

■ 論 文 ■

# 고속도로 통행시간 예측을 위한 과거 통행시간 이력자료 구축에 관한 연구 (지점 검지기를 중심으로)

A Study on the Construction of Historical Profiles for Freeway Travel Time Forecasting

**김 동 호**

(한국교통연구원 연구원)

**노 정 현**

(한양대학교 도시대학원 교수)

**박 동 주**

(서울시립대학교 교통공학과 교수)

**박 지 형**

(한국교통연구원 책임연구원)

**김 한 수**

(서울시립대학교 교통공학과 박사과정)

## 목 차

- I. 서론
  - II. 선행연구 고찰
    - 1. 통행시간 예측 방법에 관한 연구
    - 2. 과거 통행시간 이력자료 이용에 관한 연구
    - 3. 주요 쟁점
  - III. 통행시간 예측을 위한 과거 통행시간 이력자료 기준 설정 방법론
    - 1. 적정 대표값 결정 방법
    - 2. 적정 과거 데이터량 결정 방법
  - IV. 사례 분석
    - 1. 분석 구간 설정
    - 2. 적정 대표값 결과 분석
    - 3. 적정 과거 데이터량 결과 분석
  - V. 결론 및 향후 연구과제
- 참고문헌

Key Words : 통행시간 이력자료, 대표값, 과거 데이터량, 통행시간, 예측오차  
Historical Profile, Representative Value, Size Of Historical Data, Travel Time, Forecasting Error

## 요 약

고속도로에서의 지점검지체계로부터 수집·가공·처리된 과거 통행시간 이력자료를 이용한 통행시간 예측시, 사용되는 대표값과 과거 데이터량에 따라 예측의 정확성이 결정되나 이에 대한 체계적인 연구가 없는 실정이다. 따라서 본 연구의 목적은 신뢰성 있는 통행시간 예측을 위해 통행시간 이력자료의 적정 대표값과 과거 데이터량을 결정하기 위한 방법론을 제시하였다. 과거 통행시간 이력자료의 적정 대표값은 예측오차의 평균이 가장 적은 대표값을 선정할 수 있으며, 적정 과거 데이터량은 비슷한 속성을 가진 과거 통행시간 이력자료의 개별간의 차이 또는 집단간의 차이를 최소화하는 CVMSE(Cross Validated Mean Square Error)방법을 이용하여 결정할 수 있다. 한국도로공사의 고속도로 지점검지기 자료에 적용한 결과, 적정 대표값은 중앙값으로 분석되었으며, 통행시간 예측을 위한 적정 과거 데이터량은 60일로 분석되었다.

The objective of this study is to propose methods for determining optimal representative value and the optimal size of historical data for reliable travel time forecasting. We selected values with the smallest mean of forecasting errors as the optimal representative value of travel time pattern data. The optimal size of historical data used was determined using the CVMSE (Cross Validated Mean Square Error) method. According to the results of applying the methods to point vehicle detection data of Korea Highway Corporation, the optimal representative value were analyzed to be median. Second, it was analyzed that 60 days' data is the optimal size of historical data used for travel time forecasting.

## I. 서론

첨단교통정보체계(Advanced Traveler Information System: ATIS)는 교통정보이용자들에게 실시간으로 교통정보를 제공하여 최적의 경로를 선택하도록 유도하며 교통시설 이용효율을 극대화하는 중요한 역할을 하고 있다. 이러한 첨단교통정보체계가 제 역할을 수행하기 위해서는 수집된 검지기 자료를 적절히 처리하여 정확한 교통정보를 제공하는 것이 무엇보다 중요하다.

현재 고속도로상의 교통정보는 루프검지기, 영상검지기 등을 이용한 지점검지체계와 AVI, GPS, Beacon, TCS 등을 이용한 구간검지체계로부터 수집되고 있다. 지점검지체계로부터 수집·가공·처리된 과거 통행시간 이력자료(Historical Profile)<sup>1)</sup>는 교통정보 운영·관리 측면에서 그 중요성이 부각되고 있다. 예를 들면, 과거 통행시간 이력자료는 현재 제공되고 있는 통행시간 정보와 비교·평가의 대상이 되어 고속도로 운영에 대한 새로운 전략을 제공해 줄 수 있다(Turochy, 2005). 또한 오류 데이터의 결손처리와 실시간 데이터의 보정에도 중요한 자료가 된다(Chien et al. 2002; 한국도로공사, 2004).

특히 과거 통행시간 이력자료는 고속도로에서의 장기 통행시간 예측 측면에서 그 활용도가 매우 높다. 그러나 선행 통행시간 예측 연구는 대부분 실시간 데이터를 이용한 단기 통행시간 예측과 관련되어 있다(Stathopoulos and Karlaftis, 2002). 통행시간 예측시 실시간 데이터를 이용할 경우 예측대상 시간대가 현재 시각에서 멀어질수록 통행시간 예측오차는 증가한다. 반면 과거 통행시간 이력자료를 이용할 경우 통행시간 예측오차는 예측 기간에 상관없이 일정하게 나타난다(Park, 1999). 따라서 고속도로에서의 장기 통행시간 예측은 과거 통행시간 이력자료를 이용하는 것이 유리하다.

과거 통행시간 이력자료를 이용하여 통행시간을 예측할 때 사용되는 대표값과 과거 데이터량에 따라 정확성의 차이가 다르게 나타난다. 기존 통행시간 예측연구에서는 과거 통행시간 이력자료를 이용하여 통행시간 예측시 이에 대한 적정 기준을 제시하지 못하였다.

이에 본 연구에서는 고속도로에서의 장기 통행시간 예측이 필요한 경우 정확성을 높이기 위해 과거 통행시

간 이력자료의 적정 대표값과 과거 데이터량을 결정하기 위한 방법론을 제시하고 실제 사례에 적용해 보고자 한다.

## II. 선행 연구 고찰

### 1. 통행시간 예측 방법에 관한 연구

#### 1) 통행시간 예측모형의 특징

통행시간 예측 방법으로는 과거 통행시간 이력자료 접근법, 시계열 모형, 신경망 모형, 비모수 회귀분석, 동적 통행배정 모형, Traffic Simulation 모형 등이 있다. 이들 방법론의 특징은 <표 1>과 같다.

이러한 방법론을 통한 통행시간 예측은 단기 예측과 장기 예측으로 구분할 수 있으며, 예측기간에 따라 사용되는 방법도 달라진다. 단기 통행시간 예측은 현 교통류 상황과 크게 관련되므로 실시간 데이터를 이용하는 기법이 적합하다. 반면 장기 통행시간 예측은 현재 교통류를 이용하기에는 시간적 차이가 크므로 다소 무리가 있다.

<표 1> 통행시간 예측 방법의 특징 비교

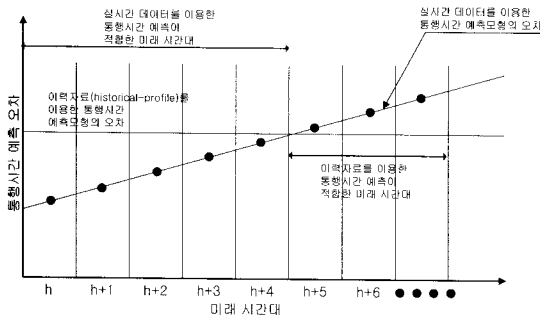
예측방법	특징
과거 통행시간 이력자료 접근법	과거 통행시간의 대표값과 교통류의 주기적 특성을 고려
시계열 모형	시간 순서대로 배열된 통계관측치의 집합으로 통행시간 예측
신경망 모형	교통류 패턴에서 현재통행시간 추정방법을 신경망의 학습능력을 통해 도출
비모수 회귀분석	현재상황과 유사한 과거사태집단을 파악하는 패턴인식기법
동적 통행배정 모형	시간에 따라 달라지는 교통수요 변화를 모형이 규정하는 시간 범위에서 수학적의 연산을 통해 도출
Traffic simulation 모형	이산시간 모형을 기초로 하여 시간을 알려진 일련의 시간 단위로 분절화한 시뮬레이션으로 도출

자료 : 박희원 외(1999)

#### 2) 과거 통행시간 이력자료를 이용한 통행시간 예측 방법의 필요성

기존 통행시간 예측모형은 크게 현재 교통류 상황을 반영한 실시간 데이터를 이용하는 경우와 과거 통행시간 이력자료를 이용하는 경우로 구분된다. 일반적으로 실시

1) 과거 통행시간 이력자료(Historical Profile)는 과거 일정기간 동안 각 링크의 시간집계간격(time interval)마다 관측된 통행시간의 대표값을 말함.



〈그림 1〉 예측기법에 따른 적정 예측 모형

간 데이터를 이용할 경우 통행시간 예측오차는 예측대상 시간대가 현재 시각에서 멀어질수록 증가한다. 이는 통행시간 예측모형의 입력변수(최근의 통행시간변화 패턴)와 출력변수(예측된 통행시간)의 상관관계가 시간이 지남에 따라 감소하기 때문이다. 반면에 과거 통행시간 이력자료를 이용한 예측모형은 예측 기간에 상관없이 오차가 일정하게 나타난다.

따라서 일정 시점 이후(예: 1시간 이후)의 장기 통행시간 예측은 실시간 데이터를 이용하기보다는 과거 통행시간 이력자료를 이용하는 것이 합리적이다(Park, 1999).

## 2. 과거 통행시간 이력자료 이용에 관한 연구

선행연구에서 과거 통행시간 이력자료는 크게 통행시간 예측, 결손처리, 교통 운영 평가에 사용되어 왔다. 특히 과거 통행시간 이력자료를 이용하여 통행시간 예측시 과거 통행시간 이력자료를 정규분포로 가정하고 평균값을 사용하였다. 일부 선행연구에서는 대표값으로 평균값을 사용할 때 과거 통행시간 이력자료로부터 많은 영향을 받아 대표값의 편이가 발생하기 때문에 최빈값을 사용하기도 하였다. 또한 과거 데이터량은 일반적으로 주 단위, 월 단위, 년 단위로 구분하여 연구마다 다르게 적용하였다. 그러나 사용된 대표값과 구축된 과거 데이터량에 따라 통행시간 예측값은 다르게 나타난다.

결론적으로 선행연구에서는 통행시간 예측시 대표값과 과거 데이터량에 대한 적정 기준을 제시하지 못하였으며 합리적인 연구나 검증 없이 이력자료를 구축하였다. 〈표 2〉는 선행연구에서 적용된 과거 통행시간 이력자료의 대표값과 과거 데이터량을 정리한 것이다.

〈표 2〉 과거 통행시간 이력자료 이용에 관한 연구

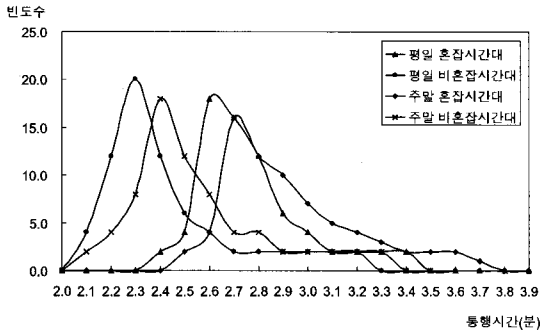
	구분	대표값	과거 데이터량
통행시간 예측	Hoffman and Janko, (1990)	평균값	제시 없음
	Shbaklo et al. (1992)	평균값	제시 없음
	Stathopoulos and Karlaftis, (2002)	평균값	3년 (요일 구분)
	Chien and Kuchipudi, (2003)	평균값	7일 (요일 구분)
	Vanajakshi, (2004)	평균값	4일 (평일만)
	이의은·김정현, (2002)	최빈값	30일 (요일 구분)
결손처리	Turner et al. (1999)	평균값	제시 없음
	한국도로공사, (2003)	평균값	30일 (요일 구분)
	Chien and Kuchipudi, (2003)	평균값	7일 (요일 구분)
	한국도로공사, (2004)	평균값	제시 없음
교통 운영 평가	Smith, (2002)	평균값	30일 (요일 구분)
	Wu, (2003)	평균값	제시 없음
	Turochy, (2005)	평균값	30일 (요일 구분)

## 3. 주요쟁점

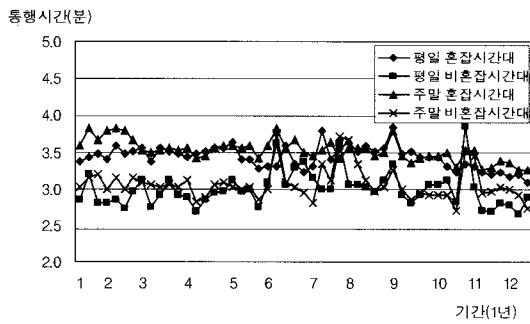
첫째, 과거 통행시간 이력자료를 이용하여 통행시간을 예측할 때 어떠한 과거 통행시간 대표값을 사용하는 것이 오차를 가장 최소화할 수 있는지 고려되어야 한다. 기존 연구에서는 일반적으로 과거 통행시간 이력자료의 분포를 정규분포로 가정하고 평균값을 사용하였다. 과거 통행시간 이력자료의 분포가 정규분포이면 대표값으로 평균값, 중앙값, 최빈값 중 어떠한 것을 사용하더라도 통행시간 예측시 오차는 동일하게 나타난다. 그러나 〈그림 2〉와 같이 과거 통행시간 분포는 오른쪽 꼬리분포를 나타내고 있다. 즉 각각의 대표값으로 평균값, 중앙값, 최빈값을 제공하였을 경우 예측오차는 다르게 나타난다.

또한 과거 몇 일의 데이터를 이용하여 이력자료를 구축하는 것이 통행시간 예측 오차를 최소화할 수 있는지 고려되어야 한다. 〈그림 3〉은 1년간 즉암-청원 구간의 평일/주말, 혼잡/비혼잡시간대 과거 통행시간 패턴을 나타낸 것이다. 〈그림 3〉과 같이 과거 통행시간의 데이터

2) 경부고속국도 서울~대전 구간을 30개 구간으로 구분하여 2004년(1년)의 과거 통행시간 분포를 살펴보면 일반적으로 오른쪽 꼬리분포를 나타냄



〈그림 2〉 과거 통행시간 자료의 빈도 분포



〈그림 3〉 과거 통행시간 패턴 분포

량이 많을 경우에는 안정된 통행시간 패턴을 확보할 수 있지만, 장래 통행시간 예측시에 적합하지 않은 통행시간 패턴이 포함될 수 있다. 반면에 과거 통행시간의 데이터 양이 적을 경우에는 안정된 통행시간 패턴을 포함하지 못하기 때문에 부정확한 통행시간 예측정보를 제공할 수 있다.

따라서 이를 고려하여 과거 통행시간 이력자료를 활용한 통행시간 예측시 정확도를 높이기 위해서는 적정 대표값과 과거 데이터량 선정을 위한 적정 기준이 필요하다.

### Ⅲ. 통행시간 예측을 위한 과거 통행시간 이력 자료 기준 설정 방법론

#### 1. 적정 대표값 결정 방법

##### 1) 기본개념

본 연구에서는 통행시간 예측을 위한 과거 통행시간 이력자료의 적정 대표값으로 평균값, 중앙값, 최빈값만 고려하였다. 평균값은 모든 과거 통행시간 이력자료 값

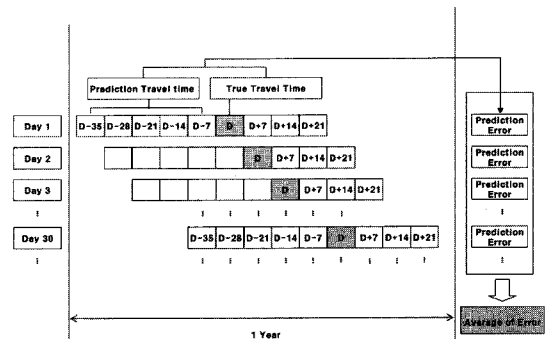
이 사용되며, 매우 크거나 작은 값에 영향을 받는다. 반면, 중앙값은 자료를 크기순으로 나열할 때 가운데 위치하는 값으로, 매우 크거나 작은 값에 영향을 받지 않는다. 또한 최빈값은 과거 통행시간 이력자료에서 도수가 가장 많은 값으로 중앙값과 마찬가지로 매우 크거나 작은 값에 영향을 받지 않는다.

이러한 과거 통행시간 이력자료의 적정 대표값은 예측오차를 근거로 결정될 수 있다. 즉 예측오차가 작으면 작을수록 정확도가 높은 대표값이다. 따라서 적정 대표값은 대표값을 이용한 예측 통행시간과 실제 관측된 통행시간과의 예측오차를 최소화하는 것을 선정할 수 있다.

#### 2) 과거 통행시간 예측오차 분석 및 평가 방법

과거 통행시간 이력자료의 적정 대표값(예측값)은 과거 일정 기간 동안 각 링크의 시간대별(예, 10분 간격) 관측된 통행시간에 대해 평균값, 중앙값, 최빈값을 산출한 것이다. 적정 대표값(예측값)을 산출하기 위해 일정기간 동안 각 시간대별 산출되는 대표값(예측값)과 실측값 차이의 예측오차를 평균(Average of Error)하여 평균오차가 최소화되는 대표값을 적정 대표값으로 결정할 수 있다.

〈그림 4〉는 과거 통행시간 이력자료의 적정 대표값을 결정하기 위한 예측오차 분석 및 평가 방법을 나타낸 것이다.



〈그림 4〉 통행시간 예측오차 분석 및 평가 방법 (예: 30일 과거 통행시간 이력자료 이용시)

#### 2. 적정 과거 데이터량 결정 방법

##### 1) 기본개념

과거 통행시간 이력자료의 적정 과거 데이터량은 대표값에 의해 예측된 통행시간과 실제 관측된 통행시간과의 차이를 최소로 하는 과거 데이터량을 의미한다. 이는

과거 통행시간 이력자료의 특성을 파악하여 비슷한 속성을 가진 데이터로 그룹화하는 개념으로 설명하는 것이 적절하다(Vanajakshi, 2004). 즉 비슷한 속성을 가진 과거 통행시간 이력자료는 과거 통행시간 이력자료의 개별간의 차이 또는 집단 간의 차이를 통해 군집될 수 있다. 이는 과거 통행시간 이력자료의 개별간의 차이 또는 집단 간의 차이를 cross-validated<sup>3)</sup> 개념을 통해 계량화될 수 있음을 의미한다(Gajewski et al. 2001). 따라서 과거 통행시간 이력자료의 적정 과거 데이터량은 cross-validated 개념을 도입하여 예측 통행시간과 실제 관측된 통행시간과의 차이를 최소화하는 과거데이터량으로 결정할 수 있다.

2) 결정 모형

CVMSE(k) : 과거 통행시간 이력자료의 대표값과 예측된 통행시간의 오차 기대값

$\hat{x}_{ij}(k)$  : k시점 이후 i번째 요일, j번째 집계간격의 예측된 통행시간(과거통행시간 대표값)

$x_{ij}(k)$  : k시점 이후 i번째 요일, j번째 집계간격의 과거 통행시간

$\bar{x}_{ij}(k)$  : i번째 일, j번째 집계간격을 제외한 예측된 통행시간

여기서, I : 일(one-day)

j : 집계간격 크기

D : 총 일수

I : 총 집계간격 수

CVMSE (Cross Validated Mean Square Error) 개념은 식(3)과 같이 나타낼 수 있다.

$$CVMSE(k) = E\left(\frac{\sum_{i=1}^p \sum_{j=1}^I (x_{ij}(k) - \hat{x}_{ij}(k))^2}{DI}\right) \quad (1)$$

식(1)을 재구성하면 식(2)와 같이 나타낼 수 있다.

$$CVMSE(k) = E\left(\frac{\sum_{i=1}^p \sum_{j=1}^I [x_{ij}(k) - \bar{x}_{ij}(k) + \bar{x}_{ij}(k) - \hat{x}_{ij}(k)]^2}{DI}\right) \quad (2)$$

식(2)의 의미를 파악하기 위해 식(3)과 같이 나타낼 수 있다.

$$CVMSE(k) = E\left(\frac{\sum_{i=1}^p \sum_{j=1}^I (x_{ij}(k) - \bar{x}_{ij}(k))^2 + (\bar{x}_{ij}(k) - \hat{x}_{ij}(k))^2}{DI}\right) + \frac{2(x_{ij}(k) - \bar{x}_{ij}(k)) \cdot (\bar{x}_{ij}(k) - \hat{x}_{ij}(k))}{DI} \quad (3)$$

식(3)에서 두 항목의 곱으로 표현된 부분은 각 항목의 기대값이 0이고 두 항목이 독립적이라 할 수 있기 때문에 무시될 수 있다(유소영 외 2인, 2004). 따라서 식(3)은 식(4)와 같이 정리된다.

$$CVMSE(k) = E\left(\frac{\sum_{i=1}^p \sum_{j=1}^I (x_{ij}(k) - \bar{x}_{ij}(k))^2 + (\bar{x}_{ij}(k) - \hat{x}_{ij}(k))^2}{DI}\right) \quad (4)$$

식(4)의 CVMSE는 크게 두 개의 항목으로 구분된다. 첫 번째 항목은 MSE로 통계학에서 Precision을 의미한다. 이는 과거 통행시간 이력자료의 데이터량이 많을수록 오차가 줄어드는 경향이 있다.

두 번째 항목은 통계학에서 Bias를 의미하며, 과거 이력자료의 데이터량에서 예측치와 i번째 일, j번째 인터벌의 개별 통행시간을 제외한 예측치와의 차이를 나타낸다. 이 Bias 부분은 과거 이력자료의 데이터량에서 대표값에 대한 신뢰도를 측정할 수 있으며, 데이터량이 많을수록 표본수가 증가하기 때문에 오차가 줄어드는 경향이 있다.

따라서 통행시간 예측을 위한 과거 통행시간 이력자료의 적정 데이터량은 과거 데이터량을 고려한 Precision과 Bias의 합을 최소로 하는 CVMSE이다. Precision은 이력자료의 데이터량이 많을 수록 작아지고, Bias는 그 반대이므로, 적정 데이터량은 Precision과 Bias의 상쇄관계에 의해 결정된다고 할 수 있다.

IV. 사례 분석

1. 분석 구간 설정

본 연구에서는 단위 링크 통행시간 산정을 위해 경부고속도로 서울~대전 구간을 30개 구간으로 구분하였다. 이 중 교통량이 많고 혼잡이 발생하여 평균 서비스 수준

3) cross-validated는 일정기간 과거 통행시간 이력자료에서 특정 과거 통행시간을 제외한 기대값을 의미한다.

이 D인(V/C비≥0.70이상) 판교-달래네고개, 서울-판교, 죽전-서울, 죽암-청원, 회덕-신탄진 5개의 구간을 분석구간으로 설정하였다.

해당 구간에 대해 2004년 1월에서 2004년 12월까지의 VDS 데이터를 요일별로 구분(한국도로공사, 2004)하여 10분 집계간격(유소영 외 2인, 2004)으로 가공·처리하였다. 이 중 추석, 설날 등의 명절과 어린이날 등의 이벤트일은 본 연구에서 제외하였다. 또한 VDS 데이터의 가공·처리 과정에서 비합리적인 데이터의 경우 통행시간 패턴을 왜곡시키므로, 한국도로공사(2003)에서 적용한 중위절대편차(MAD) 이상치 알고리즘을 적용하여 제거하였다.

위 5개 구간에서 침두시인 오전 7:00~9:30과 오후 5:00~9:00를 혼잡시간대로 설정하고 그 외 시간대는 비혼잡시간대로 설정하였다.

**2. 적정 대표값 결정 결과 분석**

**1) 예측오차 결과 분석**

통행시간 예측오차를 측정하는 방법으로는 평균오차(ME), 평균제곱오차(MSE), 평균제곱근오차(RMSE), 평균절대비율오차(MAPE) 등이 있다. 측정방법에 따라 양의 오차와 음의 오차가 상쇄되어 버리는 경우와 오차가 극단치의 영향을 많이 커지거나 작아지거나 하는 경우가 있다. 또한 오차의 제곱을 취하기 때문에 오차가 클수록 부여되는 가중치가 커지는 결과를 가져오기도 한다.

본 연구에서는 사용되는 예측오차 측정방법에 따라 다르게 나타날 수 있는 결과를 고려하여 예측오차 측정방법 중 평균절대백분율오차(MAPE: Mean Absolute Percentage Error)와 평균제곱오차(MSE: Mean Square error)를 동시에 사용하였다. 평균절대백분율오차(MAPE)와 평균제곱오차(MSE)에 대한 기본 식은 식(5)와 식(6)과 같다.

$$MAPE = \frac{\sum_{i=1}^D \sum_{j=1}^I \frac{|ts_{ij} - HP_{ij}|}{ts_{ij}}}{DI} \times 100 \tag{5}$$

$$MSE = \frac{\sum_{i=1}^D \sum_{j=1}^I (ts_{ij} - HP_{ij})^2}{DI} \tag{6}$$

여기서,  $ts_{ij}$  : i번째 요일의 j번째 집계간격의 통행

시간 관측값

$HP_{ij}$  : i번째 요일의 j번째 집계간격의 과거

통행시간 대표값

D : 일 수

I : 집계간격 수

과거 통행시간 이력자료는 요일별 통행특성이 다르기 때문에 요일별로 구분하여 예측오차 평균을 산출하였다(한국도로공사, 2004). 분석 결과를 살펴보면 혼잡/비혼잡 시간대일 경우에는 요일수별, 구간별에 따라 차이가 있지만 일반적으로 중앙값의 예측오차 평균이 가장 적게 나타났다. 주말의 경우에는 평일보다 각 대표값의 예측오차 평균이 높게 나타났다. 이는 주말의 과거 통행시간 패턴이 평일보다 패턴의 변동이 크기 때문이라 판단된다. <표 3>은 7일, 30일, 60일, 90일, 120일, 150일, 180일, 210일간의 과거 통행시간 이력자료를 요일별, 구간별, 혼잡/비혼잡 시간대로 구분하여 각 대표값의 예측오차 평균을 나타낸 것이다.

<표 3> 분석 구간 설정

구분	차로당 1시간 평균 교통량	V/C
판교-달래네고개	1,324	0.78
서울-판교	1,674	0.76
죽전-서울	1,785	0.81
죽암-청원	1,131	0.71
회덕-신탄진	986	0.72

**2) 과거 통행시간 이력자료의 분포에 따른 대표값과의 관계**

앞에서도 언급되었듯이 과거 통행시간 이력자료 분포는 skewed pattern을 나타나고 있다.

통계학에서 중앙값은 과거 통행시간 이력자료 분포의 위치에 따라 결정되므로 자료의 극한치에 영향을 전혀 받지 않게 된다. 반면에 기존 연구에서 사용된 평균값은 과거 통행시간 이력자료 하나하나로부터 영향을 받는다. 이에 과거 통행시간 이력자료 중에서 아주 작거나 큰 값이 있을 경우 이 값에 의해 영향을 받게 된다. 즉, 대표값으로서 평균값이 주어질 때 혼잡/비혼잡일 경우에서의 극단적인 통행시간(오른쪽 꼬리 부분)이 포함되기 때문에 통행시간 예측에 크게 영향을 끼칠 수 있다.

따라서 과거 통행시간 이력자료 분포가 skewed pattern 일 때 중앙값은 평균값보다 대표값으로서 더 큰 의미를 갖는다.

〈표 4〉 통행시간 예측오차 평균

요일수	대표값	월		화		수		목		금		토		일		
		MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	
혼잡 시간 대	74)	중앙값	87.2	8.3	57.4	6.5	63.4	6.6	74.0	7.6	41.6	5.7	129.8	11.3	155.2	14.4
		최빈값	104.5	9.3	63.5	7.1	70.2	7.5	90.5	8.9	47.8	5.8	141.9	11.7	212.1	17.3
		평균값	100.2	9.1	58.2	7.6	72.8	7.7	82.3	9.5	57.1	6.7	158.5	12.5	174.5	17.7
	30	중앙값	86.3	8.3	57.1	6.0	62.4	6.5	72.3	7.4	41.2	5.4	128.1	10.9	154.7	15.2
		최빈값	102.1	8.9	62.1	7.1	69.5	7.4	88.6	8.5	45.9	5.8	138.1	11.7	209.9	17.4
		평균값	91.5	8.3	52.9	6.6	65.9	6.9	74.9	7.7	50.1	5.8	143.2	10.9	163.2	15.9
	60	중앙값	79.5	7.8	53.4	6.0	53.3	5.8	68.6	7.3	38.2	5.4	95.2	8.0	135.2	14.3
		최빈값	97.2	8.9	59.9	6.8	65.3	6.7	78.6	8.0	52.5	6.3	110.3	8.9	179.3	16.4
		평균값	84.3	8.1	51.3	6.3	55.2	6.1	73.1	7.4	38.8	5.7	110.9	8.6	143.0	14.9
	90	중앙값	86.9	8.4	54.3	6.4	56.0	6.3	77.2	7.8	46.3	5.6	102.7	7.6	147.8	15.0
		최빈값	109.6	9.5	65.1	7.0	70.9	7.1	92.5	8.7	70.8	6.9	130.3	8.9	222.9	18.4
		평균값	93.0	8.7	58.1	6.6	59.1	6.3	80.7	8.0	49.0	5.8	98.3	7.9	158.8	15.7
	120	중앙값	88.9	8.6	56.6	6.2	79.0	7.5	80.4	7.9	62.3	6.9	123.0	9.8	180.2	16.2
		최빈값	111.4	9.7	70.3	7.0	97.6	8.4	95.7	8.8	79.3	7.7	167.2	11.1	256.7	18.9
		평균값	97.6	8.9	59.2	6.3	83.4	7.6	83.8	8.2	65.9	7.0	134.4	10.2	202.9	17.1
	150	중앙값	92.7	8.8	64.1	6.5	84.9	7.9	89.5	8.1	70.9	7.4	133.3	10.2	215.6	17.5
		최빈값	112.7	9.6	76.7	7.3	104.8	8.9	102.4	8.9	86.4	8.2	163.9	10.9	293.3	20.3
		평균값	97.0	8.9	65.5	6.7	89.9	8.0	92.5	8.2	77.2	7.8	141.4	10.4	260.2	18.6
	180	중앙값	101.0	9.1	73.4	7.1	89.0	8.2	94.0	8.4	108.2	9.2	133.0	10.2	227.7	17.6
		최빈값	130.7	10.6	89.9	8.0	113.1	9.4	109.6	9.2	123.1	9.1	162.6	10.9	295.1	19.8
		평균값	110.5	9.6	78.1	7.3	95.9	8.4	97.4	8.6	119.6	9.8	139.0	10.5	262.2	19.1
	210	중앙값	107.4	9.5	82.7	7.6	96.5	8.4	102.7	9.1	105.4	8.3	153.2	11.5	281.1	19.2
		최빈값	131.1	10.3	99.6	8.4	118.2	9.5	127.9	10.3	121.9	9.1	195.1	12.6	391.0	21.6
		평균값	113.0	9.6	86.8	7.7	103.2	8.6	105.4	9.3	114.7	8.6	164.3	11.9	330.9	20.3
비혼 잡시 간대	7	중앙값	57.5	5.3	50.6	4.4	46.1	4.9	54.1	5.5	46.5	5.2	114.4	10.2	145.6	12.6
		최빈값	62.3	5.6	59.2	5.9	51.3	5.8	62.0	6.1	55.5	5.7	129.7	11.6	188.3	14.7
		평균값	65.6	6.0	56.5	5.6	52.7	5.9	61.3	6.2	54.9	5.8	139.6	11.8	172.6	14.6
	30	중앙값	57.9	5.3	49.1	4.9	45.6	4.9	53.6	5.4	46.0	5.1	113.3	10.1	144.2	12.5
		최빈값	62.1	5.7	57.8	5.7	50.4	5.4	60.8	6.0	54.4	5.7	127.2	11.3	186.6	14.2
		평균값	59.4	5.5	51.5	5.2	48.0	5.2	55.7	5.6	49.9	5.4	127.8	10.7	156.9	13.1
	60	중앙값	49.8	4.9	44.9	4.9	40.5	4.6	42.7	4.8	37.0	4.5	94.6	8.4	131.5	12.1
		최빈값	57.5	5.4	54.7	5.5	49.3	5.1	49.6	5.3	47.0	5.5	113.4	9.3	174.6	14.4
		평균값	52.0	5.0	45.0	5.1	42.0	4.8	45.5	5.0	42.3	5.0	103.1	8.8	138.0	13.2
	90	중앙값	50.7	5.1	49.0	5.2	41.4	4.8	53.5	5.5	49.2	5.4	102.6	8.3	132.5	12.5
		최빈값	64.4	5.9	60.6	5.8	51.0	5.3	59.5	6.0	57.9	5.9	122.9	9.2	180.3	14.3
		평균값	53.7	5.3	48.8	5.3	42.5	5.0	55.0	5.7	51.6	5.6	105.0	8.6	144.0	13.2
	120	중앙값	52.5	5.1	51.5	5.4	47.3	5.1	57.3	5.7	53.4	5.4	125.6	11.6	157.7	13.1
		최빈값	59.4	5.5	65.6	6.1	55.9	5.6	65.6	6.1	61.3	6.0	169.9	13.3	199.1	14.9
		평균값	52.2	5.3	53.1	5.5	48.9	5.4	59.5	5.9	56.9	5.8	136.8	12.3	166.5	13.8
	150	중앙값	54.0	5.4	50.1	5.3	48.1	5.3	60.3	5.8	56.8	5.3	143.0	12.5	162.6	13.5
		최빈값	65.8	6.0	63.5	5.9	55.7	5.7	69.5	6.3	63.9	5.8	187.3	14.0	213.4	15.1
		평균값	55.4	5.5	49.9	5.5	50.1	5.4	62.6	6.1	60.2	5.5	155.6	13.2	177.4	14.3
	180	중앙값	58.8	5.6	51.5	5.2	49.3	5.2	63.0	6.0	60.4	5.6	144.2	13.0	179.1	14.6
		최빈값	66.6	6.0	56.5	5.7	53.4	5.5	73.4	6.5	66.3	6.1	190.5	14.5	199.9	16.0
		평균값	60.1	5.7	50.3	5.5	51.4	5.5	67.4	6.3	64.4	5.9	158.2	13.8	196.6	15.2
	210	중앙값	57.4	5.4	59.5	5.6	47.1	5.0	70.9	6.5	68.2	6.2	172.4	15.2	201.1	15.7
		최빈값	58.3	5.6	71.7	6.4	50.2	5.2	77.8	6.9	85.5	6.9	220.3	16.8	273.1	17.5
		평균값	56.2	5.7	59.6	5.8	49.4	5.3	73.2	6.7	73.6	6.3	193.0	16.5	236.0	16.0

4) 7일은 요일별 1개, 30일은 요일별 4~5개, 60일은 요일별 8~10개, 90일은 요일별 12~15개, 120일은 요일별 16~20개, 150일은 요일별 20~25개, 180일은 요일별 24~30개, 210일은 요일별 28~35개의 과거 통행시간 자료를 포함함

### 3. 적정 과거 데이터량 결과 분석

7일, 30일, 45일, 60일, 75일, 90일, 120일, 150일, 180일, 210일 과거 통행시간 이력자료를 제시한 모형에 적용한 후 CVMSE를 최소화하는 과거 데이터량을 결정하였다.

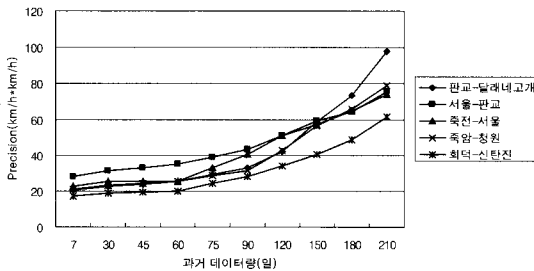
#### 1) 평일 혼잡/비혼잡 시간대의 적정 과거 데이터량 결정

〈그림 5〉와 〈그림 6〉은 평일 혼잡/비혼잡 시간대 Precision을 나타낸 것이다. 과거 데이터량이 많을수록 Precision은 증가하는 경향을 보이고 있다. 〈그림 7〉과 〈그림 8〉은 Bias를 나타낸 것이다. Bias는 대체적으로 과거 데이터량이 많을수록 감소하고 있다. 〈그림 9〉와 〈그림 10〉은 평일 혼잡/비혼잡 시간대 CVMSE를 나타

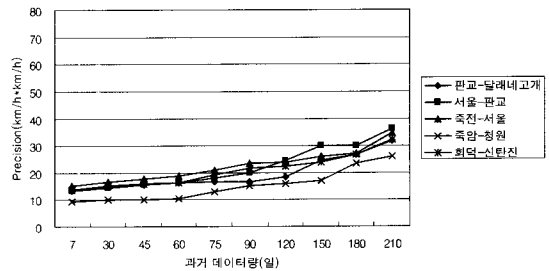
낸 것으로 Precision과 Bias의 합으로 나타난다. 혼잡/비혼잡시간대에서의 CVMSE는 60일 정도의 과거 데이터량에서 최소값을 가진다.

#### 2) 주말 혼잡·비혼잡 시간대의 적정 과거 데이터량

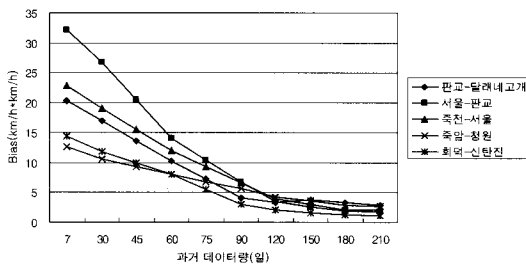
〈그림 11〉과 〈그림 12〉는 주말 혼잡/비혼잡 시간대 Precision을 나타낸 것이다. 평일과 마찬가지로 과거 데이터량이 많을수록 증가하는 경향을 보이고 있다. 〈그림 13〉과 〈그림 14〉는 Bias를 나타낸 것으로 과거 데이터량이 많을수록 0에 가까워진다. 〈그림 15〉과 〈그림 16〉은 CVMSE를 나타낸 것이다. CVMSE는 혼잡/비혼잡 시간대일 경우 구간별 차이가 존재하지만 대체적으로 60일 정도의 과거 데이터량에서 최소 예측오차를 보이고 있다.



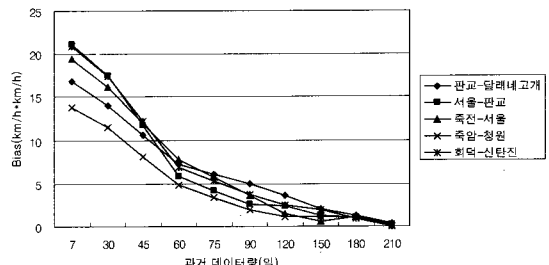
〈그림 5〉 평일 혼잡시간대 Precision



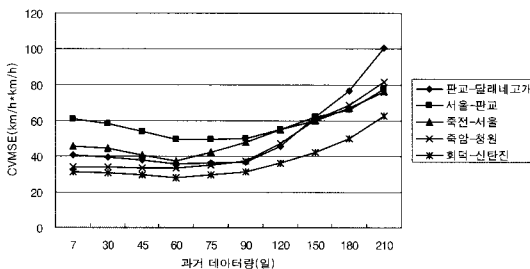
〈그림 6〉 평일 비혼잡 시간대 Precision



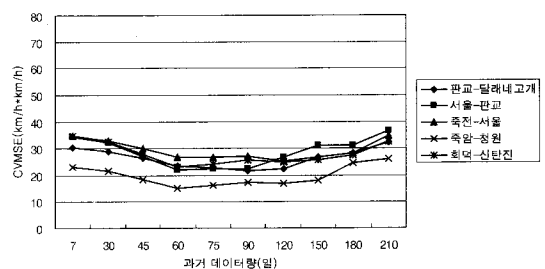
〈그림 7〉 평일 혼잡시간대 Bias



〈그림 8〉 평일 비혼잡시간대 Bias

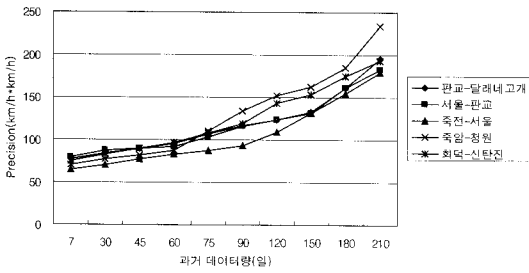


〈그림 9〉 평일 혼잡시간대 CVMSE

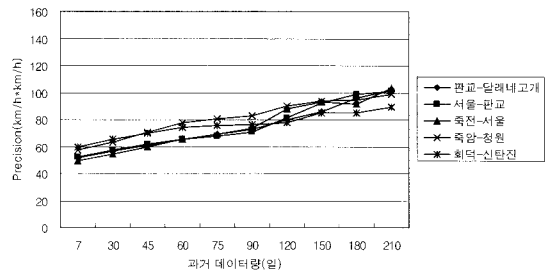


〈그림 10〉 평일 비혼잡시간대 CVMSE

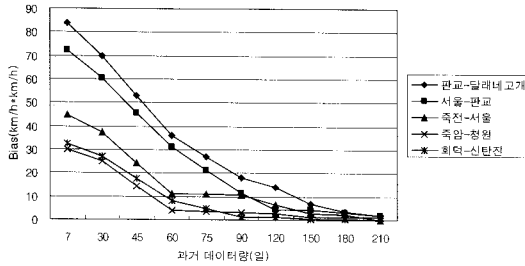




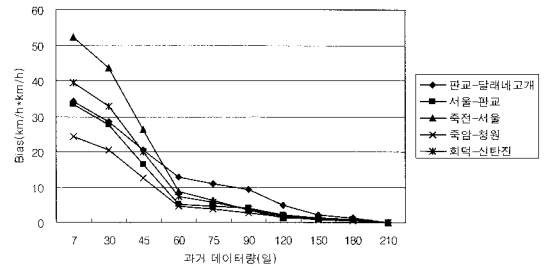
〈그림 11〉 주말 혼잡시간대 Precision



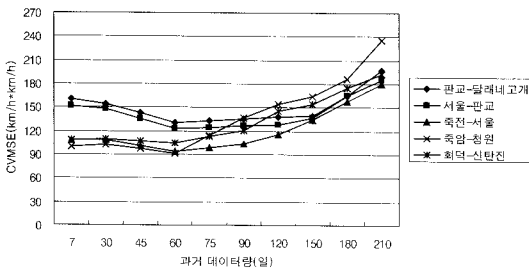
〈그림 12〉 주말 비혼잡시간대 Precision



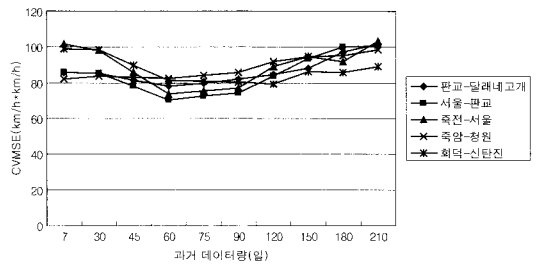
〈그림 13〉 주말 혼잡시간대 Bias



〈그림 14〉 주말 비혼잡시간대 Bias



〈그림 15〉 주말 혼잡시간대 CVMSE



〈그림 16〉 주말 비혼잡시간대 CVMSE

## V. 결론 및 향후 연구 과제

과거 통행시간 이력자료를 이용하여 통행시간을 예측할 때 사용되는 대표값과 과거 데이터량에 따라 정확성의 차이가 다르게 나타난다. 기존 통행시간 예측 연구에서는 적정 대표값과 과거 데이터량에 대한 적정 기준을 제시하지 못하였으며, 합리적인 연구나 검증없이 과거 통행시간 이력자료를 구축하였다. 따라서 본 연구에서는 통행시간 예측의 정확성을 높이기 위해 과거 통행시간 이력자료의 적정 대표값과 과거 데이터량을 결정하는 방법론을 제시하였다.

과거 통행시간 이력자료의 적정 대표값은 일정기간 동안 각 시간대별 산출되는 대표값(예측값)과 실측값 차이의 예측오차를 평균하여 평균오차가 최소화되는 대표값을 적정 대표값으로 결정할 수 있다. 또한 적정 과거

데이터량은 비슷한 속성을 가진 과거 통행시간 이력자료의 개별간의 차이 또는 집단 간의 차이를 최소화하는 CVMSE(Cross Validated Mean Square Error)방법을 이용하여 결정할 수 있다.

제시된 방법론을 고속도로 평균 서비스 수준이 D인 구간에 적용하여 적정 대표값과 과거 데이터량을 분석하였다. 적용 결과를 요약하면 다음과 같다. 첫째, 평일/주말, 혼잡/비혼잡시간대 통행시간 예측을 위한 적정 대표값은 중앙값으로, 기존 연구에서 사용되었던 평균값보다 오차가 적은 것으로 분석되었다. 둘째, 평일/주말 혼잡/비혼잡시간대 통행시간 예측을 위한 적정 과거 데이터량은 60일로 분석되었다.

본 연구에서 도출된 적정 대표값과 과거 데이터량은 고속도로 서비스 수준이 D인에 구간에 해당되는 것으로, 적용된 기간과 구간의 교통 특성에 따라 변할 수 있다.

즉 네트워크의 변화, 수요변화, 유고발생 등으로 과거 통행시간 이력자료의 대표값과 적정 데이터량은 변할 수 있다. 또한 본 연구에서는 루프검지기로부터 수집된 지점자료를 기초로 추정된 링크통행시간을 바탕으로 적정 대표값과 과거 데이터량을 결정하였다. 지검검지기인 루프검지기로부터 구간통행시간을 추정하는 과정에서는 필연적으로 오차가 발생한다 이러한 문제는 링크의 길이가 커질수록 심각해진다 본 연구의 결과는 링크통행시간의 이러한 한계를 포함하고 있다 따라서 향후에는 이러한 점을 고려하여 과거 통행시간 이력자료를 구축해야 할 필요성이 있다

또한 통행시간 예측의 정확성을 높이기 위해서는 적정 대표값 및 과거 데이터량 외에 갱신간격(Update Time Interval)에 대한 적정 기준을 정립할 필요가 있다.

## 참고문헌

1. 유소영 · 노정현 · 박동주(2004), "통행시간 추정 및 예측을 위한 루프검지기 자료의 최적 집계간격 결정", 대한교통학회지, 제22권 제6호, 대한교통학회, pp. 109~119.
2. 한국도로공사(2004), "위치기반형 동적경로탐색기법 및 유통프레임워크 개발", 기술기반보고서, pp.34~49.
3. 한국도로공사(2003), "고속도로 우회도로 ITS 구축 1단계 실시계획 보고서(1/2)"
4. 이의은 · 김정현(2002), "시간차집현상을 고려한 장거리구간 통행시간 예측 모형 개발", 대한교통학회지, 제20권 제4호, 대한교통학회, pp.51~61.
5. 김남선 · 이승환 · 오영태(2002), "신경망을 이용한 고속도로 여행시간 추정 및 예측모형 개발", 대한교통학회지, 제18권 제1호, 대한교통학회, pp.47~59.
6. 박희원 · 박창호 · 전경수 · 이성모(1999), "비모수 회귀분석을 이용한 실시간 통행시간예측에 관한 연구", 대한교통학회 제36회 학술발표회, 대한교통학회, pp.452~457.
7. Road E. Turochy(2006), "Enhancing Short-Term Traffic Forecasting With Traffic Condition Information", Journal of Transportation Engineering, Vol. 132, No. 6, pp.469~474.
8. Lelitha Vanajaskshi(2004), "Travel Time Prediction Using Support Vector Machine Technique", Texas A&M University College Station.
9. Peng Wu(2003), "Automated Data Collection, Analysis, and Archival", University of Utah, MPC Report No. 03-153.
10. Stathopoulos and Karlaftis(2002), "Temporal and Spatial Variations of Real-time traffic data in urban areas", Transportation Research Board 83th Annual meeting, Washigton D.C.
11. Smith and Turoch(2002), "Measuring Variability in Traffic Conditions using Archived Traffic Data", Transportation Research Board 81th Annual meeting, Washigton D.C.
12. Steven I.J. Chien, Chandra Mouly Kuchipudi (2002), "Dynamic Travel Time Prediction With Real-Time and Historical Data", Transportation Research Board 81th Annual meeting, Washigton D.C.
13. Gajewski B.J., Turner. S.M, Eisele, W.L., Spiegelman, O.H.(2001), "ITS Data Archiving : Statistical Techique for Determining Optimal Aggregation Widths for Inductance Loop Detector Speed Data", Transportation Reserch Record 1719, pp.85~93.
14. Shawn M. Turner, William L. Eisele, Byron J. Gajewski, Luke P. Albert, and Robert J. Bens(1999), "ITS Data Archiving : Case Study Analyses of San Antonio TransGuide Data", Texas Transportation Institute, Report No. FHWA-PL 099-024.
15. Park, D. and Rillet, L.,(1999), "Forecasting Freeway Link Travel Times with a Multilayer Feedforward Neural Network", International Journal of Computer-Aided Civil and Infrastructure Engineering, Special Issue on Advanced Computer Technologies in Transportation Engineering, Blackwell Publishers, Malden, MA and Oxford, UK, Volume 14, pp.357~367.
16. Saad Shbaklo, Chandra Bhat, Frank Koppelman (1992), "Short-Term Travel Time Prediction Review of Literature and Methods", ADVANCE Project Report TRF-TT-01.

17. Hoffmann, J. Janko(1990), "Travel Time As A Basic Of The LISB Guidance Strategy". In

Proceedings Of IEEE Road Traffic Control Conference, IEEE, Ner York, pp.6~10.

- ♣ 주 작성자 : 김동호
- ♣ 교신저자 : 박동주
- ♣ 논문투고일 : 2008. 2. 8
- ♣ 논문심사일 : 2008. 3. 15 (1차)  
2008. 10. 15 (2차)  
2008. 3. 10 (3차)  
2008. 8. 12 (4차)  
2008. 9. 21 (5차)
- ♣ 심사판정일 : 2008. 9. 21
- ♣ 반론접수기한 : 2009. 2. 28
- ♣ 3인 익명 심사필
- ♣ 1인 abstract 교정필