

상시조사 교통량 자료의 결측 보정에 관한 연구

이상협* · 신재명**

Lee, Sang Hyup* · Shin, Jae Myong**

A Study on Imputing the Missing Values of Continuous Traffic Counts

ABSTRACT

Traffic volumes are the important basic data which are directly used for transportation network planning, highway design, highway management and so forth. They are collected by two types of collection methods, one of which is the continuous traffic counts and the other is the short duration traffic counts. The continuous traffic counts are conducted for 365 days a year using the permanent traffic counter and the short duration traffic counts are conducted for specific day(s). In case of the continuous traffic counts the missing of data occurs due to breakdown or malfunction of the counter from time to time. Thus, the diverse imputation methods have been developed and applied so far. In this study the applied exponential smoothing method, in which the data from the days before and after the missing day are used, is proposed and compared with other imputation methods. The comparison shows that the applied exponential smoothing method enhances the accuracy of imputation when the coefficient of traffic volume variation is low. In addition, it is verified that the variation of traffic volume at the site is an important factor for the accuracy of imputation. Therefore, it is necessary to apply different imputation methods depending upon site and time to raise the reliability of imputation for missing traffic values.

Key words : Traffic volume, Missing value, Imputation, Applied exponential smoothing method, Coefficient of variation

초 록

교통량은 교통망 계획, 도로 설계, 도로 관리 등에 직접적으로 활용되는 중요한 기초자료이다. 교통량은 고정식 교통량조사 장비를 설치하여 연속적인 자료를 수집하는 상시조사와 특정일을 조사하는 수시조사로 구분되어 조사되고 있다. 상시조사의 경우 조사 지점에 설치되어 있는 장비의 고장이나 오작동 등으로 인하여 교통량 자료의 결측이 발생하며, 이러한 결측을 보정하기 위하여 다양한 방법이 적용되어 왔다. 본 연구에서는 결측 발생일 전·후의 자료를 활용하는 응용 지수평활화법을 제안하였으며, 평가 결과 교통량 변동계수가 낮은 경우 보정의 정확성이 제고됨을 알 수 있었다. 게다가 지점의 교통량 변동성이 결측 보정의 정확성에 영향을 미치는 중요한 요인으로 작용한다는 것을 확인하였다. 따라서 교통량 결측 보정의 신뢰성을 높이기 위해서는 지점별-시기별 결측 보정 방법이 달리 적용되어야 할 것이다.

검색어 : 교통량, 결측, 보정, 응용 지수평활화법, 변동계수

1. 연구 배경 및 목적

교통량이란 도로의 특정 지점 또는 구간을 단위 시간 동안 통과하는 각종 차량의 수를 의미한다. 따라서 교통량조사(Traffic Counts)는 도로를 이용하는 각종 차량의 단위 시간 당 통과 대수를 종류별, 방향별 및 시간대별로 관측하는 조사이다. 이렇게 조사된 교통량(Traffic

* 정회원 · 교신저자 · 한국건설기술연구원 연구위원

(Corresponding Author · Korea Institute of Construction Technology Research Fellow · infohi2@hanmail.net)

** 명지대학교 교통공학 박사과정 (firstrain1@naver.com)

Received December 14, 2012/ revised January 3, 2013/ accepted April 26, 2013

Volume) 자료는 도로 설계, 포장 설계, 도로 운영 및 유지관리 등에 폭넓게 이용되며 도로망 계획, 교통망 계획 및 교통관리 계획 수립을 위한 기초적 정보가 된다.

이러한 교통량 자료를 수집하기 위하여 각 조사 지점에 고정식 교통량조사 장비(Permanent Traffic Counter)를 설치하여 1년 이상의 장기간에 걸쳐 그 지점을 통과한 차량의 차종별 대수 및 총 대수를 조사하는데 - 이러한 유형의 교통량조사는 상시조사(Continuous Traffic Counts)로 치칭된다. - 이 조사 장비는 제어기와 센서 등 기계적인 시스템으로 구성되어 있으며 차량이 다니는 도로변과 도로표면에 설치되어 있어서 온도, 습도, 먼지, 진동, 비락, 압력 등에 항상 노출되어 있다. 따라서 노후화가 빨리 진행되며 고장이 자주 발생하게 된다(Lee et al., 2012).

이처럼 제어기나 센서가 고장 나거나 파손될 경우 교통량 자료를 수집할 수 없는 상태, 즉 결측이 발생하게 된다. 결측은 기간에 따라 장기와 단기로 구분할 수 있는데 제어기 고장의 경우는 부품 교체나 제어기 교체 등으로 단기간 내에 정상 가동이 가능하여 결측 기간이 짧을 수 있으나 센서 불량, 감도 저하나 단선 등 센서 고장의 경우는 도로포장 면 절개 등을 수반하는 센서 교체 공사가 이루어져야 하므로 단기간 내에 정상 가동이 불가능하여 결측 기간이 길게 된다.

따라서 상시조사의 경우 현장 조사장비에서 수집되는 원시자료(Base Data)는 대부분의 경우에 결측이 발생하게 되며 이 원시자료는 자료의 진실성(Truth in Data)을 위하여 결측에 대한 보정 없이 수집된 상태로 유지되어야 한다. 그러나 연간 교통량 통계 작성 등의 분석을 위한 단계에서는 필수적으로 원시자료에 대한 결측 보정이 이루어져야 한다(AASHTO, 1992; Zhong et al., 2006).

현재 상시조사 교통량의 결측 보정은 결측 기간별로 단기 결측 보정과 장기 결측 보정으로 구분하고 있으며, 단기 결측 보정의 경우는 선형 보간법, 차종별 자료를 시간대별로 배분하는 방법, 상하행 비율을 적용하는 방법 등이 적용되고 있으며, 장기 결측 보정에는 Factor법이 적용되고 있다. 이런 결측 교통량의 보정은 이해하기 쉽고 적용이 용이하다는 측면에서 장점이 있으나 결측 발생 시점 기준으로 정상적으로 수집된 과거자료에 의존한 보정을 수행하므로, 단기적으로 변화하는 추세변동 또는 순환변동을 명확히 설명하기에는 어려움이 있다.

따라서, 본 연구에서는 이해하기 쉽고, 시스템 구현이 용이하며, 통계적 신뢰성이 확보된 시계열적 요소를 고려한 지수평활화법을 응용하여 결측 발생 시점 전과 후에 정상적으로 수집된 자료를 이용한 보정방법을 제시한다. 그리고 기존 방법과 제시된 보정방법의 정확성 비교평가는 2010년 국도 상시조사 지점 중 8개 지점에 대하여 단기(1주일) 교통량이 결측된 것으로 가정하

후 이 기간의 교통량에 대하여 결측 보정을 실시한 후 실제 자료와의 비교를 통하여 보정의 정확성을 평가한다. 정확성 평가기준은 평균절대백분율오차(MAPE: Mean Absolute Percent Error)를 사용한다.

2. 기존 연구 고찰

2.1 국내외 연구사례

조사장비를 이용하여 수집된 교통량 조사 자료에 대한 결측 보정은 그 자료의 용도에 따라 ITS용 실시간 교통량 보정과 통계자료 작성용 보정으로 구분될 수 있다. 국내의 경우에는 자료에 대한 제한적인 접근으로 인하여 극히 한정적인 연구만 수행되었다(Korea Institute of Construction Technology, 2001).

실시간 교통량 보정 연구사례를 살펴보면, Lee et al.(2003)는 일반국도 3호선에 설치된 영상검지기 자료를 분석하였으며 과거에 정상 수집된 이력자료(일반적으로 5주간 평균값)를 이용한 프로파일 자료를 활용한 보정과 유사 검지기의 시점별 자료를 이용한 회귀보정을 수행하였다.

통계자료 작성용 보정 연구사례를 살펴보면, Ha et al.(2007)는 상시조사 지점에 대하여 계절시계열 모형, 자기회귀모형을 적용하였으며 보정 오차는 지점 평균으로 ARIMA 모형이 26.4%, 자기회귀모형이 30.5%를 보이고 있다. ARIMA 모형을 적용하여 보정한 결과를 살펴보면 시간대별 추정방법은 요일별 변동특성을 반영하지 못한 결과를 보이고 있으며, 자기회귀모형의 경우 ARIMA 모형보다 다소 높은 신뢰성을 나타내고 있다. 그러나 이를 적용하기 위해서는 인근에 교통량 변동이 유사한 조사지점이 항상 존재하여야 하므로 현재의 상시조사와 같은 조사체계에서는 적용하기에 다소 무리가 있을 것이다.

Sharma(2003)와 Zong et al.(2004)과 Zong and Sharma(2009)는 ARIMA 모형과 유전자회귀 모형, 신경망 모형을 이용하여 캐나다 Alberta주의 6개 상시조사 지점에서 수집된 교통량 자료에 대한 결측 보정을 수행한 결과 유전자 회귀모형이 가장 우수한 것으로 나타났다고 하였다. 또한 모형의 정확성은 도로의 종류와 기능에 따라 다르게 나타난다는 것을 알게 되었다고 하였다.

이외에도 Hu et al.(1998)은 시뮬레이션을 통하여 단기 수시조사 자료를 이용하여 교통량을 추정하는 연구를 하였으며, Lingras et al(2003)은 교통관리체계를 위한 자료 보정전략을 연구하였다. Weijermars(2007)는 clustering를 통하여 도시교통 패턴의 분석을 시도하였으며 Wright and Hu(1997)는 교통량의 변동성을 규명하려 시도하였다.

2.2 각국의 교통량 결측 보정 방법

상시조사 지점과 같은 교통량조사 체계를 구축하여 장기간 교통량 자료를 수집·관리하고 있는 각국의 교통량 결측 보정방법을 살펴보면 Table 1과 같다. Table 1에 명기된 대부분의 나라들에서는 교통량 결측이 발생할 경우 동일 지점의 과거 수집 자료를 활용하여 보정 값을 산출한다. 과거 수집 자료의 종류로는 과거 연도의 동 월 동 일 동 시간 자료, 당해 연도의 동 월 전·후 주, 동 요일 동 시간 자료가 있으며, 이 자료를 평균으로 적용하거나 변동계수의 조정을 통해서 결측 값을 산출한다. 전자의 경우는 선형 보간법, 후자의 경우는 Factor법이 대표적이며, 국내에서는 선형 보간법이나 Factor법이 주로 적용되고 있다.

특이하게 네덜란드의 경우는 정상인 인접 지점의 교통량을 이용하는데 이는 도시부에 단거리로 구축된 검지기의 공간적 특성이 반영된 결과라 할 수 있다. 또 다른 특징은 결측된 교통량을 보정하기 위하여 복잡하거나 어려운 알고리즘을 적용하지 않는다는 것이다. 이는 수집되는 자료의 방대한 양, 시스템 구현의 용이성, 교통량의 시간적 변동 특성이 반영된 결과일 것이다. 즉 특정지점 교통량의 시간적 변동은 외부적인 여건의 변화가 발생하지 않으면 단기간에 급격한 변동은 발생하지 않으며, 기간에 따라 서서히 변화하는 특성이 반영된 결과일 것이다.

3. 교통량 결측 보정방법의 개발

3.1 교통량의 시간적 변동

일반적으로 상시조사 지점의 교통량은 장기적으로 계절별, 월별로 변화하며, 중기적으로는 요일별로 변화하며, 단기적으로는 일중 시간대 및 방향별로 변화한다. Figs. 1-5는 일반국도 38호선 경기도 안성시 보개면 양북리 소재 상시조사 지점(지점번호 10002)의 교통량의 시간적 변동을 나타낸 그림들이다.

교통량의 계절별 변동은 일반적으로 겨울철보다 여름철에서 높은 값을 나타내고 있으며, 요일별 교통량은 지점별 특성에 따라 다르게 나타날 수 있으나 대부분 주중 교통량은 비슷한 특징이 있다. 월별-요일별 교통량은 요일별로 유사한 특성을 보이고 있으나 추세변동은 월변동이 반영된다. 요일별-시간대별 교통량 변동은 주중은 유사하게 오전과 오후 첨두를 보이고 있으나 주말의 경우는 오전 첨두는 나타나지 않으며 오후 첨두도 뚜렷하게 나타나지는 않는다. 일중 방향별 시간대별 변동은 방향별로 첨두시간이 명확히 구분되며 시간대별 방향별 교통량은 서로 다르다.

또한, 장기적 관점에서는 외부적인 환경변화가 없다면 연도별로 비슷한 교통량이 반복되어 나타나는 순환변동의 특성을 보이고, 중기적으로는 주별로 비슷한 변동특성이 반복적으로 나타나는 순

Table 1. Imputation Models for Missing Values Used in the Practice(Zhong and Sharma, 2009)

Agency	Model Inputs	Prediction Function and Output
Alabama	Hourly volume from the previous year or other part of the month	Directly use input as output
Alberta	Do not impute missing hourly volumes, but use historical data to estimate monthly average daily traffic(MADT)	ATRs with missing data will be assigned to ATRs without missing data, and a ratio is used to convert MADTs from ATRs without missing data to those for ATRs with missing data
Central Transport Group, England	Valid data of last week collected from the same site	Historical data are multiplied by a factor taken from nearby sites that worked correctly
Delaware	Hourly volumes from the same hour of the month before and the month after	Take average of the two hourly volumes
France	Hourly volume from the same hour in the previous month	Directly use hourly volume from the previous month
Indiana	Data from the previous year	Directly use input as output
London, England,	Historical hourly volumes from the same hours of the same days in the previous weeks	Historical values are averaged with the weights calculated based on their time lags with missing values
Montana	Historical data from the same location	Expanded by factors if there are changes in the area
Netherlands	Data from other location	“Smart” linear interpolation
Oklahoma	Data from the same day of the weeks of the month	Directly use input as output
Saskatchewan (Manitoba)	Hourly volume from the same hour on the same day of the week in the last year	Directly use input as output
South Dakota	Hourly volumes from the same hours on the same day of the week in the previous years	Take average of hourly volumes from previous years as output
Vermont	Data from the same day of the same month in the previous year	Directly use input as output

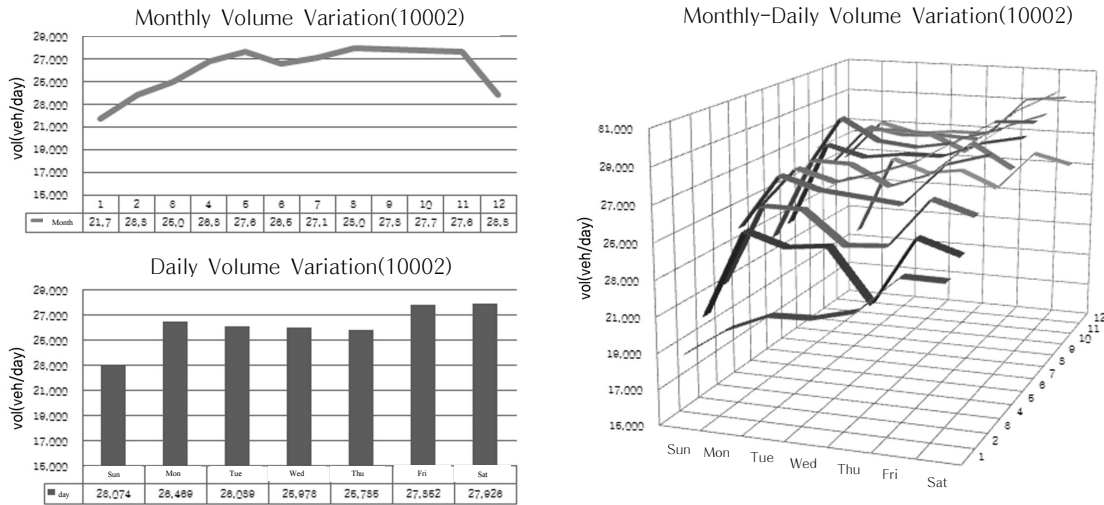


Fig. 1. Example of the Temporal Volume Variation Patterns(Site 10002)

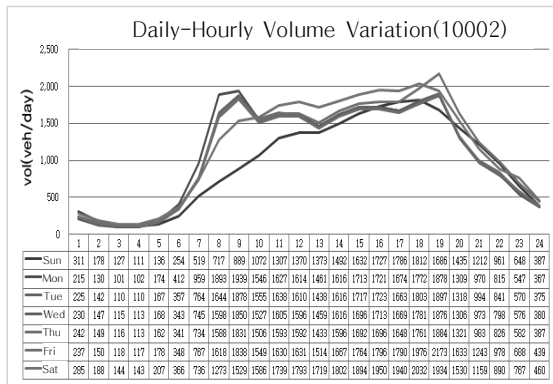


Fig. 2. Daily-Hourly Volume Variation Patterns

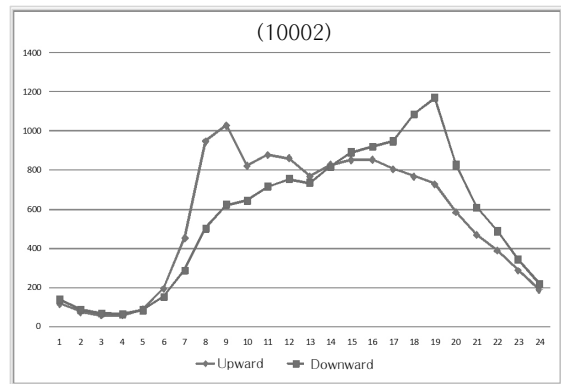


Fig. 3. Directional-Hourly Volume Variation of a Day

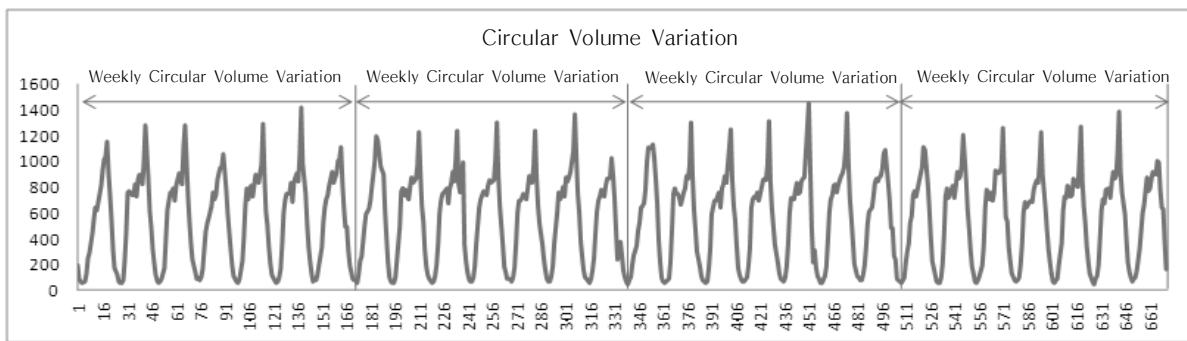


Fig. 4. Example of the Weekly Circular Volume Variation Patterns

환변동의 특성을 보인다고 할 수 있다. 또한 단기적인 일 교통량은 주중에서 비슷하게 반복적으로 나타나고, 시간교통량도 주중 요일 별로 비슷한 특성을 보이고 있다.

이상으로 살펴본 특성을 요약하면 특정일의 시간 교통량은 동일

한 월의 동일한 요일의 시간 교통량이 가장 유사하다. 이러한 점에 비추어 볼 때 특정일의 교통량 결측 보정에는 결측일의 전후주의 동일한 요일의 교통량을 기준으로 모형을 개발하는 것이 타당할 것이다.

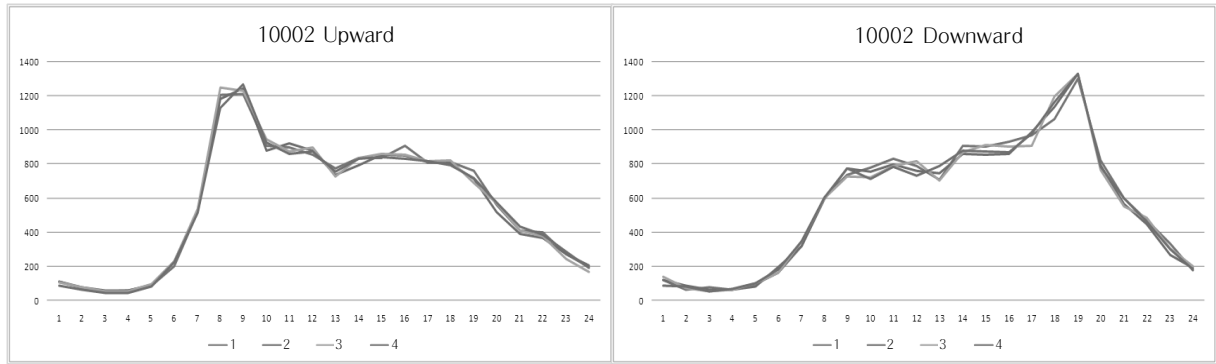


Fig. 5. Hourly Volume Variations of the Same Days of a Month

3.2 지수 평활화법을 이용한 교통량 결측 보정방법의 설계

3.2.1 입력자료의 설계

일교통량 자료는 Fig.6에서 볼 수 있듯이 시간의 흐름에 따른 주별(w), 일별(t)로 시계열적 특성을 보이고 있으며, X축에 따라서는 요일별 시간별로 변화하며, Y축인 주별(w)에서는 이들 요일의 특성이 반복적으로 나타난다.

따라서, 신규로 개발되는 결측 보정의 참조자료는 Table 2에서 볼 수 있듯이 요일별 방향별 시간교통량을 기준으로 선정하며, 월별 추세변동을 반영하기 위해서 최소 2개월의 연속자료를 입력자료로 선정한다.

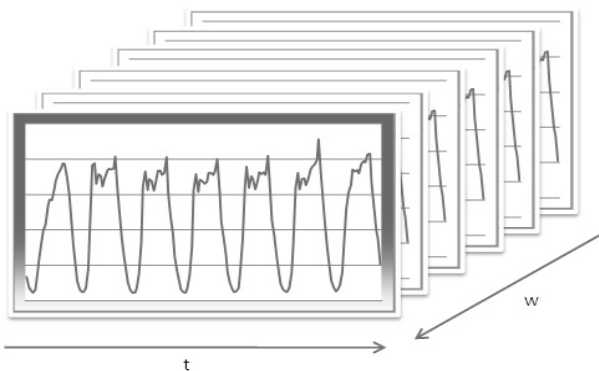


Fig. 6. Temporal Characteristics of Traffic Counts Data

Table 2. Model Inputs

Inputs	Outputs
<ul style="list-style-type: none"> Data which can reflect circular variation, trend variation and so forth - Apply the monthly trend and the weekly-hourly circular variation - Use the data of two months or eight weeks at least in case of reflecting the monthly trend variation 	Directional-hourly traffic volume

3.2.2 단순 지수평활화법

지수평활화법은 과거의 관측치를 이용하여 미래의 값을 예측할 때 최근의 자료에 더 많은 가중치를 부여하여 예측(보정)하는 기법으로 시간이 흐름에 따라 지수적으로 감소하는 가중치를 적용한다.

평활화 계수 α 는 0과 1사이의 값이며, 1에 가까우면 최근 시점의 관측치에 더 큰 가중치가 부여되며, 0에 가까우면 과거시점의 관측치에 더 큰 가중치가 부여된다.

자료의 변동이 심한 경우 α 는 작은 값을 사용하고, 자료가 특정한 패턴이 있고 임의성이 적다면 큰 값의 α 를 선정한다.

단순 지수평활화법을 이용한 결측 보정값은 Eq. (1)과 같다.

$$\widehat{x_{w,s(t)}} = \alpha x_{w-1,s(t)} + (1-\alpha)\alpha x_{w-2,s(t)} + (1-\alpha)^2 x_{w-3,s(t)} + \dots + (1-\alpha)^{n-1} x_{w-n,s(t)} \quad (1)$$

여기서, $\widehat{x_{w,s(t)}}$ = w주 s요일의 t시간대 결측보정값(단방향)

$x_{w,s(t)}$ = 과거 w주 s요일의 t시간대 단방향 교통량

α = 평활화 계수 ($0 < \alpha < 1$), 0.5 적용

3.2.3 응용 지수평활화법

본 연구에서 대상이 되는 상시조사 교통량 수집 자료는 결측 기간이 발생한 전후에 정상적으로 수집된 자료를 활용할 수 있는 장점이 있다. 즉, 단순히 과거 자료에 의존한 기존 방식과 달리 이후 자료도 함께 적용할 수 있으므로 시간에 따른 추세변동을 설명할 수 있다.

따라서, 결측일을 기준으로 이전 자료와 이후 자료의 동일한 요일의 동시간대의 교통량을 시간의 흐름에 따라 가중치를 부여하는 방식으로 순방향과 역방향의 지수평활화 값의 산술평균을 보정값으로 적용하며 Eq. (2)와 같다.

$$\widehat{x_{w,s}(t)} = \frac{1}{2}[\alpha x_{w-1,s}(t) + (1-\alpha)\alpha x_{w-2,s}(t) + (1-\alpha)\alpha^2 x_{w-3,s}(t) + \dots + (1-\alpha)\alpha^{n-1} x_{w-n,s}(t) + \alpha x_{w+1,s}(t) + (1-\alpha)\alpha x_{w+2,s}(t) + (1-\alpha)\alpha^2 x_{w+3,s}(t) + \dots + (1-\alpha)\alpha^{n-1} x_{w+n,s}(t)] \quad (2)$$

여기서, $\widehat{x_{w,s}}$ = w주 s요일의 t시간 결측보정값(단방향)
 $x_{w+n,s}$, $x_{w-1,s}$ = 이전 주, 이후 주 s요일의 t시간대 단방향 교통량
 α = 평활화 상수 ($0 < \alpha < 1$), 0.5 적용

여기서, GF_i = 과거 i연도의 성장계수
 교통량 = 과거 i연도의 동월, 동요일, 동시간대의 교통량
 N = 과거연도 수

지수평활화법의 경우 결측 발생 전 자료의 지수평활화 값을 적용하며 응용 지수평활화법의 경우 결측 발생 전후 자료의 지수평활화 값의 평균을 적용한다. 지금까지 언급한 결측 보정 방법들이 Table 3에 요약되어 있다.

정확성 평가기준은 식 (4)의 평균절대백분율오차(MAPE : Mean Absolute Percent Error)를 사용하였으며 관측 값은 결측 기간 동안 실제 수집된 자료를 활용하였다.

$$MAPE = \frac{1}{N} \sum \frac{(\widehat{Y}_i - Y_i)}{\widehat{Y}_i} \times 100 \quad (4)$$

여기서, \widehat{Y}_i = 보정값, Y_i = 관측값, N = 보정일

평가 대상 지점은 Table 4에서 볼 수 있듯이 2010년 상시조사 지점 중 일반국도 38호선 경기도 안성시 보개면 양북리, 일반국도 31호선 강원도 태백시 동점동, 일반국도 19호선 충청북도 영동군 학산면 봉소리, 일반국도 32호선 충청남도 서산시 팔봉면 어산리, 일반국도 26호선 전라북도 군산시 개정면 발산리, 일반국도 13호선 전라남도 강진군 성전면 영풍리, 일반국도 20호선 경상북도 청도군

4. 교통량 결측 보정 방법 평가

4.1 알고리즘 평가 계획

결측 보정 알고리즘의 평가는 전통적인 결측 보정 방법인 선형보간법, Factor법, 지수평활화법과 응용 지수평활화법을 대상으로 하며, 동일한 기간의 교통량 결측 보정을 수행하여 모형별 보정 오차를 비교한다. 선형보간법의 경우 결측 발생일 전후 정상 수집된 교통량의 평균값을 적용하며 Factor법의 경우 이전 연도 자료의 Factor를 적용하며 Eq. (3)에 의하여 구한다.

$$\text{보정 값} = \sum_{i=1}^N (\text{교통량}_i \times GF_i) / N \quad (3)$$

Table 3. Characteristics of Missing Value Imputation Methods

Missing Value Imputation Methods	Characteristics
Exponential smoothing method	Exponentially smoothed values of the data collected before missing has occurred
Applied exponential smoothing method	Average of exponentially smoothed values of the data collected before and after missing has occurred
Factor analysis method	Apply the factor the data from the previous year
Linear interpolation method	Average of the data collected on the days before and after missing has occurred

Table 4. Continuous Traffic Counts Sites for Evaluating the Accuracy of Missing Value Imputation(Year 2010)

Site ID	Address	National Highway Route No.	No. of Lanes	AADT (veh/day)
10002	Yangbok-Rhi, Bogae-Myun, Ansung-Shi, Kyunggi-Do	38	4	26,169
20351	Dongjeom-Dong, Taebaek-Shi, Kangwon-Do	31	2	3,281
30478	Bongso-Rhi, Haksan-Myun, Youngdong-Khoon, Choongchungbuk-Do	19	4	2,138
40009	Eosan-Rhi, Palbong-Myun, Seosan-Shi, Choongchungnam-Do	32	4	22,701
50002	Balsan-Rhi, Gaejeong-Myun, Khoonsan-Shi, Jeonrabuk-Do	26	4	15,089
60008	Youngpoong-Rhi, Seongjeon-Myun, Kangjin-Khoon, Jeonranam-Do	13	4	12,396
70014	Buya-Rhi, Chungdo-Myun, Chungdo-Khoon, Kyungsangbuk-Do	20	2	4,035
80534	Deoknam-Rhi, Chilbuk-Myun, Haman-Khoon, Kyungsangnam-Do	5	4	10,700

Source : Ministry of Land, Transport and Maritime Affairs, 2010 Statistical Yearbook of Traffic Volume, 2011

Table 5. Category of the Data Missing Period Subjected to Evaluation

Category	Missing Period	Site ID	CV	Characteristics
1	Consecutive 7 days (May 30 ~ June 5, 2010)	10002	0.08	The variation of short-term traffic volume is small.
		20351	0.11	
		30478	0.11	
		40009	0.11	
		50002	0.07	
		60008	0.09	
		70014	0.10	
		80534	0.09	
2	Consecutive 7 days (Aug. 1 ~ Aug. 7, 2010)	10002	0.08	The variation of short-term traffic volume is large.
		20351	0.29	
		30478	0.24	
		40009	0.25	
		50002	0.07	
		60008	0.27	
		70014	0.28	
		80534	0.11	

Note : CV(Coefficient of Variation)= $\frac{\text{Standard Deviation of Daily Volume}}{\text{AADT or MADT}} \times 100$

청도면 부아리, 일반국도 5호선 경상남도 함안군 칠북면 덕남리의 8개 지점을 선정하였으며, 평가 자료는 Table 6에서 볼 수 있듯이 총 2가지로 지점별 교통량의 변동이 작은 6월 1주 결측과 변동이 비교적 큰 8월 1주 결측으로 구성하였다.

4.2 결측 보정의 평가 결과

총 보정 횟수는 8개 지점에 대해 4가지 결측 보정 방법을 2가지 기간으로 구분하여 방향별로 구분하여 수행하였다.

모형별 보정 오차는 Fig. 7의 왼편 그림에서 볼 수 있듯이 2010년 6월의 경우 4가지 결측 보정 방법이 16-19%의 오차를 나타내고

있으며 2010년 8월의 경우 19-25%의 오차를 나타내고 있다. Factor법을 제외하고는 교통량의 변동이 작은 6월이 결측 보정 오차가 작게 나타나고 있다. 6월과 8월의 경우를 모두 고려해볼 때 4가지 결측 보정 방법 중 어느 것이 우수하더라는 결론을 도출하기는 어렵다. 왜냐하면 응용지수평활화법의 경우 교통량의 변동이 작을 때에는 오차가 16%로 4가지 결측 보정 방법 중 가장 우수한 결과를 나타내지만 교통량의 변동이 클 때에는 오차가 25%로 가장 나쁜 결과를 나타내기 때문이다.

Fig. 7의 오른편 그림은 8개 지점 각각에 대하여 4가지 결측 보정 방법이 나타내는 오차를 보여주고 있다. 각각의 막대그래프는

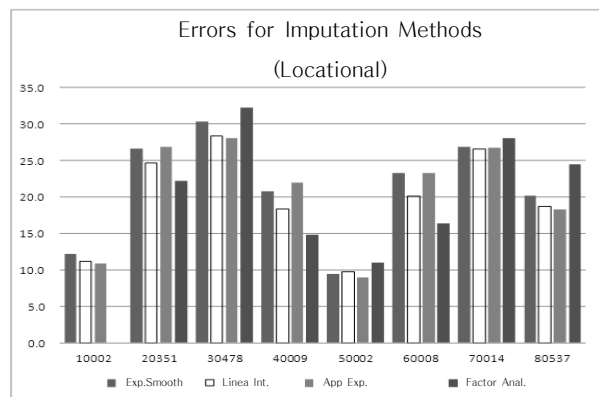
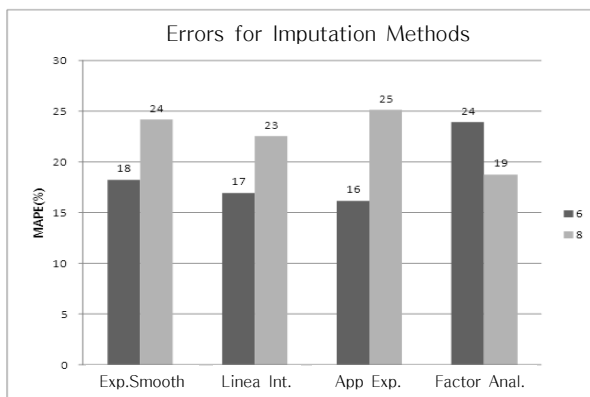


Fig. 7. Errors for the Four Imputation Methods(MAPE : %)

Table 6. Errors for the Four Imputation Methods(MAPE : %)

Site ID	Exponential Smoothing Method	Linear Interpolation Method	Applied Exponential Smoothing Method	Factor Analysis Method	Average
10002	12.2	11.3	10.9	-	11.5
20351	26.7	24.7	26.9	22.2	25.1
30478	30.3	28.3	28.1	32.2	29.7
40009	20.8	18.4	22.0	14.8	19.0
50002	9.4	9.8	8.9	11.1	9.8
60008	23.4	20.2	23.3	16.4	20.8
70014	26.8	26.6	26.8	28.1	27.1
80537	20.2	18.7	18.3	24.5	20.4
Average	21.2	19.7	20.6	21.3	20.7

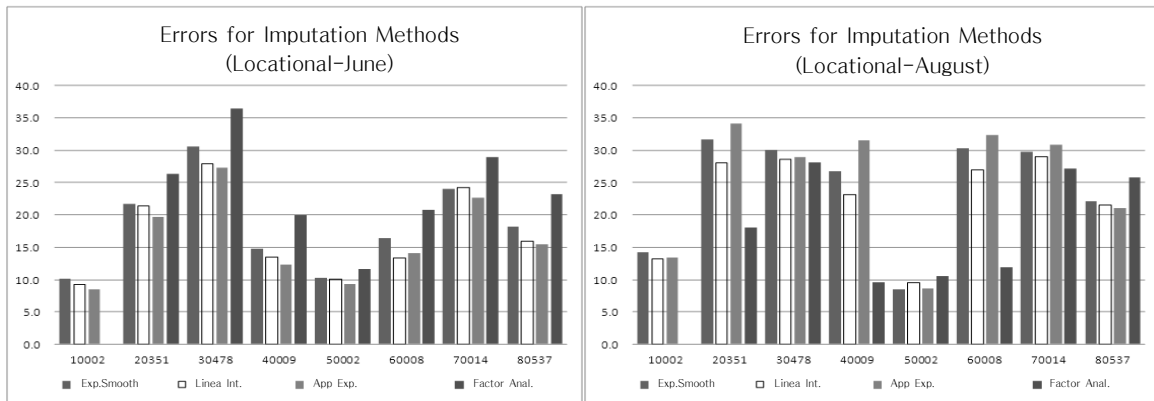


Fig. 8. Monthly-Local Errors for the Four Imputation Methods(MAPE : %)

Table 7. Monthly-Local Errors for the Four Imputation Methods(MAPE : %)

Site ID	Exponential Smoothing Method		Linear Interpolation Method		Applied Exponential Smoothing Method		Factor Analysis Method		Average	
	June	Aug.	June	Aug.	June	Aug.	June	Aug.	June	Aug.
10002	10.2	14.2	9.3	13.2	8.5	13.4	-	-	9.3	13.6
20351	21.7	31.7	21.4	28.0	19.7	34.1	26.4	18.0	22.3	27.9
30478	30.6	30.0	28.0	28.7	27.3	28.9	36.4	28.1	30.6	28.9
40009	14.8	26.7	13.5	23.2	12.4	31.6	20.0	9.7	15.2	22.8
50002	10.3	8.6	10.1	9.5	9.3	8.6	11.7	10.5	10.3	9.3
60008	16.4	30.3	13.3	27.0	14.1	32.4	20.8	12.0	16.2	25.4
70014	24.0	29.7	24.3	29.0	22.6	30.9	28.9	27.2	25.0	29.2
80537	18.2	22.2	15.9	21.5	15.5	21.0	23.2	25.8	18.2	22.6
Average	18.3	24.2	17.0	22.5	16.2	25.1	23.9	18.8	18.7	22.8

6월과 8월의 오차를 평균한 값을 보여주고 있다. 이 막대그래프가 보여주는 결과를 Table화한 것이 Table 6이다.

Table 6에서 볼 수 있듯이 8개 지점을 지점별로 살펴보면 전체 평균은 선형보간법이 보정오차가 19.7%로 가장 작은 것으로 나타

났으며 Factor법이 21.3%로 가장 큰 것으로 나타났으나, 세부 지점별로는 응용 지수평활화법이 4개 지점, Factor법이 3개 지점, 선형보간법이 1개 지점에서 다소 우수한 것으로 나타났다. 이러한 결과를 놓고 보면 본 연구의 대상인 4가지 결측 보정 방법 중

어느 방법이 확실하게 우수하다라는 결론을 내리기에에는 무리가 있을 것이다.

본 연구의 대상인 4가지 결측 보정 방법의 상대적 우위를 분석하기 위하여 지점별 월별로 결측 보정 오차를 좀 더 상세하게 살펴보면 Fig. 8과 Table 7에서 볼 수 있듯이 교통량의 변동이 작은 6월의 경우는 대부분 응용 지수평활법(MAPE : 16.2%), 선형 보간법(MAPE : 17.0%)이 우수한 것으로 나타났으며, 교통량의

변동이 큰 8월은 Factor법(MAPE : 18.8%)이 우수한 것으로 나타났다.

원인은 Factor법은 과거 연도의 자료를 기반으로 적용하는 반면 나머지 보정 방법은 당해 연도의 자료를 입력자료로 활용하기 때문일 것이다. 즉, 8월 1주의 경우 하절기 휴가가 집중적으로 나타나며 이런 휴가철은 연도별로 반복적으로 나타나기 때문이다. 결측 보정 모형과 교통량 변동계수(CV)와 관계를 살펴보면

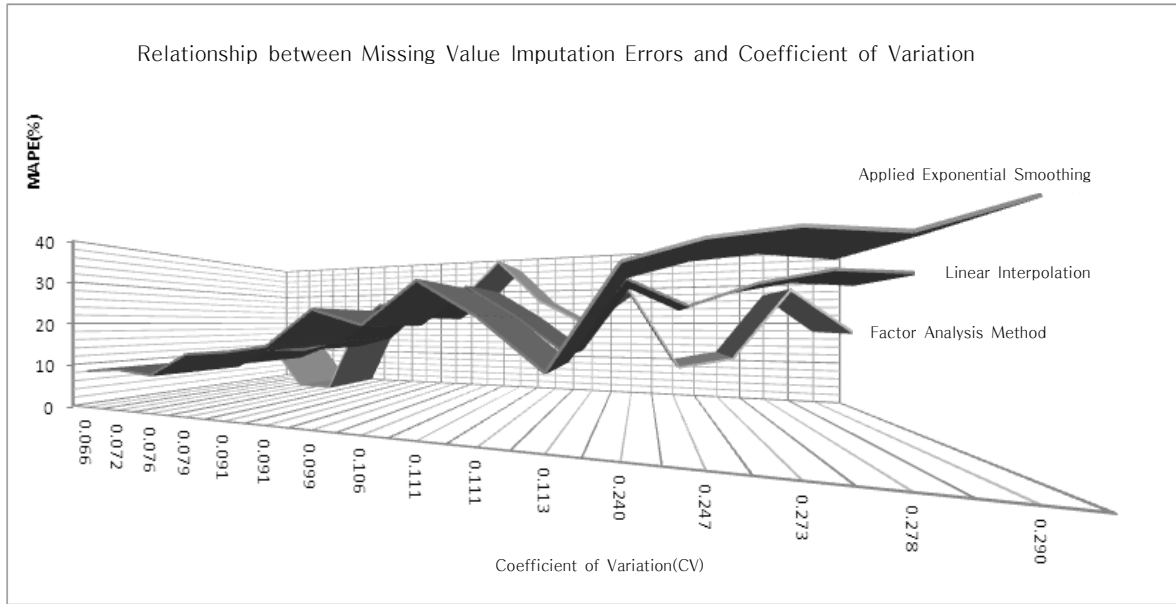


Fig. 9. Relationship Between Missing Value Imputation Errors and Coefficient of Variation

Table 8. Relationship Between Missing Value Imputation Errors and Coefficient of Variation(MAPE : %)

CV	Applied Exponential Smoothing Method	Linear Interpolation Method	Factor Analysis Method	Average
0.066	8.6	9.5	10.5	9.5
0.072	9.3	10.1	11.7	10.4
0.076	8.5	9.3	-	8.9
0.079	13.4	13.2	-	13.3
0.091	14.1	13.3	20.8	16.1
0.091	15.5	15.9	23.2	18.2
0.099	22.6	24.3	28.9	25.3
0.106	19.7	21.4	26.4	22.5
0.111	27.3	28.0	36.4	30.6
0.111	21.0	21.5	25.8	22.8
0.113	12.4	13.5	20	15.3
0.240	28.9	28.7	28.1	28.6
0.247	31.6	23.2	9.7	21.5
0.273	32.4	27.0	12	23.8
0.278	30.9	29.0	27.2	29.0
0.290	34.1	28.0	18	26.7

Fig. 9와 Table 8에서 볼 수 있듯이 결측 보정 오차는 변동계수와의 관계는 선형적으로 비례함을 알 수 있다. 제안된 응용 지수평활화법은 교통량 변동계수(CV)가 0.1이하에서 우수하며, 0.2이상의 경우에는 Factor법이 우수한 것을 확인할 수 있다.

따라서, 지점별-결측 시기의 교통량 변동에 따라 결측 보정을 구분하여 적용하는 것이 보정의 신뢰성을 높일 수 있는 방법일 것이다.

5. 결론 및 향후 연구과제

5.1 결론

각 조사 지점에 고정식 교통량조사 장비를 설치하여 1년 이상의 장기간에 걸쳐 그 지점을 통과한 차량의 차종별 대수 및 총 대수를 조사하는 상시조사는 365일 수행되는 것을 원칙으로 하지만 조사 장비가 차량이 다니는 도로변과 도로표면에 설치되어 있어서 온도, 습도, 먼지, 진동, 비락, 압력 등에 항상 노출되어 있으며 따라서 고장이 자주 발생하게 된다. 이처럼 제어기나 센서가 고장 나거나 파손될 경우 교통량 자료를 수집할 수 없는 상태, 즉 결측이 발생하게 된다.

따라서 상시조사의 경우 현장 조사장비에서 수집되는 원시자료는 대부분의 경우에 결측이 발생하게 되며 이 원시자료는 자료의 진실성을 위하여 결측에 대한 보정 없이 수집된 상태로 유지되어야 한다. 그러나 연간 교통량 통계 작성 등의 분석을 위한 단계에서는 필수적으로 원시자료에 대한 결측 보정이 이루어져야 한다.

이러한 결측 교통량 자료를 보정하는 방법 중 과거연도 자료를 활용하는 Factor 법, 당해 연도 수집 자료를 활용하는 지수평활화법과 응용 지수평활화법, 최근 1개월 이내의 수집 자료를 기초로 한 선형보간법을 대상으로 결측 보정의 정확성 평가를 실시하였다.

평가 결과 해당 지점의 결측 시기별 교통량 변동계수에 따라 보정 오차는 서로 다르게 나타났으며, 변동 계수가 낮은 경우 제안된 응용 지수평활화법이 다소 우수한 결과를 보이고 있다. 그러나 교통량 변동계수가 큰 경우 또는 장기적인 순환변동(여름 휴가철, 특별수송기간)이 강한 경우는 Factor법이 상대적으로 우수하다는 것을 알 수 있었다.

이는 특정한 단일 모형이 모든 지점에서 최적 결과를 보일 수 없으며, 교통량 결측 보정의 신뢰성을 높이기 위해서는 결측 지점, 결측 시기별로 구분하여 적용하는 것이 타당할 것이다.

게다가 교통량 결측 보정 방법은 정상적으로 수집된 과거 자료를 활용하므로 특정한 보정시점의 우연한 변동은 설명하기 어려운 한계가 있으며, 과거 자료의 정제 정도도 중요한 요소로 작용한다.

5.2 향후 연구과제

교통량은 시간과 공간에 따라 끊임없이 변하고 있으며, 이런 변동은 월별, 요일별, 시간대별로 반복적인 현상으로 나타나기도 한다. 또한, 교통량은 날씨, 공사, 사고 등 외적인 환경요소에 의해서도 영향을 많이 받는다. 현재의 결측 보정 방식은 단순히 과거 자료만을 고려한 보정 방식이므로 결측일의 상황을 고려할 수 없다. 예를 들면 폭우, 폭설이 내린 날의 교통량은 기존 교통량보다 일반적으로 작게 나타나며, 공사나 사고 시의 경우도 마찬가지로 나타난다.

따라서, 결측 교통량 보정의 정확성을 제고하기 위해서는 교통량 변화와 외적인 환경요소와의 관계 정립이 필요하며, 상황변수가 포함된 결측 보정 모형에 관한 연구가 필요할 것이다.

본 논문은 한국건설기술연구원의 2013년 주요사업 「효율적인 ITS 성능평가를 위한 Universal 기준장비 개발」의 지원으로 작성되었습니다.

References

- American Association of State Highway and Transportation Officials (AASHTO) (1992). *Guidelines for traffic data programs*.
- Ha, J. A., Park, J. H. and Kim, S. H. (2007). "Missing data imputation using permanent traffic counts on national highways." *Journal of Korean Society of Transportation*, Vol. 25, No. 1, pp. 121-132 (in Korean).
- Hu, Patricia S., Wright, Tommy and Esteve, Tony (1998). "Traffic count estimates for short-term traffic monitoring sites: Simulation Study." *Transportation Research Record*, Vol. 1625, pp. 26-34.
- Korea Institute of Construction Technology (2001). *Developing the Missing Data Imputation Model for Travel Time Estimation and Traffic Volume Statistics* (in Korean).
- Lee, J. Y., Do, M. S., Kim, S. H. and Ryu, S. K. (2003). "Real-time adjustment of traffic volume : Based on the National Highway Route 3." *The Korean Journal of Applied Statistics*, Vol. 16, No. 2, pp. 203-215 (in Korean).
- Lee, S. H., Ha, J. A. and Yoon, T. W. (2012). "A study on improving the national highway traffic counts system : With Focus on Short Duration Counts And Continuous Counts." *Journal of the Korean Society of Civil Engineers*, Vol. 32, No. 3D, pp. 205-212 (in Korean).
- Lingras, P., Sharma, S. and Scherer, W. T. (2003). *Data imputation strategies for transportation management systems*, Research Report No. UVACTS-13-0-80.
- Ministry of Land, Transport and Maritime Affairs (2011). *2010 Statistical yearbook of traffic volume*.
- Sharma, S., Lingras, P. and Zhong, M. (2003). "Effect of missing value imputations on traffic parameters estimations from permanent traffic counts." *Presented at 82nd Annual Meeting of the Transportation Research Board*, Washington DC.

- Weijermars, W. (2007). "Analysis of urban traffic patterns using clustering." *TRAIL Thesis Series T2007/3*, The Netherlands TRAIL Research School.
- Wright, T. and Hu, P. S. (1997). *Variability in traffic monitoring data*, Final Summary Report.
- Zhong, M., Lingras, P. and Sharma, S. (2004). "Estimating of missing traffic counts using factor, generic, neural and regression techniques." *Transportation Research Part C: Emerging Technologies*, 12, pp. 139-166.
- Zhong, M. and Sharma, S. (2009). "Development of improved models for imputing missing traffic counts." *The Open Transportation Journal*, Vol. 3, pp. 35-48.
- Zhong, M., Sharma, S. and Lingras, P. (2006). "Matching patterns for updating missing values of traffic counts." *Transportation Planning and Technology*, Vol. 29, No. 2, pp. 141-156.