

# ITS 시스템의 결측 자료 보정에 관한 연구

A Study of imputation methods for missing data on ITS

**한대철**  
(한국건설기술연구원,  
연구원)

**최속양**  
(한국건설기술연구원,  
연구원)

**문학룡**  
(한국건설기술연구원,  
선임연구원)

**김성현**  
(한국건설기술연구원,  
선임연구원)

Keyword : ITS, 영상검지기, 결측, 누락, 자기회귀, 평가지표

## 목 차

### I. 서 론

#### II. 대상지점 및 데이터의 특성

1. 대상지점의 교통량 및 속도 패턴
2. 대상지점의 교통량 및 속도에 대한 상관관계
3. 손실 데이터의 판단

#### III. 평가지표

1. RMSE(Root Mean Square Error)

#### 2. MARE(Mean Absolute Relative Error)

#### 3. Equality Coefficient

#### IV. 교통량 및 속도 데이터의 보정

1. 과거 이력자료를 이용한 산술평균 보정
2. 유사지점의 데이터를 이용한 자기회귀 보정
3. 평가지표에 의한 평가

#### V. 결론 및 향후 연구과제

## I. 서 론

현재 수도권 남부의 일부 국도에는 ITS 시스템이 구축되어 운영되고 있으며, ITS 시스템을 통해서 도로의 이용자에게 국도의 교통정보를 제공하고 있다. 교통정보의 제공을 위하여 다양한 교통정보 수집장비(매설식 검지기, 비매설식 검지기, 차량번호판 인식장비 등)들이 활용되고 있다. 도로 이용자들에게 보다 신속하고 정확한 교통상황에 대한 정보를 제공하기 위해서는 정확한 자료의 수집이 가장 우선이 되어야 하나, 현실적으로 장비 및 통신의 결함, 악천우, 기타 예측 불허의 상황들로 인해서 교통 자료가 결측되거나 누락되는 일이 발생하기도 한다. 교통정보 제공의 기본이 되는 교통량 및 속도의 자료가 결측되면 정확한 교통정보의 제공이 어렵게 되므로, 결측 자료 보정 문제는 해결해야 할 중요한 문제 중 하나이다.

이에 본 연구에서는 교통정보 제공의 기본이 되는 비매설식 검지기(영상검지기)의 교통수집 자료 중 교통량과 속도 자료에 결측 및 누락이 발생할 경우 결측된 자료를 보정하는 방법에 관한 연구를 하고자 한다.

본 연구를 위해 ITS 시스템이 구축 운영 중인 수도권 남부의 일반국도 중에서 국도 3호선의 곤지암~광주 구간의 영상검지기 자료를 이용하였다. 현재 국도 3호선은 기 구축된 영상검지기를 이용하여 데이터를 수집하고, 이를 근거로 교통정보를 제공하고 있다.

이에 본 연구에서는 ITS 시스템에서 데이터의 결측 및 누

락이 발생하였을 경우 이를 보완하기 위한 방법으로써, 과거 이력자료를 이용한 산술평균 보정, 유사 검지기의 하루 전체 데이터를 이용한 자기회귀 보정을 각각 실시하고자 한다.

각 보정 로직에 따라 보정을 한 다음 실제 관측된 참값과 비교하여 서로의 보정 로직을 비교하고, 각 방법에 따른 예측 능력과 적용 가능성을 검토하고자 한다.

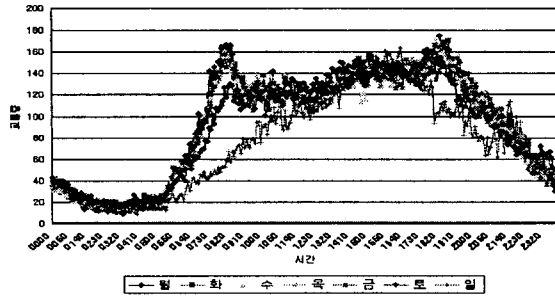
## II. 대상지점 및 데이터의 특성

본 연구에서는 현재 수도권 남부의 국도를 대상으로 운영 중인 ITS 시스템의 비매설식 검지기인 영상검지기의 자료를 이용하였다. 분석대상구간은 국도 3호선의 곤지암~광주 구간이며, 지점수는 인근 검지기의 8지점을 선정하였다.

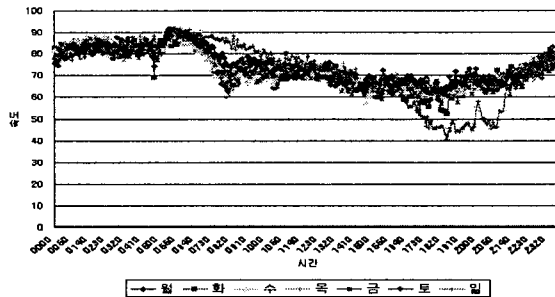
분석에 사용된 교통량 및 속도의 자료는 2004년 6월 27일부터 7월 31일까지의 자료를 이용하였으며, 데이터의 분석시간은 5주 데이터를 5분 단위로 분석하였다.

### 1. 대상지점의 교통량 및 속도 패턴

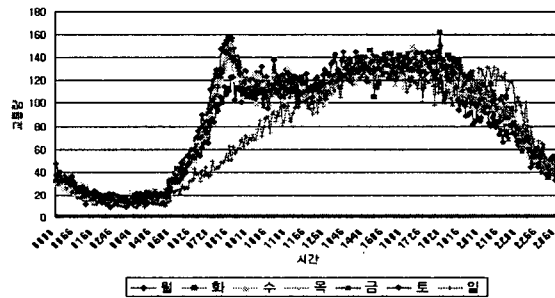
데이터의 특성을 살펴보기 위해서 시간대 및 요일별 교통량, 속도의 패턴을 분석하였다. <그림 1>, <그림 2>, <그림 3>, <그림 4>는 지점7과 지점8의 교통량 및 속도 패턴을 나타낸 것이며, 분석결과 두 지점의 패턴이 이 비슷한 것으로 분석되었다. 지점7, 지점8은 오전, 오후에 교통량은 증가한 것으로 보이거나 첨두 패턴은 없는 것으로 나타났다.



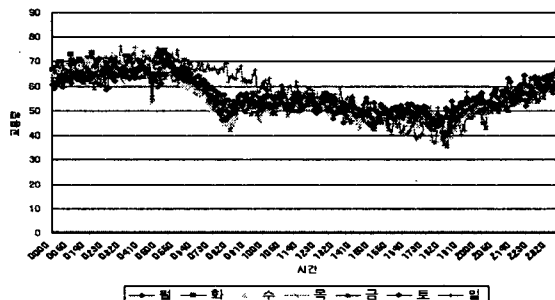
<그림 1> 지점7의 시간대 및 요일별 교통량 패턴



<그림 2> 지점7의 시간대 및 요일별 속도 패턴



<그림 3> 지점8의 시간대 및 요일별 교통량 패턴



<그림 4> 지점8의 시간대 및 요일별 속도 패턴

## 2. 대상지점의 교통량 및 속도에 대한 상관관계

인근 검지기들의 상관성을 바탕으로, 결측이 발생했을 때 인근 검지기에서 관측된 교통량 및 속도 데이터를 이용하여 회귀추정이 가능해지기 때문에 상관분석을 실시하였다. 상관분석은 평일과 주말 교통량, 속도 패턴을 고려하여 평일, 토요일, 일요일로 구분하여 분석을 실시하였다. <표 1>, <표 2>는 대상지점들의 교통량과 속도에 대한 상관관계를 나타낸

것이다.

<표 1> 대상지점의 교통량 상관계수

구분	평일	토요일	일요일
지점1-지점2	0.946	0.965	0.976
지점2-지점3	0.923	0.945	0.960
지점3-지점4	0.989	0.987	0.986
지점4-지점5	0.976	0.969	0.974
지점5-지점6	0.982	0.967	0.970
지점6-지점7	0.993	0.977	0.940
지점7-지점8	0.994	0.988	0.928

<표 2> 대상지점의 속도 상관계수

구분	평일	토요일	일요일
지점1-지점2	0.928	0.872	0.903
지점2-지점3	0.167	0.258	0.560
지점3-지점4	0.670	0.528	0.825
지점4-지점5	0.774	0.752	0.913
지점5-지점6	0.561	0.775	0.849
지점6-지점7	0.965	0.923	0.872
지점7-지점8	0.940	0.873	0.844

교통량의 경우 지점들간 상관계수가 0.9이상으로 높게 분석된 반면, 속도의 경우는 지점1-지점2(평일, 일요일), 지점4-지점5(일요일), 지점6-지점7(평일, 토요일), 지점7-지점8을 제외하고 상관계수가 낮은 것으로 분석되었다.

## 3. 손실 데이터의 판단

Kolmogorov-Smirnov 검정과 Shapiro-Wilk의 검정 결과 유의 확률이 유의수준 0.05를 초과하므로 5분 교통량 및 속도의 데이터가 정규분포를 따른다고 볼 수 있으므로 시점별로 수집되는 데이터의 평균과 표준편차를 계산하여 주어진 신뢰도에서 손실값을 판단할 수 있다. 즉, 시점  $i$  ( $i=1, \dots, 288$ )에서 직전 5주 평일, 토요일, 일요일로 얻어진 데이터 개수가  $n_i$ 이고 평균이  $\bar{x}_i$ , 표준편차가  $s_i$ 라면,

$$(\bar{x}_i - 3s_i, \bar{x}_i + 3s_i)$$

의 범위를 초과하는 교통량 데이터는 신뢰도 99%로 손실 데이터로 판단한다.

정규성 가정이 힘든 경우에는 일반적인 통계 데이터의 이상값 판단으로써, 데이터의 제1사분위 수와 제3사분위 수로 사분위 범위의 1.5배를 벗어난 데이터를 손실 데이터로 판단한다. 이때 1사분위 수는 데이터의 크기 순으로 나열했을 때 하위 25%에 해당하는 데이터이고, 제3사분위 수는 상위 25%에 해당하는 데이터이다. 그리고 사분위 범위란 제3사분위 수 - 제1사분위 수로 얻어진다. 즉, (제1사분위 수 - 1.5 × 사분위 범위, 제3사분위 수 + 1.5 × 사분위 범위)의 범위를 초과하는 데이터를 손실 데이터로 판단한다.

### III. 평가지표

보정의 정확성을 상호평가하기 위하여 관측값과 보정값의 차이를 나타내는 RMSE, MARE, Equality Coefficient를 척도로 사용하였으며, 내용은 다음과 같다.

#### 1. RMSE(Root Mean Square Error)

평균제곱오차제곱근(RMSE)은 오차가 큰 곳에 가중치를 주어 평균하는 지표로서 이는 모형의 개발이나 분석에 널리 사용되어 진다. 수식은 식 (1)과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

여기서,  $n$  = 데이터의 개수

$y_i$  = 보정값

$\hat{y}_i$  = 관측값

#### 2. MARE(Mean Absolute Relative Error)

평균절대관계오차(MARE)는 변동계수 추정치가 실제 데이터에 비해 평균적으로 얼마나 떨어져 있는가를 나타내는 지표로서, 수식으로 표현하면 식 (2)와 같다.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (2)$$

여기서,  $n$  = 데이터의 개수

$y_i$  = 보정값

$\hat{y}_i$  = 관측값

#### 3. Equality Coefficient

등가계수(EC)는 타일(H, Theil, 1966)에 의해서 개발된 방법으로 계량경제모형에서 구한 예측값의 정확성에 대한 척도로서 식 (3)과 같다.

$$EC = 1 - \frac{\sqrt{\sum_{i=1}^n [Y_i - X_i]^2}}{\sqrt{\sum_{i=1}^n Y_i^2} + \sqrt{\sum_{i=1}^n X_i^2}} \quad (3)$$

여기서,  $Y_i$  =  $i$ 시간대의 기준값

$X_i$  =  $i$ 시간대의 관측값

### IV. 교통량 및 속도 데이터의 보정

#### 1. 과거 이력자료를 이용한 산술평균 보정

과거 이력자료를 이용한 교통량 및 속도의 산술평균 보정은 직전 5주의 평일(월, 화, 수, 목, 금) 자료를 모

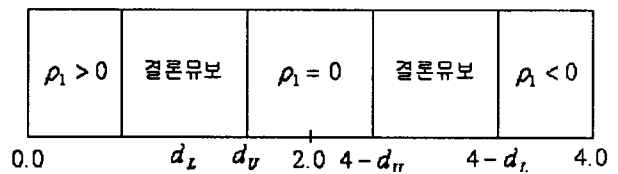
두 포함한 산술평균 보정과 직전 5주의 토요일 자료만 이용한 보정으로 구분하여 분석하였다. 산술평균에 대한 척도를 식으로 표현하면 식 (4)와 같다.

$$V_t = \frac{\sum_{i=1}^n V_i}{N} \quad (4)$$

#### 2. 유사지점의 데이터를 이용한 자기회귀 보정

지점들 간의 유사지점을 판단하기 위하여 상관분석을 실시하였으며, 상관분석 결과는 <표 1>, <표 2>와 같다. 결측이 발생한 시점의 직전 5주의 평일, 토요일, 일요일 5분 과거이력 자료를 이용하여 결측 지점의 교통량 및 속도를 종속변수로, 유사 지점의 교통량 및 속도를 독립변수로 하는 자기회귀 보정을 실시하였다. 자기회귀 보정에서 자기상관을 검정하기 위한 방법으로는 Durbin-Watson이 있으며, 통계량은 식 (5)와 같으며, 검정 기준은 그림 (5)와 같다.

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \approx 2(1 - r_1) \quad (5)$$



<그림 5> Durbin-Watson 검정 기준

자기상관에 대한 Durbin-Watson 검정방법은 오차항의 일차 자기상관만 검정할 수 있고, 또 표본의 크기가 어느 정도 이상 많아야 할 뿐만 아니라 검정결과 결론유보의 영역이 나타나는 단점을 가지고 있으나 최근까지 가장 많이 사용되는 방법이다. Durbin-Watson의 통계량은 오차항의 자기상관 모수  $\rho$ 가 0인가를 검정하는 것으로, 만약  $\rho=0$ 이면, 오차항은 서로 독립이 된다.

자기회귀의 방법은 일반적으로 많이 사용되고 있는 코크란-오커트(Cochrane-Orcutt)의 방법을 사용하였으며, 교통량의 경우 독립변수의 교통량 데이터의 변환을 통해 그 결정계수를 더 높일 수 있다. 즉, 제곱근이 변환을 통해 결정계수를 높일 수 있다. 교통량과 속도에 대한 자기회귀 모형식은 식 (6), 식 (7)과 같다.

교통량의 자기회귀 모형식은 시점  $t$ 에서의 종속변수를  $Y_t$ , 독립변수를  $X_t$ 라고 했을 때,

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad \varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (6)$$

인 자기회귀 모형을 가정한다.

여기서,  $\rho$ 는 자기 상관모수로서  $|\rho| < 1$

$u_t$ 는 새로운 오차항으로  $u_t \sim N(0,1)$ 로 가정

<그림 6>은 평일 지점8의 교통량 데이터를 종속변수  $Y_t$ , 지점7의 교통량의 제곱근을 독립변수  $X_t$ 로 했을 때 얻어지는 SPSS의 코크란-오커트방법의 결과이다. 분석결과 Durbin-Watson 의 값이 2.43671로 오차들간에 자기상관이 존재하지 않으므로 회귀분석을 이용하여도 된다.

Rho .74745601  
Standard Error of Rho .03935041

Cochrane-Orcutt Estimates

Multiple R .94505967  
R-Squared .89313778  
Adjusted R-Squared .89238523  
Standard Error 3.9570211  
Durbin-Watson 2.43671

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	37354.389	37354.389
Residuals	284	4469.381	15.737

Variables in the Equation:

	B	SEB	BETA	T	SIG T
Xt	14.647750	.3006524	.94505967	48.719890	.0000000
CONSTANT	-48.603559	2.9228615		-16.628759	.0000000

<그림 6> 지점7-지점8의 교통량 자기회귀 모형 분석

따라서 다음과 같은 자기회귀 모형을 얻을 수 있다.

$$Y_t = -48.6035 + 14.6477X_t + \varepsilon_t$$

$$\varepsilon_t = 0.7474\varepsilon_{t-1} + u_t$$

만일 시점  $t$ 에서 지점8의 교통량 데이터가 결측되면 동일 시점에서 정상적으로 얻어지는 지점7의 교통량 데이터를 이용하여 위의 자기회귀 모형에 의해 다음과 같이 보정할 수 있다.

$$\hat{y}_t = -48.6035 + 14.6477x_t + \hat{\varepsilon}_t$$

$$\text{단, } \hat{\varepsilon}_t = 0.7474\hat{\varepsilon}_{t-1}$$

$$= 0.7474[y_{t-1} - (-48.6035 + 14.6477x_{t-1})]$$

이다.

속도의 자기회귀 모형식은 결측 지점의 시점  $t$ 에서의 평균 속도를 종속변수  $Y_t$ , 유사 지점의 평균 속도를  $X_t$ , 같은 유사 지점의 교통량을  $Z_t$ , 그리고 정체 시간을 나타내는 더미변수를  $T_t$ 라고 했을 때,

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + \beta_3 T_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (7)$$

인 자기회귀 모형을 가정한다.

여기서,  $\rho$ 는 자기 상관모수로서  $|\rho| < 1$

$u_t$ 는 새로운 오차항으로  $u_t \sim N(0,1)$ 로 가정

$T_t = 1$  :  $t$ 가 정체 시간대에 포함

$T_t = 0$  :  $t$ 가 정체 시간대에 포함되지 않음

<그림 7>은 평일 지점8의 평균속도 데이터를 종속변수  $Y_t$ , 지점7의 평균속도 데이터를 독립변수  $X_t$ , 지점7의 교통량 데이터를 독립변수  $Z_t$ , 정체시간을 나타내는 더미변수를  $T_t$ 로 했을 때 얻어지는 SPSS의 코크란-오커트방법의 결과이다. 분석결과 Durbin-Watson 의 값이 1.9692027로 오차들간에 자기상관이 존재하지 않으므로 회귀분석을 이용하여도 된다.

Rho .36866822  
Standard Error of Rho .0552569

Cochrane-Orcutt Estimates

Multiple R .97401106  
R-Squared .94869754  
Adjusted R-Squared .94796984  
Standard Error 1.0953613  
Durbin-Watson 1.9692027

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	3	6268.2469	2089.4156
Residuals	282	336.9663	1.2020

Variables in the Equation:

	B	SEB	BETA	T	SIG T
Xt	.298608	.0270976	.32017992	11.019753	.0000000
Zt	-.102881	.0044393	-.67158029	-23.175222	.0000000
Tt	-4.122861	.8921995	-.06271705	-4.621008	.0000582
CONSTANT	43.642738	2.3615917		18.480222	.0000000

<그림 7> 지점7-지점8의 속도 자기회귀 모형 분석

따라서 다음과 같은 자기회귀 모형을 얻을 수 있다.

$$Y_t = -43.6427 + 0.2986X_t - 0.1028Z_t - 4.1228T_t + \varepsilon_t$$

$$\varepsilon_t = 0.3686\varepsilon_{t-1} + u_t$$

만일 시점  $t$ 에서 지점8의 속도 데이터가 결측되면 동일 시점에서 정상적으로 얻어지는 지점7의 속도 데이터를 이용하여 위의 자기회귀 모형에 의해 다음과 같이 보정할 수 있다.

$$\hat{y}_t = -43.6427 + 0.2986x_t - 0.1028z_t - 4.1228t_t + \hat{\varepsilon}_t$$

$$\text{단, } \hat{\varepsilon}_t = 0.3686\hat{\varepsilon}_{t-1}$$

$$= 0.3686[y_{t-1} - (-43.6427 + 0.2986x_{t-1} - 0.1028z_{t-1} - 4.1228t_{t-1})]$$

이다.

### 3. 평가지표에 의한 평가

과거 이력자료를 이용한 산술평균 보정에서 직전 5주의 평일 자료를 모두 포함한 산술평균 보정을 보정1, 직전 5주의

동요일 자료만 이용한 보정을 보정2, 유사지점의 데이터를 이용한 자기회귀 보정을 보정3이라고 명시하였다.

본 연구에서 제시된 평가지표와 비교하기 위해서 실제로 검지기가 측정한 2004년 8월 1일부터 2004년 8월 7일까지의 지점7과 지점8의 교통량과 속도에 대한 데이터와 비교·분석하였다. 평가에 대한 분석결과는 <표 3>, <표 4>, <표 5>, <표 6>과 같다.

<표 3> 지점7의 교통량 평가 결과

구분		보정1	보정2	보정3
8월 1일(일)	RMSE	21.822	19.032	18.709
	MARE	0.166	0.142	0.138
	EC	0.909	0.917	0.918
8월 2일(월)	RMSE	13.809	13.281	14.347
	MARE	0.109	0.102	0.112
	EC	0.944	0.947	0.942
8월 3일(화)	RMSE	16.037	17.295	16.506
	MARE	0.107	0.118	0.109
	EC	0.936	0.932	0.935
8월 4일(수)	RMSE	10.206	11.708	10.179
	MARE	0.067	0.078	0.060
	EC	0.959	0.953	0.959
8월 5일(목)	RMSE	9.535	11.120	9.444
	MARE	0.063	0.073	0.056
	EC	0.962	0.955	0.962
8월 6일(금)	RMSE	9.666	12.008	9.469
	MARE	0.078	0.099	0.075
	EC	0.960	0.932	0.961
8월 7일(토)	RMSE	14.826	13.459	13.435
	MARE	0.111	0.101	0.093
	EC	0.940	0.946	0.946

<표 4> 지점8의 교통량 평가 결과

구분		보정1	보정2	보정3
8월 1일(일)	RMSE	23.387	19.361	18.911
	MARE	0.189	0.168	0.169
	EC	0.895	0.912	0.914
8월 2일(월)	RMSE	15.883	15.488	16.380
	MARE	0.124	0.122	0.129
	EC	0.931	0.933	0.929
8월 3일(화)	RMSE	13.241	14.143	13.233
	MARE	0.101	0.107	0.099
	EC	0.942	0.939	0.942
8월 4일(수)	RMSE	11.926	6.191	12.555
	MARE	0.087	0.015	0.090
	EC	0.948	0.974	0.945
8월 5일(목)	RMSE	12.292	13.301	12.255
	MARE	0.087	0.093	0.086
	EC	0.947	0.942	0.947
8월 6일(금)	RMSE	11.309	13.441	11.197
	MARE	0.106	0.121	0.105
	EC	0.949	0.878	0.950
8월 7일(토)	RMSE	26.922	28.070	19.456
	MARE	0.242	0.257	0.111
	EC	0.882	0.876	0.917

<표 5> 지점7의 속도 평가 결과

구분		보정1	보정2	보정3
8월 1일(일)	RMSE	7.7075	12.4718	12.3007
	MARE	0.0836	0.1428	0.1403
	EC	0.9459	0.9104	0.9117
8월 2일(월)	RMSE	6.5899	6.0886	6.7123
	MARE	0.0709	0.0646	0.0730
	EC	0.9533	0.9572	0.9524
8월 3일(화)	RMSE	6.6040	6.7251	6.6373
	MARE	0.0734	0.0772	0.0736
	EC	0.9534	0.9526	0.9532
8월 4일(수)	RMSE	7.5192	7.6124	7.5045
	MARE	0.0823	0.0834	0.0810
	EC	0.9473	0.9466	0.9474
8월 5일(목)	RMSE	6.6784	6.8750	6.6972
	MARE	0.0755	0.0758	0.0747
	EC	0.9529	0.9516	0.9528
8월 6일(금)	RMSE	6.6249	7.4607	6.7103
	MARE	0.0778	0.0853	0.0796
	EC	0.9522	0.9308	0.9516
8월 7일(토)	RMSE	8.5172	9.0135	19.9149
	MARE	0.1066	0.1116	0.2676
	EC	0.9389	0.9358	0.8389

<표 6> 지점8의 속도 평가 결과

구분		보정1	보정2	보정3
8월 1일(일)	RMSE	7.2256	7.2670	7.4631
	MARE	0.1343	0.1390	0.1422
	EC	0.9294	0.9297	0.9277
8월 2일(월)	RMSE	5.4774	6.0355	5.6188
	MARE	0.0974	0.1101	0.1010
	EC	0.9466	0.9421	0.9452
8월 3일(화)	RMSE	5.3902	5.6400	5.5665
	MARE	0.0972	0.1015	0.0983
	EC	0.9471	0.9450	0.9454
8월 4일(수)	RMSE	4.6466	4.7116	4.6839
	MARE	0.0743	0.0730	0.0758
	EC	0.9550	0.9541	0.9547
8월 5일(목)	RMSE	5.1118	5.5888	5.2377
	MARE	0.0870	0.0919	0.0863
	EC	0.9502	0.9453	0.9490
8월 6일(금)	RMSE	5.9750	6.4072	5.9094
	MARE	0.1025	0.1112	0.1022
	EC	0.9417	0.9344	0.9423
8월 7일(토)	RMSE	6.5067	7.3232	7.2384
	MARE	0.1195	0.1360	0.1330
	EC	0.9360	0.9284	0.9292

교통량의 경우 토요일과 일요일은 보정3이 가장 유의한 것으로 분석되었다. 평일에도 보정3이 유의한 것으로 분석되었지만, 지점7의 경우 월요일에는 보정2가, 화요일에는 보정 1이 유의한 것으로 나타났으며, 지점8의 경우 월요일과 수요일에 보정2가 유의한 것으로 나타났다. 속도의 경우 지점7과 지점8에서 보정1이 월요일과 수요일을 제외하고는 가장 유의한 것으로 분석되었다.

그러므로, 교통량은 평일, 토요일, 일요일에 대해서 보정3을 적용하는 것이 적합하다고 판단되어지며, 속도의 경우는 보정1을 적용하는 것이 적합하다는 결론을 얻을 수 있다.

## V. 결론 및 향후 연구과제

본 연구는 ITS 시스템의 요소인 비매설식 영상검지기를 대상으로 교통량과 속도에 대한 결측 및 누락 데이터를 보정하기 위한 연구이다. 영상검지기는 교통정보 수집장비로써 도로의 이용자들에게 국도의 교통정보를 제공하기 위해 활용되고 있다. 도로 이용자들에게 보다 신속하고 정확한 교통상황에 대한 정보를 제공하기 위해서는 정확한 자료의 수집이 무엇보다도 중요하나 장비 및 통신의 결함, 기상여건 등 예측불허의 상황들로 인해 교통자료가 결측되거나 누락되는 일이 발생한다.

이에 본 연구에서는 교통정보 제공의 기본이 되는 영상검지기의 교통수집 자료 중 교통량과 속도 자료의 결측이나 누락이 발생했을 때 보정하기 위하여 연구를 수행하였다.

보정방법은 과거 이력자료를 이용한 산술평균 보정과 유사 지점의 데이터를 이용한 자기회귀 보정방법으로 분석을 실시하였다. 분석결과 교통량은 유사지점의 데이터를 이용한 자기

회귀 보정이 과거 이력자료를 이용한 산술평균 보정 방법보다 유의한 것으로 나타났으며, 속도는 과거 이력자료를 이용한 산술평균 보정 중 직전 5주의 평일 자료를 이용한 방법이 가장 유의한 것으로 분석되었다. 교통량의 경우 데이터가 정상적인 패턴을 유지하는 상황이라면 결측된 지점에 대해서 자기회귀 보정 방법을 이용하여 유사지점의 교통량 데이터를 보정할 수 있다고 여겨지며, 속도의 경우는 과거 이력자료를 이용한 산술평균 보정 방법으로 결측된 지점의 속도 데이터를 보정할 수 있다고 판단된다.

본 연구에서는 교통량과 속도에 대해서 과거 이력자료를 이용한 산술평균 보정 방법과 자기회귀 보정 방법만 이용하여 결측된 데이터의 보정을 고려하였으나, 향후 신경망 모형 등 다양한 방법을 이용하여 결측 및 누락 데이터에 대한 보정 방법의 연구가 이루어져야 될 것으로 보인다.

## 참고 문헌

1. 건설교통부, 한국건설기술연구원, 영남대학교 공저, 통행시간 추정과 통계자료 보정을 위한 결측 데이터 보정 모형 제작, 2001
2. 장진환, 류승기, 문학룡, 변상철, 실시간 누락 교통자료의 대체기법에 관한 연구, 2003
3. 이지연, 도명식, 김성현, 류승기, 실시간 교통량 데이터의 결측값 보정, 2003
4. 유지성, 오창수 공저, 현대통계학, 1995
5. 김명학, 손상호 공저, 한글 SPSSWIN을 이용한 통계분석, 2001
6. Satish Sharma et.al. TRB 2003 Annual Meeting, Effect of Missing Value Imputation on Traffic Parameters Estimations from Permanent Traffic Count, 2003
7. Chao Chen et al. TRB 2003 Annual Meeting, Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems, 2003