■ 論 文 ■

차량 검지자료 결측 보정처리에 관한 연구
(이력자료 활용방안을 중심으로)

A Study on the Imputation for Missing Data in Dual-loop Vehicle Detector System

김 정 연
(서울대학교 환경대학원 교통관리전공 석사과정)

이 영 인
(서울대학교 환경대학원 교통관리전공 교수)

백 승 걸
(한국도로공사 도로교통기술원 책임연구원)

남 궁 성
(한국도로공사 도로교통기술원 수석연구원)

목 차

Ⅰ. 서론
1. 연구의 배경 및 필요성
2. 연구의 목적
Ⅱ. 국·내외 결측 보정처리 현황 및 선행연구고찰
1. 결측 보정처리 정의
2. 국내·외 결측 보정처리 현황
3. 선행연구 고찰
Ⅲ. 방법론 정립
1. 기본 전제
2. 적용 방법론 소개
3. 평가지표
Ⅳ. 결측 보정처리 적용
1. 적용 사례지점 선정
2. 자료 구축
3. 방법론별 보정처리 결과 분석
4. 분석결과 및 시사점
Ⅴ. 결론 및 향후연구과제
1. 결론
2. 향후 연구과제
참고문헌

Key Words : 결측자료, 보정처리, 지능형 교통시스템, 고속도로 교통관리시스템, 차량검지시스템
missing data, imputation, ITS, FTMS, VDS

요 약

교통정보는 운영 중인 VDS(Vehicle Detector System)를 통해 수집된 교통량, 속도, 점유율 자료를 바탕으로 가공되어 제공된다. 수집된 자료의 응용범위와 이용자 및 활용분야는 점점 증가하는 추세에 있다. 수집된 차량 검지자료 내의 결측자료는 대상의 속성을 포함하지 않은 채 전송되는 일련의 빈 데이터를 말한다. 결측자료는 속성 값을 갖고 있지 않은 관계로 데이터 가공 처리대상에서 제외된다. 결국 수집 데이터 내의 결측자료의 비율이 증가할수록 해당 지점의 교통상황을 반영하는데 있어 신뢰성이 결여되는 문제점을 갖게 된다. 본 연구에서는 결측자료에 대해 인접지점 참조방식과 이력자료를 활용한 방법론을 적용한 보정처리 결과를 제시하였다. 현재 운영 중인 서해안/경부고속도로의 VDS 자료에 임의의 결측자료 비율을 처리 후 보정처리 방법론을 적용하였다. 보정단위는 차로단위-30초 주기로 수행 하였으며 오전/오후/일단위로 구분하여 원데이터 대비 보정된 값의 오차를 분석하였다. 분석결과 인접지점 참조방식에 비해 이력자료를 활용한 보정처리 방법이 원데이터에 가까운 값을 도출하는 것을 알 수 있었다.

The traffic information is provided, which based on the volume of traffic, speed, occupancy collected through the currently operating Vehicle Detector System(VDS). In addition to the trend in utilization field of traffic information is increasing gradually with the applied various fields and users. Missing data in Vehicle detector data means series of data transmitted to controller without specific property. The missing data does not have a data property, so excluded at the whole data process. Hence, increasing ratio of missing data in VDS data inflicts unreliable representation of actual traffic situation. This study presented the imputation process due out which applied the methodologies that utilized adjacent stations reference and historical data utilize about missing data. Applied imputation process methodologies to VDS data of SeoHaeAn/Kyongbu Expressway, currently operation VDS, after processes at missing data ratio of an option. Imputation process held presented to per lane-30seconds-period, and morning/afternoon/daily time scope ranges classified, and analyzed an error of imputed data preparing for actual data. The analysis results, an low error occurred relatively in the results of the imputation process way that utilized a historical data compare with adjacent stations reference methods.

Ⅰ. 서론

1. 연구의 배경 및 필요성

교통정보는 운영 중인 VDS(Vehicle Detector System)를 통해 수집된 교통량, 속도, 점유율을 바탕으로 목적에 따라 다양하게 가공되어 제공된다. 제공정보는 구간통행 속도 및 시간을 비롯하여 전방 교통상황, 돌발상황 관리, 연간 교통량 자료에 이르기까지 그 응용범위와 이용자 및 활용분야가 점점 증가하는 추세이다.

차량 검지자료 내의 결측자료(missing data)는 대상의 속성(data property)을 포함하지 않은 채 전송되는 일련의 빈 데이터(non-responsive, missing)를 의미하며 루프 및 제어기와 같은 검지 관련 장비의 노후, 통신단선 등과 같은 다양한 원인에 의해 발생된다. 결측자료는 속성 값을 갖고 있지 않은 관계로 데이터 가공 처리대상에서 제외된다. 결국 수집 데이터 내의 결측자료의 비율이 증가할수록 해당 지점의 교통상황을 반영하는데 신뢰성(reliability)이 결여되는 문제점이 발생하게 된다. 예를들어, 지점단위 자료를 바탕으로 하는 정보가공 체계에 있어 차로단위의 결측발생 비중이 증가하게 될 경우 가공시 분석의 기본이 되는 샘플 데이터 수의 결여가 발생하게 된다. 이로 인해 해당 지점의 대푯값과 구간정보 산출시 실제 상황에 비해 왜곡된 가공결과가 도출될 가능성이 높아지게 된다.

2. 연구의 목적

결측자료에 대한 보정처리는 결측내역을 신뢰성 있는 대체 자료로 처리해 주는 일련의 과정이라 할 수 있다. 보정처리를 거친 자료는 결측내역을 처리하지 않은 경우에 비해 분석의 기반이 되는 데이터 셋(data set)의 기본 총량을 갖추게 되므로, 자료 가공시 완성도가 높은 결과를 산출하게 된다. 본 연구에서는 기 적용 중인 국/내외의 다양한 결측보정 방법론과 제안된 방법론에 대한 적용 사례와 특성을 제시하며, 보정처리방식에 있어 기반이 되는 데이터를 기준으로 하여 방법론을 분류한다. 또한 기 운영 중인 VDS를 통해 수집된 자료를 대상으로 결측내역에 대한 보정처리를 수행하며 도출된 결과를 비교·평가하는 과정을 제시한다.

Ⅱ. 국내·외 결측 보정처리 현황 및 선행연구고찰

1. 결측 보정처리 정의

1) 차량 검지자료의 결측자료

VDS를 통해 수집된 자료는 완전한 데이터 셋을 갖추어야 하지만 결측자료를 포함하고 있는 경우도 발생한다. 오류자료(erroneous data)의 경우 수집이후 이상치(outlier)로 분류되는 반면 결측자료(missing data)는 수집 자체가 전혀 이루어지지 않은 값이라는 점에서 서로 구분되어진다. VDS에서의 결측은 주로 통신단선, 장비 노후 및 기타 다른 요인에 의해 발생한다. VDS에서의 결측자료는 MCAR(Missing Completely At Random) 상황으로 어느 특정 항목이 언제 결측 되는지와 상관없이 임의로 발생하며, 다른 개체의 특성과 무관하게 시공간상에서 무작위로 발생하는 경우에 해당 된다고 볼 수 있다.

2) 결측발생 현황

결측자료는 수집 폴링주기(poll cycle) 동안 자료가 전송되지 않은 경우로, 수집주기 동안 차량이 통과하지 않을 경우 전송하는 기본값(default value)인 교통량 '0' 점유율 '0'과는 다르다. 즉, 수집과정이 일절 진행되지 않음으로 인해 발생한 수집주기상의 빈 공간을 의미한다. 국내 유수한 유관기관 중 하나인 한국도로공사의 고속도로교통관리시스템에서는 수집자료 내의 결측자료를 〈표 1〉~〈표 2〉와 같이 처리한다. 차량검지자료의 기본적인 시간집적 단위는 30초 주기로 진행된다. 수집된 가공전 원시자료를 자료 속성별 기준값(threshold) 범위 여부의 오류 필터링(filtering) 과정을 거친 이후 가공후 내역에 해당 결측내역을 '-999'와 같은 비 논리적인

〈표 1〉 원시자료 가공 전 내역

SAMPLE_TIME	교통량	점유율	속도
20060205162311	1	1	81
20060205162410	1	1	107
20060205162441	4	6	76

〈표 2〉 원시자료 가공 후 결측 구분 전송 내역

SAMPLE_TIME	교통량	점유율	속도
20060205162301	1	1	81
20060205162330	-999	-999	-999
20060205162401	1	1	107

〈표 3〉 경부선 연속 7개 지점 결측 발생 빈도

ID	2/4(토)	2/5(일)	2/6(월)	2/7(화)	2/8(수)	2/9(목)	2/10(금)	2/11(토)	2/12(일)
LD3552	847	-	-	32	148	1,144	40	-	24
LD3563	23,040	23,040	23,040	23,040	23,040	23,040	23,040	23,040	23,040
LD3585	6,496	4,032	4,504	5,192	4,528	6,792	6,768	5,848	64,976
LD3597	844	8	-	8	144	1,144	32	-	16
LD3624	836	16	8	112	256	1,272	136	32	64
LD3647	2,012	8	32	8	280	1,344	88	48	40
LD3658	831	-	-	8	144	1,144	48	16	8

〈표 4〉 서해선 연속 7개 지점 결측 발생 빈도

ID	2/4(토)	2/5(일)	2/6(월)	2/7(화)	2/8(수)	2/9(목)	2/10(금)	2/11(토)	2/12(일)
LD2980	660	20	12	12	120	858	36	36	18
LD2990	616	14	12	30	126	882	42	42	36
LD3012	740	63	58	307	203	965	123	123	193
LD3027	17,280	17,280	17,280	17,280	17,280	17,280	17,280	17,280	17,280
LD3042	5,783	4,855	5,029	4,705	4,541	5,556	5,027	5,027	5,006
LD3056	935	40	35	557	229	928	83	83	94
LD3073	3,188	2,405	2,614	2,576	2,631	3,361	2,682	2,682	2,540

값으로 전송하여 정상적으로 수집된 데이터와 구분한다.

〈표 3〉~〈표 4〉는 2006/02/04~2006/02/12 동안 기 운영 중인 경부선과 서해안선의 각 20개소, 총 40개소의 지점 검지기에서 결측 발생 빈도를 조사한 결과이다. 결측 발생은 1일단위로 작게는 0(건/지점)에서 경부선의 경우 최대 4개 차로 23,040(건/지점), 서해안선의 경우3개 차로 17,280(건/지점)으로 나타났다.

3) 보정처리 과정

보정처리 과정은 결측자료를 적합한 과정을 통해 얻어진 추정된 자료(estimate data)로 대체시키는 과정이다. 보정처리 단계가 존재하지 않거나 적합한 처리과정이 적용되지 않을 경우 수집자료 내의 결측내역은 가공 처리된 결과에 상당한 영향을 미치게 된다. 결측빈도가 높을 경우 지점 정보 집계 및 구간 정보 산출시 적합 표본수의 미달에 따른 수집자료 운영의 비효율을 초래하며, 결측내역을 제외한 나머지 자료를 바탕으로 가공시 실제 상황에 비해 왜곡된 결과가 산출될 가능성이 높다.

4) 보정처리 효과

〈표 5〉는 결측이 발생하지 않은 'LD3712' (안성IC 부근, 상행)검지기의 2006/02/06(월)의 오전(08:00)~오후(14:00), 총 6시간 동안 수집된 교통량 자료에 대하여 보정처리 시행에 따른 기대효과를 간략히 제시한 것이다. 결측이 발생하지 않은 자료와, 임의의 결측비율을 적용한 각각의 자료를 지점단위 5분 자료로 가공하여 결측비율에 따른 실제 대비 오차발생을 비교한 결과이다. 결측자료를 포함하는 경우의 RMSE[8]는 평균치에 비해 잔차의 분산정도가 심한 것을 알 수 있는데, 이런 경우 구간정보 산출시 올바르게 산출되지 못한 지점정보를 포함하게 되므로 왜곡된 결과가 산출될 가능성이 매우 높다. 위의 결과를 역으로 생각하여 결측이 존재하는 자료에 적절한 보정처리 과정을 수행할 경우 가공자료의 신뢰성 확보를 기대할 수 있다.

〈표 5〉 임의 결측 발생비율에 따른 오차 분석

결측 발생비율	MARE	MAE	RMSE	EC
0%	-	-	-	1.00
5%	0.13	1.03	1.49	0.90
10%	0.13	1.03	2.10	0.86
30%	0.29	2.22	3.05	0.78
50%	0.48	3.77	4.39	0.64
70%	0.59	4.50	5.27	0.57

1) Ⅲ장 3절 "평가지표" 참고

2. 국내·외 결측 보정처리 현황

1) 국내

한국도로공사 FTMS를 제외한 한국도로공사 우회국도 ITS, 서울시 교통관리 시스템, 천안/논산 교통관리 시스템에서는 차로단위 수집자료를 지점단위로 집계하는 과정에서 결측보정 처리를 수행하고 있다. 주로 수행하는 방법은 연속된 지점의 공간적 추세를 활용하는 방법과 이력자료를 바탕으로 해당 결측된 지점을 보정처리하는 방법을 주로 적용하고 있다.

2) 국외

미국의 Utah ATMS 및 Virginia ADMS, California PeMS 등 에서는 인접검지기를 참조하는 공간적 추세활용 및 이력자료와 통계적 처리를 통한 결측보정 처리 방식을 적용하고 있다[9),14)]. 국내에 비해 눈여겨 볼만한 점은 결측 발생시 교통량, 점유율, 속도에 대해 각각의 데이터 프로파일을 형성하여 이를 별도로 관리하고 있다는 점이다. 결측 발생에 대한 패턴 자료를 구축하며 자정(子正)이 지난 이후 하루 동안 수집된 전날 자료에 대해 EM (Expectation Maximization) 알고리듬과 같은 별도의 심화된 처리과정을 수행하여 이력데이터의 기반자료를 별도로 관리한다는 점이다. 즉 실시간으로 결측 보정과정을 수행하고, 자정이후 on-line 상의 실시간으로 보정처리 된 자료와는 별개로 off-line 상으로 보정처리를 수행하며 1일치 데이터 내의 결측내역에 대한 기록을 별도로 관리하고 있다.

3. 선행연구 고찰

1) 최근 연구동향

수집된 자료는 완전한 구성을 갖추어야 하나 일부 결측(missing, non-responsive) 문제를 포함하고 있다. 결측자료 보정처리 문제는 사회과학을 비롯한 다양한 분야에서 설문 및 조사 분석에 따른 개체 및 항목 무응답에 대한 해결방안에서 시작되었다. 자료수집 전반에 있어 발생하는 결측자료 문제는 교통 분야에서도 주요 이슈로 거론되고 있으며, 최근 국외의 일부 연구에서 진행되고 있다.

AASHTO(American Association of State Highway and Transportation Officials)의 "AASHTO Guidelines for Traffic Data Programs"(1992)에서는 수집된 교통자료내의 결측에 대한 보정처리를 권고사항에서 제외하고 있다. 이는 수정되거나 보정처리로 인한 오차는 정량화 될 수 없음을 전제로 하고 있기 때문이다[16)].

2000년을 전후로 하여 결측 보정처리에 있어 Little and Rubin(1998)[13)] 등은 EM(Expectation Maximization) 과 같은 심화된 통계학 적인 방법론을 제기하였다. 이후 ITS를 비롯하여 교통자료 저장체계의 중요성이 대두됨에 따라, 결측자료를 고려하지 않는 가공처리문제에 대해 반하는 의견을 제시하는 연구가 진행되기 시작 하였다. 미국 TTI (Texas Transportation Institute)(2004)[8)]를 비롯하여 Virgina ADMS(2005)[15)], California PeMs, UTAH DOT[9)]와 같은 연구기관과 여러 연구논문[4),5),6),7),10),11)]에서 데이터 수집체계내의 결측자료 문제와 보정처리의 방안 및 필요성이 제시되었다.

2) 기존문헌 고찰

차량검지자료내의 결측자료에 대한 보정처리에 관한 선행연구는 다음과 같다. James H. Conklin 등은 두 편의 논문에서(2003) 차량 검지자료 내 결측에 대한 처리를 이동평균 및 과거자료 대체를 비롯하여 EM과 같은 통계적 보정처리를 적용한 결측보정 결과를 제시하였다. 보정처리 과정을 통해 처리방법에 대한 수행속도 및 처리결과에 대한 추정력을 제시하였다[4),7)]. 또한 과거 차로 이용률 정보와 적용 가능한 인접 지점의 검지자료를 통해 보정 처리를 수행하였다[5)]. Chao Chen 등(2003)은 전후 지점간의 통행패턴의 연속성을 전제로 하는 선형회귀분석을 통해 지점 값을 추정하였다[6)]. Daiheng Ni 등(2005)은 차량검지자료내의 결측내역에 대한 다양한 보정처리 방법론을 소개하고 구분하였다[10)]. 그리고 AASHTO의 "Guidelines for Traffic Data Programs" (1992)[16)]에서 수집자료 내의 결측에 대해 "결측 발생량은 정량화 될 수 없으므로 보정처리(imputation)는 권고사항에서 제외한다."라는 입장에 대하여 보정처리의 신뢰성과 필요성에 대한 이슈와 근거를 연구결과를 통해 제시하였다[4),10)].

Ⅲ. 방법론 정립

1. 기본 전제

본 연구에서 수행하는 방법론은 다음과 같은 내용을

전제조건으로 하고 있다. 또한 돌발 상황 및 천재지변이 발생하지 않은 정상적인 상황아래 수집된 자료에 대해 보정 처리과정을 수행하였다.

보정처리 방식은 실시간(on-line)으로 이루어진다는 점에서, 사전에 결측발생 비율을 예측하기 어려운 실제 상황과 유사하게끔 구성되었으므로, 결측 빈도수에 따른 보정처리 여부와 off-line부문은 추후 진행과제에 포함하기로 하였다.

1) 보정처리 대상자료 및 방식

자료의 단위는 30초 주기로 24시간 동안 수집된 차로별 교통량, 속도, 점유율 자료이며, 유출입이 존재하지 않는 폐쇄구간의 연속된 지점의 검지기 자료를 대상으로 수행한다. 2006/04/01~04/30 서해안 고속도로 서해대교~서평택 IC간 상행방향 연속된 4개 지점을 대상으로 조사한 결과 결측발생은 〈표 6〉과 같이 지점단위 전 차로에서 일괄적으로 발생하는 경우가 대부분으로 나타났다. 그러나 루프 상태에 따라 차로단위를 포함하는 경우도 존재하므로 지점이 아닌 차로단위로 수행하였다.

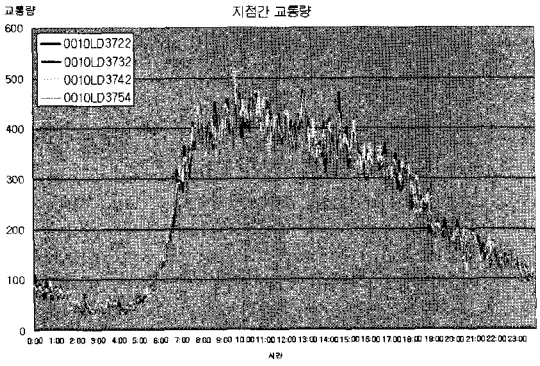
〈표 6〉 지점 단위 및 차로단위 결측 발생 예시

구분	결측발생 일 수/조사기간	지점단위	차로단위
0150LD2826	30일/30일	26건	4건
0150LD2835	30일/30일	12건	18건
0150LD2845	30일/30일	23건	7건
0150LD2858	26일/30일	29건	1건

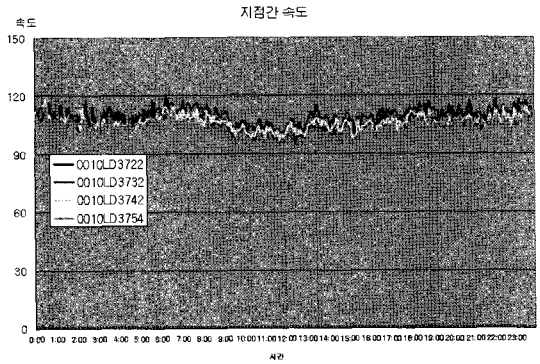
2) 연속된 지점간의 교통패턴

결측발생시 인접지점의 결측이 발생하지 않은 자료를 참조하게 되는 전제조건은, 유출입이 존재하지 않은 폐쇄구간내의 교통량은 일정하며 기본 수집주기인 30초 동안 전후 지점 간 교통량과 속도의 패턴이 유사하다 것이다. 〈그림 1〉~〈그림 2〉는 경부선 안성 분기점↔오산 IC간 2006/04/01~04/30일 동안 수집된 자료 중에서 결측 발생이 없는 연속된 지점 4개소(0010LD3722, 3732, 3742, 3754)의 4/4(화)의 결과를 나타낸 것이다. 지점간 거리는 1~1.2km이며 차로당 30초 자료를 정보가공 최소단위인 5분주기의 지점단위 자료로 가공한 후 상행선 방향을 비교한 것이다.

〈그림 1〉~〈그림 2〉를 통해 결측이 발생하지 않은 4/4(화)의 폐쇄구간 내 연속된 4개 지점간의 24시간



〈그림 1〉 연속된 4개 지점의 5분 단위 지점 교통량



〈그림 2〉 연속된 4개 지점의 5분 단위 지점 속도

동안 시간대별 통행패턴이 거의 유사함을 알 수 있으며, 이를 바탕으로 지점참조를 적용하였다.

3) 이력자료 구축

분석에 적용되는 이력자료는 2006/04/01~04/30일 동안 수집된 자료 중 동일요일에 대해 별도로 구축하여 적용하였다. 이력자료는 30초 주기로 차로단위로 구축하였으며, 앞선 3주간의 요일별·시간별·차로별 자료의 평균값을 적용하였다. 차량검지 자료내의 결측처리에 이력자료를 사용하는 선행연구에서는 대푯값으로 과거 데이터의 평균값을 적용하였으며 과거 데이터량의 범위를 7일~30일로 제시하였다. 김동호(2006)의 연구에 의하면 교통패턴을 예측함에 있어 60일의 데이터량이 적합하다고 제시하였다[2].

2. 적용 방법론

결측 보정처리 방법론은 연속된 인접지점의 수집자료를 활용하는 방식과 이력자료를 기반으로 하는 보

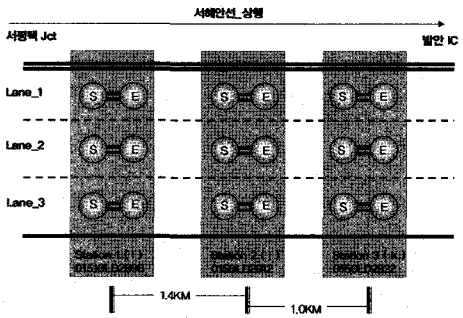
정처리로 크게 나눌 수 있다. 기존 방법으로 전후지점 동일주기 이동평균법, 이전지점 동일주기, 이력자료 동일주기 보정처리 방식이 있으며, 본 연구는 지점간 이동소요 주기를 고려하는 방법 외에 지점간 연속성을 고려한 회귀분석 및 교통량에 대하여 차로별 이용률을 적용한 보정처리방법을 제안하였다.

〈표 25〉 보정처리 방법론 구분

구분	방법론	적용범위	비고
인접지점 자료활용	전후지점 동일주기	on-ine /off-line	기존
	이전지점 동일주기	on-ine /off-line	기존
	지점간 소요주기	on-ine /off-line	제안
이력자료 활용	이력자료 동일주기	on-ine /off-line	기존
	회귀분석	on-ine /off-line	제안
	차로이용률	on-ine /off-line	제안

1) 인접지점 차량검지자료 활용기반 보정처리

유출입이 없는 폐쇄구간내의 연속된 지점상의 교통패턴이 유사하다는 전제아래, 정상류 상태에서 결측 발생시 인접 지점의 자료를 참조하는 방법으로, 본 연구에서의 해당 지점은 〈그림 3〉과 같이 3차로의 dual-loop체계로 구성되어있다.



〈그림 3〉 대상 지점 3개 차로, Dual-loop 체계

(1) 전후지점 동일주기 이동평균 적용

결측발생 지점 전후의 동일주기의 수집자료의 평균값으로 결측자료를 보정처리하는 방법으로 다음과 같다.

$$\widehat{V_{j,l,t,d}} = \frac{V_{i,l,t,d} + V_{k,l,t,d}}{2}$$

$$\widehat{S_{j,l,t,d}} = \frac{S_{i,l,t,d} + S_{k,l,t,d}}{2}$$

$$\widehat{O_{j,l,t,d}} = \frac{O_{i,l,t,d} + O_{k,l,t,d}}{2}$$

$\widehat{V_{j,l,t,d}}$: 지점j , 차로l, 시간t, 날짜d의 교통량 추정치

$\widehat{S_{j,l,t,d}}$: 지점j , 차로l, 시간t, 날짜d의 속도 추정치

$\widehat{O_{j,l,t,d}}$: 지점j , 차로l, 시간t, 날짜d의 점유율 추정치

$V_{i,l,t,d}$: 지점 i , 차로l, 시간t, 날짜d의 교통량 관측치

$S_{i,l,t,d}$: 지점i , 차로l, 시간t, 날짜d의 속도 관측치

$O_{i,l,t,d}$: 지점i , 차로l, 시간t, 날짜d의 점유율 관측치

(2) 이전지점 동일주기 적용

결측이 발생한 지점의 상류부의 인접한 첫번째 지점 자료의 동일주기로 보정처리 하는 방법이다.

$$\widehat{V_{j,l,t,d}} = V_{i,l,t,d}$$

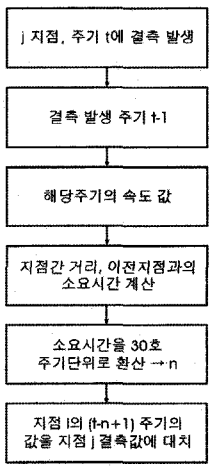
$$\widehat{S_{j,l,t,d}} = S_{i,l,t,d}$$

$$\widehat{O_{j,l,t,d}} = O_{i,l,t,d}$$

(3) 지점간 이동 소요주기 적용

연속된 지점간의 통행패턴이 유사하다는 전제아래 속도자료와 지점간의 거리를 바탕으로 이전 상류 지점에서의 해당 주기를 찾아 결측자료를 보정처리 하는 방법으로 다음과 같다.

고속도로의 경우 주행속도가 높은 관계로 대부분 한 주기 다음에 하류부 지점에 도착하지만, 저속차량의 경우 1개주기 이상이 소요되므로 보정값 추정에 있어 동일 주기 보정처리에 비해 효율적이라 할 수 있다. 조사대상구간의 1개 주기 이상이 소요되는 80km/h 이하의 경우 1, 2차로에 비해 3차로에서 하루 평균 약 10% 가량이 발생하는 것으로 나타났다.



$$\widehat{V_{j,l,t,d}} = V_{i,l,(t-n+1),d}$$
$$\widehat{S_{j,l,t,d}} = S_{i,l,(t-n+1),d}$$
$$\widehat{O_{j,l,t,d}} = O_{i,l,(t-n+1),d}$$
$$n = \frac{L}{\left[V_{j,l,(t-1),d} \times \frac{1000}{3600}\right] \times \frac{1}{timeinterval}}$$

$n \geq 0$

L : 지점간 거리(km)

n : 소요시간, 수집주기 단위로 환산

timeinterval : 수집주기, 30초

2) 이력자료 활용기반 보정처리

이력자료는 인접지점이 모두 결측자료를 갖고 있어 보정 처리시 참조가 불가능 할 경우, 사전에 구축된 이력자료를 바탕으로 하여 처리한다.

(1) 이력자료 동일주기 적용

결측이 발생한 지점의 동일 주기자료를 이력자료내 동일 주기자료로 보정처리 하는 방법이다.

$$\widehat{V_{j,l,t,d}} = V_{histj,l,t,d'}$$
$$\widehat{S_{j,l,t,d}} = S_{histj,l,t,d'}$$
$$\widehat{O_{j,l,t,d}} = O_{histj,l,t,d'}$$

$V_{histj,l,t,d'}$: 지점j , 차로l, 시간t, 요일d′, 교통량 이력자료

$S_{histj,l,t,d'}$: 지점j , 차로l, 시간t, 요일d′, 속도 이력자료

$O_{histj,l,t,d'}$: 지점j , 차로l, 시간t, 요일d′, 점유율 이력자료

(2) 회귀분석 적용

상류부 지점과 하류부 지점간의 연속성과 상관관계를 전제로 하여, 상류부지점을 독립변수 로 하류부 지점을 종속변수로 하는 회귀식을 통해 얻어진 추정치로 보정처리 하는 방법이다. 서평택 분기점~발안 IC간 2006/04/01~04/30일 동안 수집된 자료 중 해당 연속지점의 매주 화요일을 대상으로, 차로단위 교통량/속도/점유율을 3시간단위 지점 간 상관분석을 시행하였다. 분석결과 지점 간 상관관계는 작게는 0.622 크게는 0.914로 도출되었다. 교통량의 경우 상관계수가 크게 나타나며, 점유율, 속도 순으로 나타났다.

$$\widehat{V_{j,l,t,d}} = a_{l,v}(i,j) + \beta_{l,v}(i,j) \times V_{i,l,t,d}$$
$$\widehat{S_{j,l,t,d}} = a_{l,s}(i,j) + \beta_{l,s}(i,j) \times S_{i,l,t,d}$$
$$\widehat{O_{j,l,t,d}} = a_{l,o}(i,j) + \beta_{l,o}(i,j) \times O_{i,l,t,d}$$

a와 β 회귀계수는 3주간의 이력자료내의 3시간 단위로 분석한 지점 i와 j간 차로단위 교통량/속도/점유율을 각각 회귀분석을 수행하여 도출하였으며, 방법론 적용시 결측이 발생한 지점의 상류지점을 독립변수로 적용하여 결측자료에 대한 보정처리를 수행하였다.

(3) 차로별 이용률 적용

교통량 추정에 있어 전후지점 참조가 불가능한 경우, 교통량을 추정함에 있어 결측이 발생한 주기의 해당지점의 이전 주기들의 차로이용율과 이력자료 내의 지점단위 집락 된 교통량을 활용하여 차로단위 교통량을 추정하는 방법이다. 이력자료내의 차로이용률이 아닌 현 결측발생 시점 이전 10주기의 차로이용률의 평균값을 반영한다는 점에서 차로 및 지점단위 교통량 결측발생시 효율적이라 할 수 있다.

$$\widehat{V_{j,l,t,d}} = V_{hist,j,all-lane,t,d} \times Lu_{j,l}$$

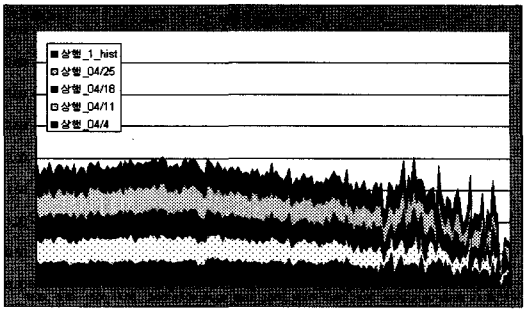
1차로의 경우, 차로이용률은 다음과 같다.

$$Lu_{j,l1} = \frac{1}{n}\sum_{k=1}^{n} \frac{V_{j,l1,t-k,d}}{V_{j,l1,t-k,d} + V_{j,l2,t-k,d} + V_{j,l3,t-k,d}}$$

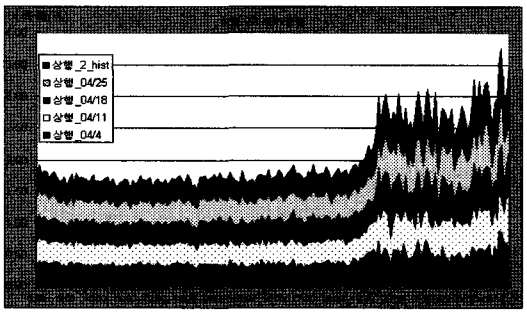
$V_{histj,all-lane,t,d}$: 지점 j, 지점단위자료, 시간 t, 요일d′,교통량이력자료

$Lu_{j,l1}$: 지점 j, 1차로이용률, n : 참조주기 수

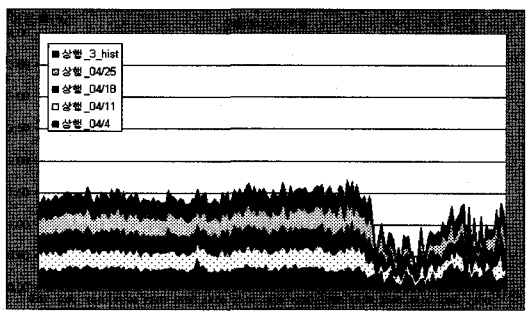
〈그림 4〉~〈그림 6〉은 2006년 4월 매주 화요일에 수집된 교통량 자료를 5분 단위 차로별 자료로 가공 후 매주 화요일(4/4, 11, 18, 25)과 4월 화요일분의 이력자료와의 차로별 이용률을 비교한 것이다. 이를 통



〈그림 4〉 00150LD2912(stationII), 1차로 이용률



〈그림 5〉 00150LD2912(stationII), 2차로 이용률



<그림 6> 00150LD2912(stationII), 3차로 이용률

해 각 차로별 이용률의 시간대별 패턴이 유사하게 진행됨을 알 수 있다.

전반적으로 1차로의 이용률(35~40%)이 가장 높았으며 2차로, 3차로 순으로 나타났다. 19시 이후 2차로의 이용률 증가와 같은 패턴의 변화는 야간 주행시 측방 여유 공간을 두고 주행함에 따른 영향으로 추정할 수 있다.

3. 평가지표

보정처리 결과의 추정력을 평가하기 위해 도출된 추정치만을 대상으로 평가지표를 적용한다. 평가시 참값을 기준으로 오차 및 유의성을 검증하기 위해 다음과 같은 지표를 적용하였다. 여러 방법론 중 최적의 대안을 선택할 경우 잔차에 의한 모형 선택뿐만 아니라 예측오차의 최소화역시 고려하는 것이 바람직하다. 이들 평가지표를 간단히 설명하면 MAE와 RMSE는 값이 작을수록, EC는 값이 클수록 모형의 추정력이 좋다고 볼 수 있다

1) MARE (Mean Absolute Relative Error)

평균절대상대오차(MARE)는 잔차의 참값에 대한 평균적인 오차비율을 의미한다. 즉 도출된 추정치가 참값에 비해 평균적으로 어느 정도의 비율로 차이가 나는지를 설명해 주는 지표이다.

$$MARE = \frac{1}{N}\sum_{i=1}^{n}\frac{|x(i)-\hat{x}(i)|}{x(i)}$$

$x(i)$: 참값, $\hat{x}(i)$: 추정치
N: 대상자료 개수

2) MAE (Mean Absolute Error)

절대평균오차는 추정값과 관측값 차이의 절대값의 평균을 의미한다. MAE는 0에 가까울수록 좋은 추정결과임을 의미한다.

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|x(i)-\hat{x}(i)|$$

3) RMSE (Root Mean Square Error)

평균제곱근오차는 관측값과 추정값의 상호간의 편차를 나타내는 지표로 추정치와 관측치가 평균적으로 어느 정도 떨어져 있는지를 나타내는 지표이다.

$$SE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|x(i)-\hat{x}(i)|^2}$$

4) EC (Equality Coefficient)

EC는 타일(Theil, 1966)에 의해서 개발된 방법으로 계량경제모형에서 구한 예측값의 정확성에 대한 척도이다. U는 타일의 부등계수로서 U가 1일 경우는 자

료가 완전히 불일치하는 것을 의미하며, U가 0일 경우는 완전히 일치 하는 것을 의미한다. 따라서 EC는 1에 가까울수록 추정력이 좋은 것을 의미한다. U는 U_m(편이비율,bias), U_s(분산비율,variance), U_c(공분산비율)로 구성되며 셋의 합은 1이 된다.

$$EC = 1 - U$$

$$U = \frac{\sqrt{\sum_{i=1}^{n}|x(i)-\hat{x}(i)|^2}}{\sqrt{\sum_{i=1}^{n}x(i)^2}-\sqrt{\sum_{i=1}^{n}\hat{x}(i)^2}}$$

5) Paired T-test

관측치와 추정치 간에 평균의 차이를 분석할 때 적용하는 방법으로, 설정된 신뢰수준에서 두 집단간 시행전후 차이를 분석하는 방법이다. 두 집단에 차이가 없다는 것을 귀무가설로 설정하고, 분석시 양측검정기준 p값이 일정수준 이상으로 도출 될 경우, 귀무가설을 지지하는 것으로 추정치가 관측치에 가깝게 도출됨을 의미한다. 본 연구에 적용된 데이터를 대상으로 확률분포 특성을 "Kolmogorov-Smirnov"검정을 수행하여 검토한 결과 오전/오후 모두 근사유의확률이 유의수준(α =0.05)보다 크게 나타났으므로(α=0.67~0.45) 정규분포임이 확인 되었다.

Ⅳ. 결측 보정처리 적용

1. 적용 사례지점 선정

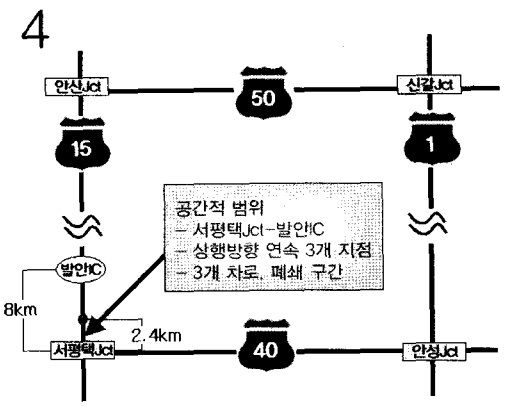
결측자료에 대해 보정처리를 수행하기 위한 대상 지점은 다음과 같다.

1) 공간적 범위

서평택 분기점~발안 IC간 약 8Km구간의 상행방향 연속된 3개 지점, 대상구간 약 2.4km의 지점별 3개 차로를 대상으로 보정처리를 수행하였으며 〈그림 7〉과 같다.

2) 시간적 범위

2006/04/01~2006/04/30 총 30일, 1일 24시간 동안 수집된 30초 주기의 차로별 자료이다. 1일 차로 단위로 2,880개의 수집영역이 이에 해당된다.



〈그림 7〉 사례지점 공간적 범위

2. 자료 구축

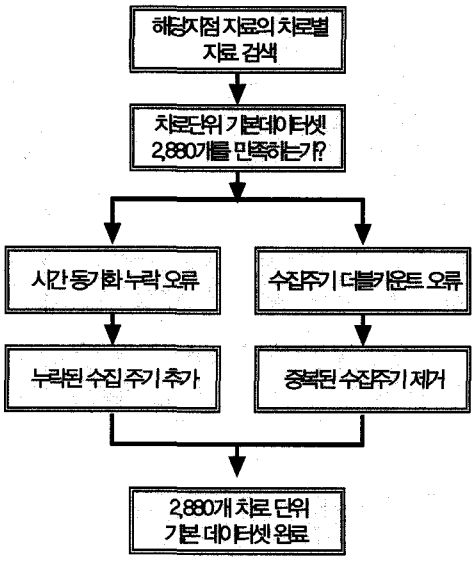
결측 보정처리를 위한 자료는 크게 다음과 같이 구분된다.

- **이력자료(historical data)**
  : 조사기간 동안 매주 화요일(4/4, 11, 18일)에 대한자료의 30초주기 차로단위, 5분 단위 지점단위 평균값 적용
- **결측 대상 지점자료**
  : 〈그림 3〉의 지점 0150LD2912(stationⅡ)의 4/25일(화)의 08:00:00~24:00:00의 30초주기 차로단위 자료
- **결측 대상 지점 전후 지점자료(adjacent data)**
  : 〈그림 3〉의 지점 0150LD2898(stationⅠ)(상류), 0150LD2912(stationⅢ)(하류)의 4/25일(화)의 00:00:00~24:00:00의 30초주기 차로단위자료

구축된 자료는 일정량의 기본 데이터셋을 갖추어야 하며 결측자료를 포함하지 않아야 하므로 〈그림 8〉과 같은 과정을 거치게 된다.

기본 데이터셋 내의 결측자료가 존재하지 않은 시간대인 08:00:00~24:00:00를 대상자료 시간적 범위로 선정하였다.

결측자료가 존재하지 않는 04/25(화)의 지점 0150LD2912 (stationⅡ)의 상행 3개 차로에 대해, 차로별 전체 수집량의 30%에 해당되는 데이터를 결측자료로 수정하여 적용하였다. 임의 결측처리는 시간대별 무작위로 시행하였으며, 실제 상황에서 차로단위 지

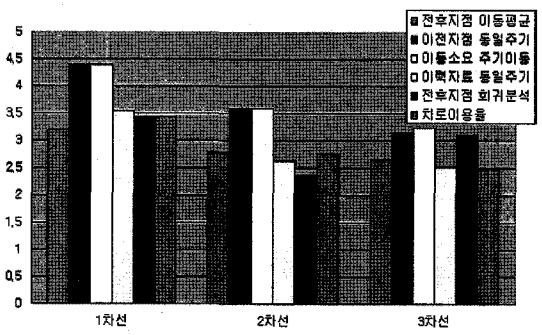


〈그림 8〉 자료구축 기본단위 처리과정

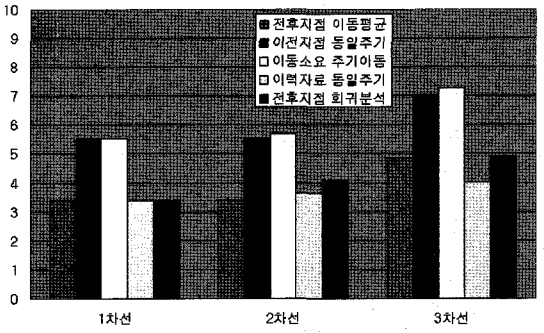
점단위로 다른 관측요인에 상관없이 무작위로 발생하는 MCAR 상황을 구현하기 위해 차로단위로 수행하였다.

3. 방법론별 보정처리 결과 분석

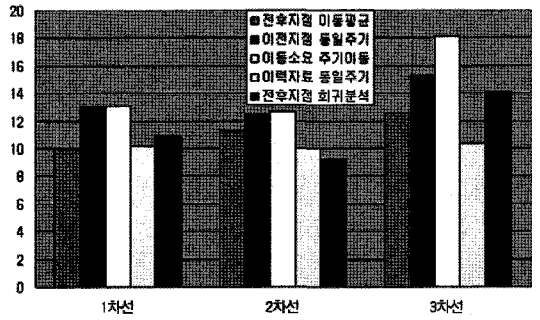
보정처리과정은 〈그림 1〉을 바탕으로 하여 1일 교통량 패턴에 따라 크게 3부문으로 나누어 각 방법론을 적용하였다. 첫째는 08:00:00~24:00:00의 전체 수집시간 동안이며, 둘째는 1일 중 가장 많은 교통량의 패턴을 나타내는 오전 시간대를 포함하는 08:00:00~10:00:00, 셋째는 한산한 저녁 시간대를 포함하는 18:00:00~20:00:00로 구분하여 수행하였다. 보정처리 결과는 〈그림 9〉~〈그림 11〉과 같다.



〈그림 9〉 지점 0150LD2912(station II)



〈그림 10〉 지점 0150LD2912(station II)
차로별 점유율 RMSE



〈그림 11〉 지점 0150LD2912(station II)
차로별 속도 RMSE

4. 분석결과 및 시사점

〈표 8〉~〈표 10〉 및 〈그림 9〉~〈그림 11〉을 통해 결측값에 대해 각 방법론별 보정처리를 거친 추정치와 관측치 간의 오차를 분석한 결과 다음을 알 수 있다. MAE와 RMSE를 통해 전반적으로 교통량은 속도와 점유율에 비해 상대적으로 참값에 가깝게 추정이 되었다. EC의 경우 속도, 교통량, 점유율 순으로 도출되었다. MAE 및 RMSE의 수치가 측정단위에 따라 변하는 점을 고려할 때 EC를 통해 이를 보완 및 참고할 수 있다. 이를 바탕으로 결측자료에 대한 보정처리를 통해 교통량과 속도 면에서는 좋은 추정치를 도출할 수 있음을 알 수 있다. 또한 인접지점을 참조하는 것에 반해, 이력자료 기반의 회귀식을 비롯하여 해당지점의 전 일정 주기의 차로이용률 등을 적용한 방법론의 오차가 상대적으로 적게 도출되었다. 이를 바탕으로 지점 간 이동시 차로변경 등과 같은 패턴의 변화가 지점간의 공간적 연속성에 미치는 영향이 적지 않음을 알 수 있었다. 예로 〈그림 9~11〉을 통해 지점 간 소요시간을 주기로

〈표 8〉 지점 0150LD2912(station II) 방법론별 보정처리결과 - 08:00:00~24:00:00

척도	방법론	1차로			2차로			3차로		
		교통량	속도	점유율	교통량	속도	점유율	교통량	속도	점유율
MARE	전후지점 이동평균	0.44	0.07	0.69	0.38	0.08	0.60	0.56	0.10	0.90
	이전지점 동일주기	0.39	0.10	2.53	0.39	0.17	1.61	0.49	0.23	2.26
	이동소요 주기이동	0.39	0.10	2.53	0.39	0.18	1.62	0.50	0.31	2.27
	이력자료 동일주기	0.44	0.07	0.49	0.42	0.08	0.41	0.36	0.09	0.43
	전후지점 회귀분석	0.48	0.08	0.51	0.44	0.07	0.57	0.71	0.11	0.77
	차로이용율	0.36	-	-	0.40	-	-	0.35	-	-
MAE	전후지점 이동평균	2.44	7.52	2.40	2.16	8.57	2.40	2.02	9.44	3.51
	이전지점 동일주기	3.28	10.21	4.07	2.75	9.05	4.07	2.40	10.89	5.14
	이동소요 주기이동	3.28	10.21	4.07	2.74	9.17	3.97	2.48	11.89	5.29
	이력자료 동일주기	2.54	7.76	2.35	1.93	7.47	2.46	1.79	7.73	2.83
	전후지점 회귀분석	2.51	8.84	2.70	1.84	6.95	3.00	2.35	10.51	3.90
	차로이용율	2.47	-	-	2.02	-	-	1.73	-	-
RMSE	전후지점 이동평균	3.19	9.98	3.42	2.78	11.29	3.42	2.65	12.54	4.87
	이전지점 동일주기	4.40	13.05	5.51	3.59	12.48	5.51	3.15	15.30	7.04
	이동소요 주기이동	4.40	13.05	5.51	3.58	12.67	5.69	3.24	18.11	7.25
	이력자료 동일주기	3.55	10.16	3.40	2.64	10.01	3.65	2.53	10.36	4.04
	전후지점 회귀분석	3.43	10.90	3.40	2.38	9.21	4.09	3.11	14.09	4.94
	차로이용율	3.44	-	-	2.76	-	-	2.49	-	-
EC	전후지점 이동평균	0.80	0.95	0.70	0.81	0.94	0.70	0.76	0.93	0.66
	이전지점 동일주기	0.74	0.94	0.40	0.76	0.94	0.40	0.72	0.92	0.48
	이동소요 주기이동	0.74	0.94	0.40	0.76	0.94	0.53	0.71	0.90	0.46
	이력자료 동일주기	0.77	0.95	0.73	0.80	0.95	0.76	0.78	0.94	0.75
	전후지점 회귀분석	0.77	0.95	0.74	0.82	0.95	0.70	0.75	0.92	0.66
	차로이용율	0.77	-	-	0.79	-	-	0.77	-	-

〈표 9〉 지점 0150LD2912(station II) 방법론별 보정처리결과 - 08:00:00~10:00:00

척도	방법론	1차로			2차로			3차로		
		교통량	속도	점유율	교통량	속도	점유율	교통량	속도	점유율
MARE	전후지점 이동평균	0.29	0.06	0.68	0.31	0.06	0.73	0.49	0.09	0.91
	이전지점 동일주기	0.37	0.11	3.86	0.31	0.08	2.27	0.44	0.12	2.53
	이동소요 주기이동	0.37	0.11	3.86	0.30	0.09	2.29	0.45	0.12	2.39
	이력자료 동일주기	0.48	0.07	0.61	0.29	0.08	0.43	0.33	0.10	0.37
	전후지점 회귀분석	0.38	0.07	0.47	0.27	0.06	0.58	0.85	0.07	0.62
	차로이용율	0.44	-	-	0.38	-	-	0.36	-	-
MAE	전후지점 이동평균	2.73	6.66	3.28	2.69	6.03	3.28	2.30	8.79	4.37
	이전지점 동일주기	4.28	10.93	6.31	3.06	7.97	6.31	2.82	10.68	7.62
	이동소요 주기이동	4.28	10.93	6.31	2.99	8.46	5.77	3.00	11.80	7.66
	이력자료 동일주기	3.41	7.90	3.31	2.27	7.08	3.58	2.06	8.51	3.11
	전후지점 회귀분석	3.10	8.13	3.01	2.20	6.01	3.73	2.94	6.45	4.28
	차로이용율	3.39	-	-	2.52	-	-	1.93	-	-
RMSE	전후지점 이동평균	3.42	8.32	4.42	3.45	7.69	4.42	2.87	10.72	5.62
	이전지점 동일주기	5.28	12.79	7.36	3.77	10.04	7.36	3.54	13.96	9.29
	이동소요 주기이동	5.28	12.79	7.36	3.72	10.91	7.44	3.74	17.52	9.31
	이력자료 동일주기	4.42	9.90	4.45	3.15	8.78	5.08	2.67	10.69	4.03
	전후지점 회귀분석	3.86	10.04	4.06	2.90	7.31	5.43	3.70	8.58	5.65
	차로이용율	4.24	-	-	3.50	-	-	2.67	-	-
EC	전후지점 이동평균	0.84	0.96	0.70	0.81	0.96	0.70	0.79	0.94	0.68
	이전지점 동일주기	0.75	0.94	0.36	0.80	0.95	0.36	0.74	0.92	0.49
	이동소요 주기이동	0.75	0.94	0.36	0.81	0.94	0.53	0.72	0.90	0.49
	이력자료 동일주기	0.77	0.96	0.72	0.82	0.96	0.75	0.81	0.94	0.80
	전후지점 회귀분석	0.80	0.96	0.74	0.83	0.96	0.70	0.76	0.95	0.69
	차로이용율	0.78	-	-	0.79	-	-	0.80	-	-

〈표 10〉 지점 0150LD2912(stationII) 방법론별 보정처리결과 - 18:00:00~20:00:00

척도	방법론	1차로			2차로			3차로		
		교통량	속도	점유율	교통량	속도	점유율	교통량	속도	점유율
MARE	전후지점 이동평균	0.80	0.07	0.83	0.52	0.10	0.64	0.69	0.09	0.98
	이전지점 동일주기	0.58	0.09	1.21	0.56	0.10	1.03	0.49	0.12	1.49
	이동소요 주기이동	0.58	0.09	1.21	0.54	0.10	1.07	0.51	0.12	1.37
	이력자료 동일주기	0.66	0.07	0.47	0.67	0.09	0.47	0.40	0.09	0.42
	전후지점 회귀분석	1.03	0.09	0.66	0.61	0.08	0.62	0.73	0.09	0.75
	차로이용율	0.40	-	-	0.50	-	-	0.34	-	-
MAE	전후지점 이동평균	2.33	7.40	1.58	2.23	10.58	1.58	1.83	8.53	2.25
	이전지점 동일주기	2.88	10.05	1.90	3.05	10.80	1.90	2.25	10.80	2.73
	이동소요 주기이동	2.88	10.05	1.90	2.88	10.43	2.93	2.43	10.10	2.88
	이력자료 동일주기	1.55	7.48	1.10	1.40	9.08	1.23	1.20	8.15	1.58
	전후지점 회귀분석	2.40	11.05	2.80	2.18	8.13	2.90	2.30	8.68	3.18
	차로이용율	1.53	-	-	1.43	-	-	1.13	-	-
RMSE	전후지점 이동평균	2.83	8.97	2.13	2.70	13.20	2.13	2.54	11.43	3.49
	이전지점 동일주기	3.72	12.26	2.92	3.73	13.03	2.92	3.14	15.25	4.22
	이동소요 주기이동	3.72	12.26	2.92	3.57	12.46	5.49	3.26	14.82	4.33
	이력자료 동일주기	2.31	8.92	1.88	1.72	11.90	1.61	2.12	10.79	3.07
	전후지점 회귀분석	3.07	13.00	3.31	2.66	10.40	3.42	2.97	10.30	3.83
	차로이용율	2.23	-	-	2.01	-	-	1.94	-	-
EC	전후지점 이동평균	0.67	0.96	0.62	0.69	0.94	0.62	0.63	0.94	0.57
	이전지점 동일주기	0.59	0.95	0.42	0.60	0.94	0.42	0.56	0.92	0.40
	이동소요 주기이동	0.59	0.95	0.42	0.60	0.94	0.42	0.55	0.92	0.40
	이력자료 동일주기	0.70	0.96	0.67	0.76	0.94	0.74	0.68	0.94	0.67
	전후지점 회귀분석	0.62	0.94	0.57	0.66	0.95	0.58	0.59	0.95	0.58
	차로이용율	0.70	-	-	0.70	-	-	0.67	-	-

〈표 11〉 지점 0150LD2912(stationII) Paired T-test(95% 유의수준, 양측검정) 결과 - 08:00:00~24:00:00

구분		전후지점 이동평균	이전지점 동일주기	이동소요 주기이동	이력자료 동일주기	전후지점 회귀분석	차로이용률
교통량	관측치평균	17.06	17.06	17.06	17.06	17.06	17.06
	추정치평균	18.02	18.24	18.12	16.68	17.26	15.25
	유의확률	0.0001	0.0004	0.0014	0.169*	0.445*	8.0789E-11
점유율	관측치평균	6.03	6.03	6.03	6.03	6.03	6.03
	추정치평균	4.57	2.05	2.54	5.74	5.65	-
	유의확률	1.221E-29	1.496E-13	9.288E-79	0.014	0.005	-
속도	관측치평균	100.17	100.17	100.17	100.17	100.17	100.17
	추정치평균	102.27	98.22	97.86	99.45	102.06	-
	유의확률	7.908E-16	3.016E-09	5.385E-11	0.007	7.7384E-11	-

주) *은 95% 신뢰수준에서 대응 표본간 차이가 없다는 귀무가설이 채택됨

환산하여 보정처리시 1, 2차로에서는 이전지점 동일주기 적용과 비슷한 결과가 나타났지만, 3차로에서는 비슷하거나 상대적으로 낮은 결과가 도출되었다. 이는 하루 중 저속차량이 차지하는 비율이 10%안팎을 차지하고 3차로에서 1, 2차로로 이동하려는 차량에 따른 영향으로 볼 수 있다. 반면 인접지점 중 상/하류부 지점을 모두 고려하는 전후지점 이동평균 방식이 상대적으로 양호한 추정치를 도출함에 따라 1개 지점이 아닌 전후지점을 참조하는 방식이 인접지점 참조시 효율적임을 알 수 있다.

이력자료 기반의 경우 동일주기와 회귀분석을 적용한 결과 〈표 11〉의 Paired T-test 결과를 통해 교통량 추정에 있어 두 집단이 유사하다는 결론을 도출 하였다. 〈그림 9~11〉과 〈표 8~10〉의 해당내역을 통해 교통량을 비롯하여 속도 점유율 추정에 있어 상대적으로 우수한 결과를 도출함을 알 수 있다. 차로이용률을 적용하는 방법의 경우, 해당시점의 지점 데이터가 이력자료의 지점데이터와 차이가 많이 날때 오차가 커지게

된다. 1일 중 오전 첨두시 08:00~10:00의 차로이용률 적용과 저녁 첨두시 18:00~20:00의 경우 전자는 MAE, RMSE가 각각 3.39, 4.24로 도출되었으며, 후자는 1.53, 2.23으로 도출되어 보정처리에 있어 기반 데이터 및 패턴에 따라 추정치의 결과에 차이가 발생함을 알 수 있다.

V. 결론 및 향후 연구과제

1. 결론

차량검지자료의 활용과 제공분야가 점차 늘어가는 추세에 따라 수집자료 내의 결측자료를 보정처리하는 연구를 수행하였다. 가장 정확한 정보는 올바른 데이터를 바탕으로 얻어지기 때문에, 자료내의 결측내역은 정보의 신뢰성과 무한한 자료 응용의 가능성에 제동을 거는 요인이 될 수 있다. 시간대를 구분하여 도출된 차로 및 수집 항목별 보정처리결과를 통해 같은 방법론이라도 차로 및 항목별 효과가 다름을 알 수 있었다. 이를 통해 동일지점 내 시간대별 차로별 특성이 보정처리시 적지 않은 요소로 작용함을 판단 할 수 있다.

전반적으로 인접지점 참조방식의 경우가 추정력이 낮게 도출되는 반면 이력데이터기반의 보정처리 결과는 상대적으로 원데이터에 가깝게 도출되었다. 이를 통해 이력데이터가 결측 시점의 교통특성을 많이 반영하고 있음을 확인 할 수 있었다. 3가지의 이력자료 활용 방안 중 이력자료 동일주기 대체와 전후지점 회귀분석에서 좋은 추정치를 얻을 수 있었으며, 이는 조사기간 전반과 교통량이 많은 오전을 통해 알 수 있었다. 그러나 교통량이 적은 저녁시간대의 경우와 3차로에 있어서 교통량 추정에서는 차로 이용률을 고려한 방법이 상대적으로 좋은 결과를 나타내는 것을 알 수 있었다.

기 적용중인 방법론과 제안된 방법론을 기 운영중인 VDS를 통해 수집된 차량 검지자료에 적용하고, 결측자료에 대한 보정처리내역의 실제 관측 자료인 원데이터 대비 추정력을 결과로 제시함으로써 보정처리과정의 신뢰성과 필요성을 제시하였다. 이후 실제 교통관리 시스템에 적용시 이를 바탕으로 통행패턴을 고려하는 보정처리과정을 구축하며, 결측발생 내역 및 처리결과에 대한 기록을 별도로 관리하여 관리·운영상의 효율성을 재고해야 할 것이다.

2. 향후 연구과제

본 연구에서는 차량 검지자료 내의 결측자료에 대해 다양한 보정처리를 수행하여 관측자료 대비 추정치의 오차분석 결과를 제시하였다. 차로별 30초 주기단위로 인접지점과 이력데이터를 기반으로 하는 보정처리 과정을 수행한 결과를 관련 지표를 통해 분석하였다.

향후 이와 같은 처리 결과를 시간대별 구분하여 보정처리 값의 패턴과 정보가공 최소단위인 5분 데이터로 가공하여 IC-IC간 구간통행정보를 산출하는 과정을 수행해야 할 것이다. 또한 다양한 결측 비율에 따라 가공되는 정보의 정확성을 분석하는 민감도 분석이 수행되어야 하며, 관측 자료에 가까운 추정치를 도출할 새로운 연구가 진행되어야 할 것이다.

본 연구에서는 실시간(on-line) 보정처리 방식을 전제함에 따라 결측 빈도수에 따른 보정처리 필요성의 유무를 깊게 고려하지 못하였으나, 이 점을 비롯하여 off-line상의 보정처리 문제를 추후 진행되는 과제로 남겨두기로 하였다.

참고문헌

1. 한국도로공사(2004), 2세대 FTMS 구축방안 연구 최종보고서.
2. 김동호(2006), 고속도로 통행시간 예측을 위한 과거 통행시간 패턴자료 구축에 관한 연구, 한양대학교 도시대학원 석사학위 논문.
3. 안광훈·백승걸·남궁성(2005), 차로이용률을 고려한 지점 교통량 자료의 집락화 방법에 관한 연구, 한국ITS학회 논문집, 제4권 제3호.
4. B.L. Smith(2003), Exploring imputation techniques for missing data in transportation management systems, TRR(1836).
5. Conklin J.H(2002), The use of local lane distribution patterns for the estimation of missing data in transportation management systems, TRB.
6. Chao Chen, Jaimyoung Kwon, Jhoen Rice, Pravin Varaiya(2003), Detecting Errors And Imputing Missing Data for Single Loop Surveillance Systems, TRB 2003 annual

Meeting.
7. James H. Conklin(2003), Data imputation strategies for Transportation Management System, Graduate Master's of Science Thesis, university of Virginia.
8. Texas Transportation Institute(2004), TRAFFIC DATA QUALITY MEASUREMENT FINAL REPORT.
9. UTAH DOT(2003), Automated Data Collection, Analysis, and Archival-FINAL REPORT.
10. Daiheng et. al(2005), Multiple Imputation scheme for overcoming the missing values and variability issues in ITS data, journal of transportation engineering.
11. Chandra, CVSR(2005), New Algorithms for filtering and imputation of real-time and archived dual-loop detector data in I-4 data warehouse, TRB.
12. Missing Data(2002), Sage Publication, Thousand Oaks, Calif.
13. B. Rubin(2002), Statistical Analysis With Missing Data, Wiley-Interscience.
14. University of Virginia(2005), TMS Performance Monitoring, Evaluation and Reporting Handbook Final Annotated Outline.
15. Linh N. Nguyen(2003), Imputation Techniques to Account for Missing Data in Support of ITS Application.
16. AASHTO(1992), AASHTO Guidelines for Traffic Data Programs, Washington, D.C.
17. 건설교통부·한국교통연구원(2003), 검지기별 구간 소통정보 산출 및 성능평가사업 1차년도 최종보고서.

♧ 주 작 성 자 : 김정연
♧ 교 신 저 자 : 김정연
♧ 논문투고일 : 2006. 7. 2
♧ 논문심사일 : 2006. 8. 14 (1차)
2006. 10. 11 (2차)
♧ 심사판정일 : 2006. 10. 11
♧ 반론접수기한 : 2007. 4. 30