는 경우를 그림으로 표현한 것이며, Fig. 2는 면틀림과 줄틀림 TQI값들을 시계열 형태로 배열했을 때의 그래프를 보여준다.

궤도틀림 진전율(deterioration rate)이란 Fig. 2에서처럼 나타나는 TQI값이 시간에 따라 얼마나 악화되느냐를 나타내는 지표인데, 현재까지 발표된 궤도틀림 진전율과 관련한 거의 모든 국내 연구는 TQI로 표준편차를 고려하고, 그 표준편차의 시간에 따른 선형적 증가를 가정한 단순 선형 회귀분석(simple linear regression) 모형에 기초하고 있다[6-10]. 본 연구는 기존의 전통적인 회귀분석 모형의 문제점을 살펴보고, 베이지안 회귀분석과 어떤 차이가 있는지를 비교 분석함으로써, 베이지안 방법론의 적용 가능성을 검토하는데 목적이 있다.

2. 본 론

2.1 기존 진전율 연구 분석

현재까지 발표된 궤도틀림 진전율과 관련한 거의 모든 국내 연구는 TQI로 표준편차를 고려하고, 그 표준편차의 시간에 따른 선형적 증가를 가정한 아래와 같은 단순 선형 회귀분석(simple linear regression) 모형에 기초하고 있다[6-10].

$$TQI_i = \beta_0 + \beta_1 \cdot t_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

즉 i번째 TQI값은 알려져 있지 않지만 어떤 값을 갖는 파라미터(β_0, β_1)에 의해 결정되며, 이 때의 오차항(ε_i)는 평균이 0이고 분산이 σ^2인 정규분포를 따른다고 가정한다. 선형회귀분석에서는 최소자승법(least square)을 이용하여 파라미터 β_0, β_1을 추정하며, 오차항의 분포를 이용하여, 회귀분석모형의 적합성(F검정, t검정)을 판단하게 된다. 기존의 많은 연구[6-10]에서 다양한 선형회귀 결과를 제시하고 있지만, 실제 회귀 모형의 적합성에 관련된 통계량은 제시하고 있지 않은데, 일반적으로 궤도 유지보

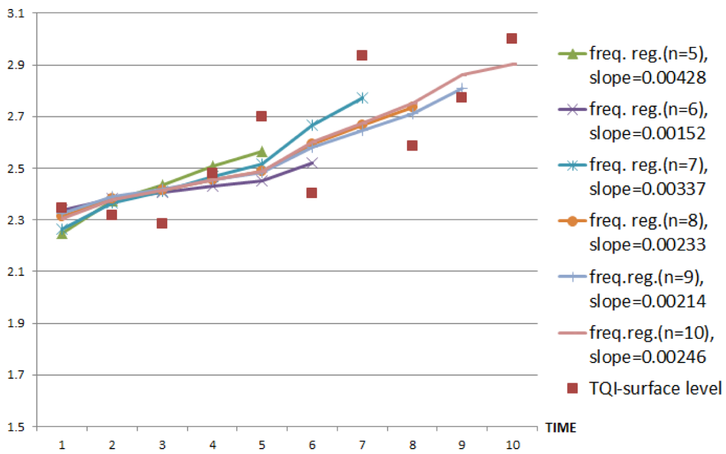

Fig. 3. Estimating the deterioration rate based on conventional linear regression.

수가 일어나는 두 시점 사이의 TQI 데이터는 매우 제한적이므로, 이러한 데이터 개수의 한계 및 데이터 자체의 변동성으로 인해, 선형식은 도출되지만 모형 적합성은 매우 낮기 때문인 것으로 추측된다. 선형 회귀식을 통해 대강의 진전율을 확인할 수 있을 뿐, 그 진전율의 통계적 적합성은 매우 떨어진다고 볼 수 있다.

뿐만 아니라 TQI값의 경우, 유지보수 이후 TQI값이 일관되게 증가하거나 감소하지 않고, 제한된 개수의 TQI값만이 존재하여 선형 회귀분석을 할 경우 현재 가지고 있는 샘플 데이터에 심하게 의존하는 결과를 도출할 수 있다는 큰 문제점이 있다. Fig. 3은 국내 고속철도 자갈궤도의 200미터 구간에서 9개월간의 면틀림 TQI값을 기초로 관측값 개수(n)의 추가에 따른 선형회귀분석 결과를 도시한 것이다. 먼저 사각형 점(“TQI-surface level”)은 면틀림 TQI값을 의미하고, n=5부터 n=10까지 샘플을 하나씩 더 추가했을 때, 회귀계수 β_1(진전율)값을 기울기로 나타낸 것이다. 그림에서 보듯이 β_1의 값이 n=5일 때, 0.00428(mm/day)에서 n=6일 때, 0.00152(mm/day)로 대폭 감소됨을 볼 수 있으며, 그림에서 보듯이 새로운 TQI값이 추가될 때 마다 진전율은 큰 폭으로 변화됨을 볼 수 있다. 새로운 검측데이터를 통해 진전율을 개선해가며 유지보수 계획을 작성해야 하는 측면에서 보면, 이러한 큰 폭의 변동성을 갖는 진전율의 관리는 의미가 없다.

유지보수가 없는 기간 동안의 평균적인 진전율은 그동안 축적된 비슷한 세그먼트 구간의 진전율을 통해 평균적인 진전율의 움직임이 예상 가능하다. 따라서 특정 구간에 발생하는 변동성은 그 변동성만을 모형에 반영하지 않고, 비슷한 구간들의 평균적인 진전율 혹은 전문가의 판단을 동시에 고려하여 진전율을 예측하는 것이 보다 합리적이다. 다음 절에서는 과거의 축적된 경험과 특정 구간에서의 데이터를 함께 반영하고, 회귀분석모형상의 회귀계수를 고정된 값이 아닌 확률 변수로 간주함으로써, 진전율에 대한 확률분포를 유도할 수 있는 베이지안 회귀분석 모형에 대해 살펴본다.

2.2 베이지안 회귀분석 모형

2.2.1 베이지안 모델

앞에서 살펴본 일반적인 선형 회귀분석에서의 회귀 계수는 알려져 있지 않을 뿐, 최소자승법과 같은 방법을 이용하여 샘플들로부터 추정해야 하는 고정 값으로서, 회귀 계수 자체의 확률적 요소를 고려하지는 않는다. 베이지안 분석에서는 기존의 통계학에서 샘플들로 추정해야 하는 파라미터가 하나의 확률 분포를 갖는 확률변수로 간주하고, 관측값을 통해 파라미터의 확률분포를 좀 더 정확히 추정하고자 하는 방법론으로 생각할 수 있다. 베이지안 분석 방법론은 다음과 같은 간단한 수식으로 표현될 수 있다.

$$f(\theta|x)=\frac{f(x|\theta)\cdot f(\theta)}{f(x)}\propto f(x|\theta)\cdot f(\theta)$$

여기서, θ는 파라미터, x는 관측값, $f(\theta)$는 파라미터 θ에 대한 사전분포(prior distribution), $f(x|\theta)$는 우도함수(likelihood), $f(\theta|x)$는 사후분포(posterior distribution)를 의미한다. 예를 들어, θ가 회귀분석 모형에서의 회귀계수이고 x가 관측값이라면, $f(\theta)$는 과거의 데이터 혹은 전문가의 식견을 통한 θ의 분포를 의미하고 이것을 이용하여 사후분포 $f(\theta|x)$, 즉 관측값이 주어졌을 때, 변화된 θ의 분포(회귀계수의 분포)를 계산하는 것을 의미한다.

그런데 이러한 사후분포를 찾기 위해서는 사전분포와 우도함수가 계산이 용이하여 닫힌 형태(closed form)형태의 확률분포를

유도할 수 있으면 좋지만, 현실 모형의 경우 다변량 분포를 고려하거나 복잡한 형태의 확률밀도함수를 이용한 모델의 경우 시뮬레이션에 의한 표본 추출 및 그 표본 값들의 평균, 중간값 등으로 분포를 간접적으로 알아낼 수 밖에 없다. 분포로부터 추출한 표본으로부터 확률분포의 통계량을 도출하는 방법론을 몬테카를로 시뮬레이션(Monte Carlo Simulation)이라 하는데, 본 연구에서는 이 중에서도 안정 상태의 상태(state)값들이 표본값으로 수렴하는 마코프 연쇄 몬테카를로 시뮬레이션(Markov Chain Monte Carlo: 이하 MCMC) 방법[11]을 활용할 것이다.

2.2.2 TQI 진전율 예측을 위한 베이지안 회귀분석 모델

n개의 궤도 품질 지수(TQI_i)와 그 때의 시간(t_i)이 주어져 있고($i = 1,2...,n$), 아래와 같은 선형관계에 있다고 가정하자.

$$TQI_i = \beta_0 + \beta_1 t_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

β_0, β_1, σ^2이 주어져 있을 경우 TQI_i의 확률 분포, 즉 TQI_i의 우도 함수(likelihood)는 아래와 같은 정규분포를 갖는다.

$$TQI_i | \beta_0, \beta_1 \sigma^2 \sim N(\beta_0 + \beta_1 t_i, \sigma^2)$$

앞에서 서술했듯이 베이지안 회귀분석에서는 β_0, β_1, σ^2가 사전분포를 갖는 확률변수로 간주된다. 사전분포를 어떤 분포로 하느냐에 따라 베이지안 회귀분석 결과는 다양한 형태로 나타날 수 있다.

사후분포가 사전분포와 동일한 분포를 갖도록 만드는 공액사전분포(conjugate prior)와 사후분포에 대해 최소한의 임무만 수행하는 무정보적 사전분포(noninformative prior)등 다양한 형태의 사전분포를 고려할 수 있다. 특히 [2]에서는 독립변수를 누적통과톤수로 했을 경우, (β_0, β_1)는 이변수 로그 정규분포로 가정하였고, σ^2은 역감마(inverse gamma) 분포를 이용하였다.

일반적으로 고려할 수 있는 베이지안 회귀분석 모형은, β_0와 β_1이 정규분포를 이루고, 두 파라미터의 상관관계가 존재하여(β_0, β_1)의 사전분포를 이변수 정규분포(bivariate normal distribution)로 가정하는 것이다. 만약 σ^2이 역감마 분포를 따른다고 가정하면, 아래와 같은 수식 전개에 의해, (β_0, β_1)와 σ^2의 결합 사후 확률 분포를 계산할 수 있다.

-β=(β_0, β_1) ~ $N_2(\bar{\beta}, \Sigma)$ 여기서, $\bar{\beta} = (\bar{\beta}_0, \bar{\beta}_1)$는 평균, $\Sigma = \begin{pmatrix} V(\beta_0) & cov(\beta_0, \beta_1) \\ cov(\beta_1, \beta_0) & V(\beta_1) \end{pmatrix}$는 분산-공분산행렬

-$1/\sigma^2 \sim Gamma(a, b)$

$$-f(\beta_0, \beta_1, \sigma^2 | TQI) \propto f(TQI | \beta_0, \beta_1, \sigma^2) \cdot f(\beta_0, \beta_1, \sigma^2) = f(TQI | \beta_0, \beta_1, \sigma^2) \cdot f(\beta_0) \cdot f(\beta_1) \cdot f(\sigma^2)$$

$$\propto (8\pi^3 \sigma^2 \sigma_{\beta_0}^2 \sigma_{\beta_1}^2)^{-1/2} \exp\left[-\frac{(TQI_i - \beta_0 - \beta_1 t_i)^2}{2\sigma^2}\right] \cdot \exp\left[-\frac{(\beta_0 - \bar{\beta}_0)}{2\sigma_{\beta_0}^2} - \frac{(\beta_1 - \bar{\beta}_1)}{2\sigma_{\beta_1}^2}\right] \cdot \frac{b^{-a}}{\Gamma} (\sigma^2)^{-a-1} \exp\left(\frac{-1}{b \cdot \sigma^2}\right)$$

최종적으로 도출된 결합확률분포는 위 수식에서 보듯이 정규분포, 이변량정규분포, 역감마분포의 곱에 비례하는 다변량 확률분포가 되며, β_0, β_1의 한계확률을 계산하려면 위 식을 다시 적분해야 하는데 이는 현실적으로 불가능하다. 앞에서 서술했듯이, 이러한 복잡한 형태의 사후분포로부터 확률변수를 효과적으로 표본 추출하여 그 확률변수의 평균이나 분산과 같은 통계량을 표본값들로부터 유추하는 대표적인 방법론으로 MCMC의 일종인 깁스 표본 기법(Gibbs Sampling)[11]을 이용할 수 있다. 깁스 표본기법은 사후분포에 존재하는 모든 파라미터들의 결합사후 확률분포로부터 직접 표본을 추출하지 않고, 완전조건부사후분포, 즉 하나의 파라미터를 제외한 나머지 파라미터들은 조건부로 설정한 사후분포로부터 표본을 추출하고, 그 표본을 기초로 또 다른 파라미터의 표본을 추출하는 단계적인 방식의 표본 추출 기법이다.

2.3 모형의 적용

본 연구에서는 국내 고속철도 자갈궤도 중에서 약 48km 구간에 대한 2년 4개월간의 면틀림량을 분석하였다. 진전율을 효과적으로 파악하기 위해 인력 탬핑 작업을 제외한 어떠한 유지보수 작업도 없는 구간을 선택하였으며, 200미터 길이의 총 31개의 자갈궤도 토공부 세그먼트를 선정하였다.

2.3.1 사전분포의 유도

TQI 진전율에 대한 사후분포를 계산하기 위해서는 β_0, β_1의 사전분포를 선정해야 한다. 본 연구에서는 이 β_0, β_1이 무정보적 사전분포를 갖는 경우와 β_0, β_1가 서로 독립으로 개별적인 정규분포를 갖는 경우, 그리고 앞에서 서술한 이변량 정규분포를 갖는 경우에 대해, 기존의 전통적인 회귀분석 방법론과의 비교 연구를 수행한다. 이를 위해 먼저, 31개의 토공부 세그먼트에 대해 전체 기간동안 TQI 진전율을 전통적인 회귀분석 방법에 의해 계산해 보았다. 이렇게 도출된 진전율(β_0) 및 y절편값(β_1)을 일종의 사전확률로 간주하고자 한다.

Table 1. Inducing prior probability.

Normality test									
	Mean	var.	Covariance	Kolmogorov-smirnova			Shapiro-wilk		
				Statistics	d.f	Sig. prob.	Statistics	d.f	Sig. prob.
β_0	1.666	0.42389	2.42E-05	0.122	31	.200	0.961	31	0.310
β_1	4.83E-04	1.37E-07		0.088	31	.200	0.969	31	0.501

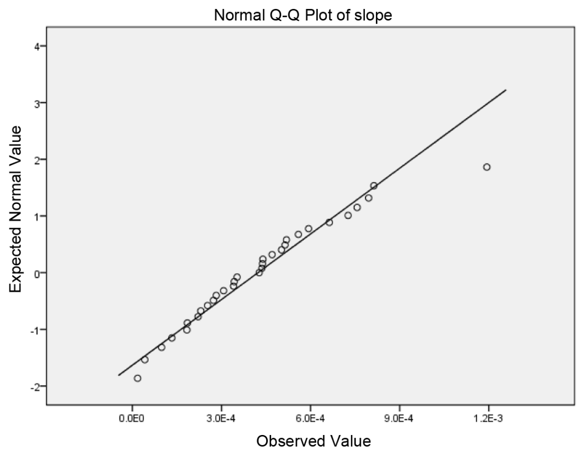

Fig. 4. Q-Q Plot for β_1

각각 31개의 진전율값과 y절편값의 평균, 분산, 공분산을 계산하면 Table1과 같으며, 특히 β_0, β_1이 정규분포를 이룬다고 가정할 수 있는지를 검토하기 위해 통계 패키지 SPSS 18을 이용한 Kolmogorov-Smirnova 분석, Shapiro-Wilk 분석을 수행하였다. 유의확률이 둘 다 0.2이상으로 유의수준을 20% 미만으로 설정할 경우, 정규분포가 아니라고 할 만한 충분한 통계적 증가가 존재하지 않아, 두 파라미터의 분포는 정규분포로 가정할 수 있다. 진전율의 정규성을 확인하기 위해, Q-Q plot을 그려보면 Fig. 4와 같다. 위 Table 1의 데이터를 기초로 $\beta=(\beta_0,\beta_1) \sim N_2(\overline{\beta},\Sigma)$를 구성할 수 있고, $1/\sigma^2$은 감마분포 $gamma(0.001,\ 0.001)$를 따른다고 가정하여 사전 분포를 구성한다.

2.3.2 사전분포에 따른 TQI 진전율의 변화

사전분포 설정에 따른 TQI의 진전율을 확인해보기 위해, 2.3.1절에서 제시한 사전분포 뿐만 아니라, 계산적인 의미에서의 최소한의 역할만을 제공하고 관측값을 최대한 고려하는 무정보적 사전분포, 두 파라미터 β_0, β_1의 공분산(covariance)가 0에 가깝다는 의미에서 두 파라미터의 사전분포로 서로 독립인 개별적인 정규분포까지 고려한 총 총 3가지 사전분포를 고려하였으며, 이러한 사전분포에 따른 진전율의 변화를 살펴보았다. 무정보적 사전분포는 일양분포(uniform distribution)을 활용할 수도 있지만, 본 연구에서는 분산이 매우 큰 정규분포를 이용하여 표본의 무작위성을 고려하였다. 본 연구에서 고려하는 세 가지 사전분포를 정리하면 아래와 같다.

- β_0, β_1의 사전분포1(무정보적 사전분포) : $\beta_0 \sim N(\overline{\beta}_0,\ 10^6)$, $\beta_1 \sim N(\overline{\beta}_1,\ 10^6)$
- β_0, β_1의 사전분포2(각각의 독립된 정규분포) : $\beta_0 \sim N(\overline{\beta}_0,\ \sigma^2_{\beta_0})$, $\beta_1 \sim N(\overline{\beta}_1,\ \sigma^2_{\beta_1})$
- β_0, β_1의 사전분포3(이변량 정규분포) : $\beta=(\beta_0,\ \beta_1) \sim N_2(\overline{\beta},\ \Sigma)$

Table 2. Gibbs sampling results in case of bivariate normal prior distribution in case that $n = 10$ in Fig. 3.

Parameter	Mean	St. dev.	MC error	2.5%	Median	97.5%	Sample
β_0	2.474	0.08426	0.00132	2.311	2.474	2.642	5000
β_1	8.59E-04	3.62E-04	5.81E-06	1.30E-04	8.68E-04	0.001559	5000

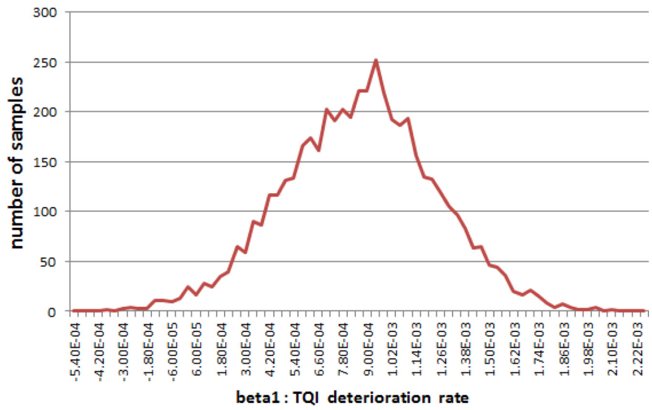

Fig. 5. Samples from the posterior distribution of using Gibbs sampler.

앞에서 서술했듯이 사후분포의 몬테카를로 시뮬레이션에 활용된 방법은 Winbugs[12]에 내장된 깁스 표본 기법을 활용하였다. Table 2와 Fig. 5는 이변량 정규 사전분포를 고려했을 때, 깁스 표본 추출의 결과를 나타낸 것이다. Table 2에 표현된 통계량들은 마코프 체인이 안정상태로 가기 위한 과정으로 5000번의 번인(burn-in)을 거친 후, 사후 분포로부터 5000개의 표본을 추출했을 때의 통계량을 나타낸 것이다. 특히, TQI 진전율에 해당하는 β_1을 살펴보면, β_1의 평균($E(\beta_1)$)은 0.000859(mm/day)이며, 그것의 95% 신용구간(credible interval)은 [0.00013, 0.001559]이다. 95% '신용구간'은 정확하게 진전율이 그 구간에 들어가 있을 확률이 95%라는 의미로서, 진전율이 속할 확률이라는 개념과는 전혀 관계가 없는 전통적인 통계학에서의 '신뢰구간(confidence interval)'과는 완전히 다른 개념이다. 즉, β_0, β_1이 앞에서 서술한 이변량 정규 사전분포를 따를 경우, β_1이 이 신용구간에 포함될 확률은 정확히 95%가 되는 것이다. Fig. 5는 이러한 5000개의 표본을 확률밀도함수(probability density function) 형태로 을 이용하여 밀도함수를 표현한 것인데, β_1의 사후분포는, 일반적인 정규분포와 유사하지만 왼쪽 꼬리가 조금 더 길며, 중간 부분이 좀 더 두꺼운 형태의 확률분포임을 알 수 있다.

Fig. 6은 Fig. 3에서 활용한 동일한 TQI값들에 대해, 전통적인 회귀분석에 의한 진전율과, 위 세 가지 사전분포를 이용하여 진전율(β_1)의 평균값들을 비교한 것이다. 먼저 막대그래프는 전통적인 회귀분석에 의한 진전율을 나타낸 것이고, 긴 점선으로 표현된 진전율은 무정보적 사전분포를 이용했을 때의 진전율을 나타낸 것이다. 이 두 가지 진전율은 왼쪽 수직 좌표값에 나타나 있으며, 그림에서 보듯이 TQI데이터가 추가됨에 따라 진전율의 변동이 굉장히 심함을 알 수 있다. 반면 진전율에 대한 기존 데이터를 적극 활용하는 베이지안 모형의 경우, 진전율이 상당히 안정적인 형태로 변화하는데, 실선으로 표현된 이변량 정규분

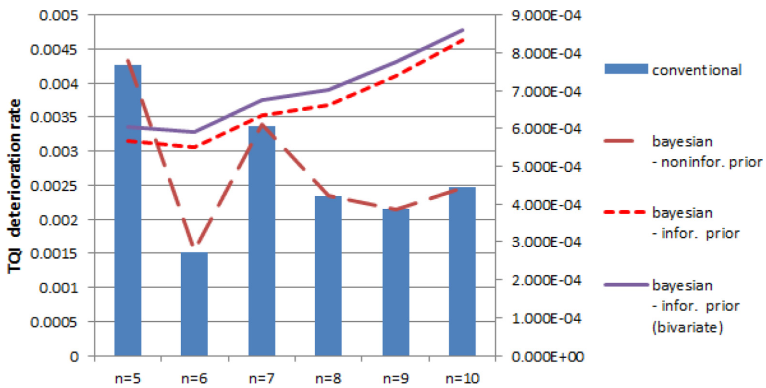

Fig. 6. Comparison of the deterioration rates: conventional vs. Bayesian.

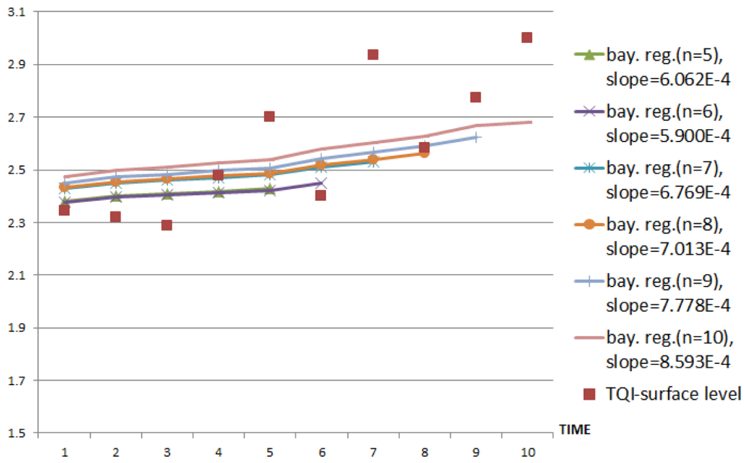

Fig. 7. Estimating the deterioration rate based on Bayesian linear regression with bivariate normal prior distribution.

포를 사전분포로 설정한 경우가 개별적인 독립된 정규분포를 가정한 경우(작은 점선)보다 진전율이 높게 나왔다. 두 경우 오른쪽 수직 좌표값이 진전율을 의미하며, 수직축의 단위가 0.0001임을 생각해서 보면 TQI데이터가 추가됨에 따라 진전율의 증가가 매우 적음을 알 수 있다. Fig. 7은 이변량 정규분포를 가정했을 때의 진전율을 TQI값과 함께 도시한 것으로, 독립적인 정규분포를 사전분포로 가정한 경우도 비슷한 형태의 그래프를 얻을 수 있었다.

3. 결론 및 추후 연구 과제

본 연구에서는 베이지안 방법론에 기초하여 궤도틀림 진전율을 예측하는 방법론을 제시하였다. 기존의 전통적 통계학(frequentist statistics)에서 다루는 선형 회귀분석과 달리 베이지안 회귀분석은 과거 데이터 혹은 전문가의 식견을 통한 진전율에 대한 사전 지식을 포함시킬 수 있으며, 진전율 자체가 확률변수로서 다루어진다는 점에서 기존 연구들과 큰 차이점이 있다. 특히 본 연구에서 핵심적으로 살펴보았듯이, 전통적인 회귀분석 방법으로는 매 시기 추가되는 TQI 값에 따라 진전율의 변동폭이 매우 커진다는 단점이 있는 반면, 본 연구에서 제시한 베이지안 회귀분석방법은 과거 데이터로부터 도출된 진전율을 사전분포 형태로 고려함과 동시에, 그 구간에 나타난 관측값을 결합하여 사후분포로 예측함으로써, 보다 안정적인 진전율을 예측할 수 있음을 보였다.

그러나 본 연구에서 제시한 방법론을 일반화하기 위해서는 세그먼트 특성에 따라 달라질 수 있는 사전분포에 대한 보다 많은 연구가 필요하며, 이를 위해서는 무엇보다도 TQI를 계산하는 궤도의 기본 단위라 할 수 있는 세그먼트를 나누는 기준에 대한 연구도 필요하다. 이 세그먼트를 어느 정도의 길이로 할 것인지, 그리고 궤도의 어떤 특성에 의해 나눌지에 대한 기존 연구는 전무하다. 예를 들어 1km 구간을 5개의 200미터 세그먼트로 구분하여 진전율을 계산하는 경우, 1km를 단일 세그먼트로 하여 진전율을 계산하는 경우, 두 진전율 값은 매우 상이함을 알 수 있었다.

마지막으로 본 연구에서 다룬 모델은 두 번의 연속적인 궤도 유지보수 작업 사이에서 시간에 따라 진행되는 TQI 진전율이지만, 실제로 보다 더 중요한 것은 유지보수 작업이 TQI 진전율에 미치는 영향을 계산할 수 있는 모델을 개발하는 것이다. 이와 관련하여 몇 가지 비선형 모형을 참조할 수 있으며[13], 어떤 모형이 한국 궤도에 적합한지에 대한 실증적 연구가 필요하다. 물론 이 모형에 대해서도 본 연구에서 제시한 베이지안 기법을 결합함으로써, 본 연구에서 서술한 전통적 선형회귀모형에서 발생하는 문제를 해결할 수 있을 것으로 기대된다.

References

[1] J. Lee, Y.B. Choi (2009) Application of track recording data for track maintenance, *Proceedings of the Korean Railway Association Spring Conference*, Gyeongju, pp. 3057-3063.

[2] A.R. Andrade and P. F. Teixeira (2012) A Bayesian model to assess rail track geometry degradation through its life-cycle, *Research in Transportation Economics*, 36(1), pp. 1-8.

[3] N. Kim, S. Lee, Y. Won, *et al*. (2009) Introduction of track quality index (TQI) methods using track induction data, *Proceedings of the Korean Railway Association Fall Conference*, Jeju, pp. 66-72.
[4] M. El-Sibaie and Y. Zhang (2004) Objective track quality indices, *Transportation Research Record : Journal of the Transportation Research Boar*d, 1863, pp. 81-87.
[5] Y. Zhang, M. El-Sibaie, Sung Lee (2004) FRA track quality indices and distribution characteristics, *Proceedings of The American Railway Engineering and Maintenance-of-way Association 2004 Annual Conference,* Nashville, pp. 1-26.
[6] M.C. Jeong, J.H. Kim, J. Lee, *et al*. (2012) Study for progress rate of standard deviation of irregularity based on track properties for the railway track maintenance Cycle Analysis, *Journal of the Korea Institute for Structural Maintenance and Inspection*, 15(3), pp. 31-40.
[7] G.C. Shin (2013) A Study on the progress of track irregularity by track structure in urban railway system, *Mater thesis, Seoul National University of Science and Technology.*
[8] H. Park, S.Y. Jang, S. Park (2014) Correlation analysis between track irregularity and maintenance of high-speed railway, *Proceedings of the Korean Railway Association Fall Conference*, Jeju, pp. 1130-1133.
[9] J.-H. Ko, M.-C. Kim, J.-H. Lee, J.-G. Cho, and Y.G. Park (2011) An analysis of the Track Irregularity Progress on the Various Track System in Urban Transit, *Proceedings of the Korean Railway Association Fall Conference*, Jeju, pp. 311-319.
[10] D.-Y. Kim *et al*. (2008) Track Deterioration Prediction and Scheduling for Preventive Maintenance of Railroad, *Proceedings of the Korean Railway Association Fall Conference*, Gwangju, pp. 1346-1357.
[11] M.S. Oh (2013) Bayesian Statistical Inference with R Monte Carlo, *Freedom Academy.*
[12] http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/ (Accessed 15 May 2016).
[13] J. Zhao *et al*. (2006) Optimizing policies of railway ballast tamping and renewal, *Transportation Research Record Journal of the Transportation Research Board,* 1943, pp. 50-56.

(Received 13 July 2016; Revised 5 August 2016; Accepted 5 August 2016)

Bum Hwan Park : beomi72@hanmail.net
Department of Railroad Management and Logistics, Korea National University of Transportation, 157 Cheoldo bangmulgwan-ro, Uiwang-si, Gyeonggi-do, 16106, Korea