

■ 論 文 ■

기종점 통행표 산출모형의 적용성 평가

Adaptability Questions of O-D Table Estimation Models

오 상 진

(충북대학교 도시공학과 대학원 박사과정)

박 병 호

(충북대학교 도시공학과 교수)

목 차

- | | |
|-------------------|------------------|
| I. 서론 | |
| II. 기존이론의 검토 | 3. 평가방법 |
| 1. 선형계수 산출모형 | IV. 모형의 적용 |
| 2. 엔트로피 모형 | 1. 완전한 링크교통량 자료 |
| 3. 통계적 모형 | 2. 불안정한 링크교통량 자료 |
| III. 분석 및 평가방법 설정 | 3. 추가정보 |
| 1. 분석방법 | V. 결론 및 제언 |
| 2. 입력자료의 설정 | 참고문헌 |

요 약

본 연구는 도로교통량을 이용한 기종점 통행표 산출모형들의 적용성 평가를 다루고 있다. 연구의 목적은 두 가지로, 첫째 기종점 통행표 산출을 위한 여러 모형들(선형계수 산출모형, 엔트로피모형 및 통계적 모형)의 예측력을 평가하고, 둘째 이들 모형중에 자료의 확보 정도에 따라 적용성이 보다 우수한 모형을 찾아내는 데 있다. 이를 위해 본 연구에서는 Sioux Falls의 가로망 자료, 교통계획 프로그램인 MINUTP의 균형배정법 및 다양한 형태의 자료확보 상황을 전제로 분석하고 있다. 주요 연구결과는 다음과 같다.

첫째, 관측 링크 통행량 자료가 모두 있는 경우엔 통계적 모형의 신뢰도가 가장 높다. 그러나 관측 통행량 자료가 부족하거나 불확실할 경우에 가장 민감한 모형 또한 통계적 모형이다.

둘째, 선형계수 산출모형은 타 모형에 비해 전체적으로 예측력이 낮다. 그러나 관측링크 통행량자료가 부족하거나 불확실할 경우에도 그 예측력은 크게 달라지지 않는다.

셋째, 엔트로피모형에서 THE/1은 링크별 관측 통행량의 충분한 확보와 자료의 일관성을 요구하지만, THE/2는 그들에 크게 영향을 받지 않는다.

마지막으로 추가정보(총통행량 자료, 존별 유입·유출량 자료)가 주어진 경우, 모형들의 예측력은 다소 차이가 있으나 모두 향상된 것으로 분석된다.

본 연구는 자료의 확보 정도에 따라 적정 기종점 산출모형을 선택하는데 도움을 줄 뿐만 아니라, 향후 모형 자체의 특성에 대한 정보를 제공하여 모형개발이나 수정에 도움을 줄 것으로 기대된다.

1. 서론

교통계획 과정에서 기중점 통행표(Origin Destination Trip Matrix)는 가장 기본적이며, 중요한 입력자료 중의 하나이다. 기중점 통행표의 정밀도가 교통계획 전체과정에 미치는 영향 또한 매우 크다. 따라서 보다 정확한 기중점 자료를 산출하는 것은 교통계획과정에서 매우 중요한 일이다.

전통적인 O-D표 산출 방법은 많은 조사인력, 시간과 비용이 소요되며, 표본자료가 가지는 오차, 편향(bias), 자료처리 시간 등 막대한 노력을 수반한다.

이러한 문제에 대한 혁신적인 방법은 다음 세 가지 방향으로 발전되고 있다. 첫 번째는 이전 가능한 계수에 기반을 둔 단순화된 기법이며, 두 번째는 도로교통량과 여타 정보로부터 O/D 매트릭스를 추정하는 모형이다. 그리고 세 번째로는 기타 단순화된 기법들, 즉 자기정산(self-calibrating) 중력모형, 부분 매트릭스 기법, Heuristic 기법 등이 여기에 속한다.¹⁾

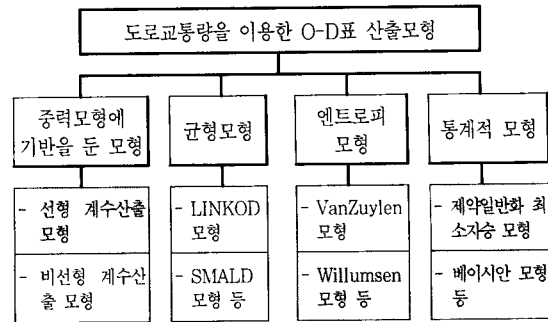
그러나 기존의 O/D표 산출모형에 관한 연구들은 각 모형을 통해 산출된 O/D표를 실제 O/D표와 비교 분석하는데 그치고 있어, 각 모형들이 입력자료의 불완전성에 어떠한 결과를 도출하는지에 관한 연구는 미흡한 실정이다.

본 연구는 도로교통량을 이용한 기중점 통행표 산출모형들의 적용성 평가를 다루고 있다. 연구의 목적은 두 가지로, 첫째 기중점 통행표 산출을 위한 여러 모형들의 예측력을 평가하고, 둘째 이들 모형중에 자료의 확보정도에 따라 적용성이 우수한 모형을 찾아내는 데 있다.

이를 위해 본 연구에서는 Sioux Falls의 가로망 자료를 기본으로 교통계획 프로그램인 MINUTP의 균형배정법, 다양한 평가지표 등을 활용하여 기중점 통행표 산출모형의 적용성을 평가한다.

II. 기존이론의 검토

1970년대 이후 도로교통량과 추가정보를 근거로 O/D 매트릭스를 추정하기 위해 다양한 모형들이 개발되어왔다. 이러한 모형들은 다음 네 가지 범주, 즉



〈그림 1〉 O-D표 산출모형의 분류

①중력모형에 기반을 둔 모형, ②균형 모형, ③엔트로피 모형, 그리고 ④통계적 모형으로 구분된다.

도로교통량을 이용하여 O-D표를 산출하는 모형에 관한 연구는 국내에서 간헐적으로 이루어지고 있지만, 대부분의 연구들은 완전한 입력자료를 가정하여 각 모형을 이론적 혹은 실험적으로 비교하는 연구들이다. 반면에 국외 연구의 경우는 입력자료가 부족하거나 불확실한 상황에서 O-D표를 예측하는 이론들이 활발히 연구되고 있다.

국외의 연구로서 Sven Erlander 외 2인(1985)은 부족한 자료와 불확실한 자료를 이용하여 순수 중력모형의 O-D표 예측을 논의한 바 있다. 기준년도의 Sample O-D표가 있다는 가정하에 그 O-D표의 일부가 부족한 상태에서 불확실한 장래의 O-D표를 예측하는 내용이다.

Kurt Jornsten(1993)은 일관성이 없는 관측 링크통행량에 확률적 프로그래밍 접근방법을 사용하여 최대한 자료의 일관성을 확보하는 방법을 제안하고 있다.

국내연구로 정필현(1991), 김동선(1989), 김영찬(1985) 등의 연구가 있다. 정필현은 베이지안 추정모형과 제약일반화 최소자승 모형, 다중목적함수 결합모형을 카이제곱 통계량을 이용하여 비교 분석하고 있다.

정의 연구에서 청주시의 조사된 가로 통행량 및 O/D 표를 이용하여 각 모형의 예측력을 평가한 결과, 베이지안 추정모형과 제약일반화 최소자승 모형이 다중목적함수 결합모형에 비해 예측력이 우수한 것으로 나타났다.

김동선은 중력모형을 기반으로 한 모형, Equilibrium모형(LINKOD), Entropy모형 및 통계적 모

1) 첫 번째 방법은 NCHRP에 의해 개발되었으며, 1980년초 이후에는 신속응답(quick response) 분석을 위한 몇몇 소형 컴퓨터 프로그램이 개발되어 사용되고 있다.

형을 체계적으로 정립하였다. 김의 연구에서는 실제 적용되는 과정을 프로그램화하면서 각 모형의 이론적 장·단점을 논의하였지만 실제 적용결과를 제시하지는 않고 있다.

김영찬은 Entropy 극대화 모형을 이용하여 Under-specification 문제를 해결하려 하였지만, 인위적인 방법에 의해서 정확한 O/D를 예측하는데는 한계가 있는 것으로 결론을 맺고 있다.

이와 같이 기존의 국내 연구들은 입력자료가 완전히 구축되어 있는 상황에서 각 모형들을 적용하여 결과를 실제 값과 비교하는데 그치고 있는 실정이다.

본 연구는 완전한 정보를 가지고 있을 경우뿐만 아니라 부족하거나 불확실한 입력자료의 경우에도 연구 대상이 된다는 점에서 기존연구와 차별된다. 즉, 각 모형의 입력자료인 관측링크 통행량이 모두 갖추어져 있지 않거나, 관측오류 등으로 인한 오차를 포함하고 있는 경우에도 각 모형의 예측력을 비교하고 있다.

본 연구에서는 중력모형에 기반을 둔 모형중 선형 계수산출 모형, 엔트로피 모형, 통계적 모형중 베이지안 모형과 제약일반화 최소자승모형을 대상으로 적용성을 평가한다.

1. 선형계수 산출모형

중력모형에 기반을 둔 모형들은 교통량을 기초로 통행분포를 수행하는 대표적인 모형이다. O-D표는 선형 회귀분석이나 비선형 회귀분석을 이용하여 추정되는데, 본 연구에서는 선형 계수산출모형을 대상으로 한다.

선형모형의 첫 시도는 D. E. Low(1972)의 중력식 선형모형으로써, 그 기본형태는 식(1)과 같다.

$$T_{ij} = \alpha P_i E_j d_{ij}^{-2} \quad (1)$$

여기서,

T_{ij} : i기점에서 j종점으로의 통행량

P_i, E_j : 존 i, j에서의 인구, 고용자수

d_{ij} : 거리 혹은 통행시간

α : 도출할 계수

본 모형에서 α 는 통행유출과 유입의 결합계수이며, 진량배정법을 이용하고 있다.

$$P_{ij}^a = \begin{cases} \text{만약 } i \text{와 } j \text{간의 통행이 링크 } a \text{를} \\ \text{이용하지 않는다면, } 0 \\ \text{만약 이용한다면, } 1 \end{cases}$$

$$\begin{aligned} \text{이때, } V_a &= \sum_{ij} T_{ij} \cdot P_{ij}^a \\ &= \alpha \sum_{ij} (P_i E_j d_{ij}^{-2}) \cdot P_{ij}^a \end{aligned} \quad (2)$$

이 모형은 요구하는 자료가 비교적 적고, 산출과정이 비교적 간단하다는 장점을 갖고 있다. 그러나 이 모형에서는 타 모형과 달리 존별 인구 및 고용자수에 관한 자료가 필요하다.

2. 엔트로피 모형(THE Model)

엔트로피 극대 그리고 정보 극소의 기법은 도시 및 지역계획과 교통계획에 모형을 구축하는 용도로 Wilson(1970)의 과업 이후 여러 해 동안 사용되어 왔다.²⁾

Van Zuylen(1980)은 링크의 관측 통행량으로부터 얻을 수 있는 정보가 완전한 통행표를 결정하기에는 불충분하다. 따라서 Van Zuylen(1980)은 식(3)에 포함된 정보에 가능한 적은 정보를 추가하여 통행표를 결정하는 것이 합리적인 것으로 보는 정보극소의 관점에서 문제에 접근하고 있다.

$$V_a = \sum_i \sum_j P_{ij}^a T_{ij} \quad (3)$$

여기서, V_a : a 링크의 교통량

P_{ij}^a : i기점에서 j종점으로의 통행이 a링크를 이용할 확률

반면에, Willumsen(1978a, 1978b)은 엔트로피 극대화 접근법을 따름으로써 동일한 문제를 다루고 있다.³⁾

본 연구에서는 교통계획 패키지인 THE(The Highway Emulator) 프로그램의 통행표 산출 모듈의 이

2) 엔트로피 모형의 예측 알고리즘은 어떤 도로교통량을 만들어 내기 위해 가장 유용직한 통행표를 산출하는 것이다.

3) 이 모형은 정산과정이 다소 어렵다는 단점이 있으나, 혼잡한 대도시 가로망 분석에도 적용이 가능하다는 장점이 있다.

론적 기반을 이루고 있는 Willumsen 모형을 중심으로 분석하되, 회전교통량 자료가 없는 경우(THE/1)와 있는 경우(THE/2)로 구분한다.

3. 통계적 모형

1) 베이시안 추정기법

베이시안 추정기법은 베이즈 정리를 이론적 배경으로 한다. 이 방법은 비례배정일 경우와 통행발생 확률을 이미 알고 있을 경우에 매우 유효하다.

필요한 입력자료는 기존에 조사된 기종점자료, 추정될 시기의 도로교통량, 기종점자료의 분산-공분산 행렬, 도로교통량의 분산-공분산 행렬 및 통행자의 도로선택확률이다. 자료를 구성하는 각 추정값의 신뢰도는 분산-공분산 행렬에 의해 조절된다.

초기 매트릭스 분포를 포아송(Poisson)분포로 보고, 이를 다변수 정규분포로 취급하여 초기평균벡터와 분산 매트릭스를 후기평균벡터와 분산매트릭스로 현실화(update)하는 최종 식은 식(4)와 같다.

$$T_e = T_0 + UP'(W + PUP')^{-1}(V_0 - PT_0) \quad (4)$$

여기서,

- T_e : 추정될 기종점 자료 ($n \times 1$)
- T_0 : 기존에 조사된 기종점 자료 ($n \times 1$)
- V_0 : 추정될 시기의 도로교통량 ($m \times 1$)
- U : T_0 의 분산-공분산 행렬 ($n \times n$)
- W : V_0 의 분산-공분산 행렬 ($m \times m$)
- P : 통행자의 도로이용 확률 ($m \times n$)
- n : 기종점 쌍의 수
- m : 교통량이 측정된 도로의 수

그러나 식(4)는 통행특성의 변화를 다루지 못하므로, 마허(Maher, 1983)는 통행특성의 변화에 따라 조절된 α^* 과 β^* 를 이용하여, 식(5)를 제안하고 있다.

$$T_e = T_0 + \beta^* uP'(\alpha^* w + \beta^* PuP')^{-1} \cdot (V_0 - PT_0) \quad (5)$$

2) 제약 일반화 최소자승법

제약 일반화 최소자승법(CGLS:constrained generalized least square model)은 일반화 최소자승법의 최적해에서 추정된 기종점자료가 0보다 작은 값을 가질 가능성에 대해 0보다 작은 값을 갖지 않도록 제약을 주는 형태이다.

$$\begin{aligned} \text{Min } L = & \frac{1}{2} (T_e - T_0)' U^{-1} (T_e - T_0) + \\ & \frac{1}{2} (V_0 - PT_0)' W^{-1} (V_0 - PT_0) + \mu'(C - T_e) \end{aligned}$$

(6)

제약조건 : $T_e \geq C$

여기서,

- L : 목적함수
- μ : 라그랑주 승수
- C : 기종점자료를 구성하는 통행량이 0보다 크게 추정되게 하기 위한 상수, $C > 0$

제약일반화 최소자승모형의 목적함수 식(6)의 최적해는 식(7)과 같다.

$$\begin{aligned} T_e = & T_0 + UP'(W + PUP')^{-1}(V_0 - PT_0) \\ & + (U - UP'(PUP' + W)^{-1}PU)\mu \end{aligned} \quad (7)$$

통행특성의 변화를 다루기 위해 베이시안 추정모형에서와 같이 상수 α^* , β^* 를 이용하면, CGLS 모형의 최적해인 식(7)은 식(8)과 같이 전환된다.

$$\begin{aligned} T_e = & T_0 + \beta^* uP'(\alpha^* w + \beta^* PuP')^{-1}(V_0 - PT_0) \\ & + (\beta^* u - \beta^* uP'(\beta^* PuP' + \alpha^* w)^{-1}\beta^* Pu)\mu \end{aligned} \quad (8)$$

III. 분석 및 평가방법 설정

본 연구는 가로 통행량으로부터 기종점 통행표를 산출하는 모형들의 실제 적용성을 평가하기 위한 연구이다. 즉, 도로교통량을 이용한 기종점 통행표를 산출할 경우에 발생하는 일반적인 문제인 불일치(inconsistency) 문제와 불충분한 조건식(under-specification) 문제의⁴⁾ 영향과 도로 교통량 이외의 추가자료가 있는

4) 이것은 여러 개의 가능 통행표가 똑같은 링크통행량을 만들어낸다는 문제이다.

경우의 적용성을 평가하는 것이다.

1. 분석방법

기중점 통행표 산출모형의 실제 적용성을 평가하기 위해 다음과 같은 분석방법을 설정한다.

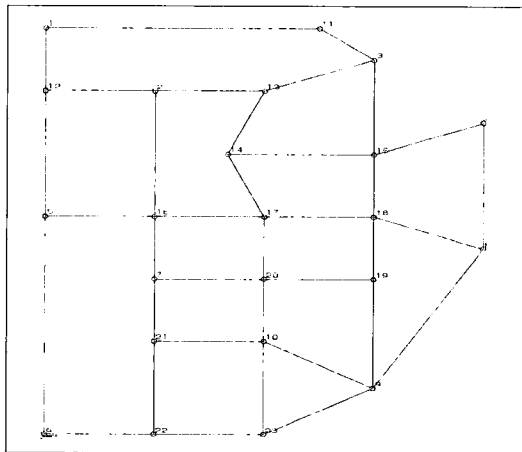
첫째, 입력자료의 확보정도에 따라 모형의 적용성을 분석하기 위해, 여러 가지 상황을 모사하도록 입력자료를 구축한다.

둘째, 각 모형의 알고리즘을 FORTRAN 프로그램으로 구현한다. 다만 엔트로피 모형은 THE 프로그램의 통행표 추정 모듈을 이용한다.

셋째, 각 모형에 의해 산출된 O-D표의 노선배정에는 MINUTP의 균형배정법⁵⁾을 이용한다.

넷째, 본 논문에서 적용할 네트워크는 <그림 2>에 나타나 있는 Sioux Falls의 가로망이다. 이것은 총 23개의 노드와 36개(일방향 72개)의 링크로 구성되어 있다. 또한 23개의 노드 모두가 존 센트로이드(zone centroid)로 구성되어 있다.

마지막으로 본 연구의 목적에 적합한 평가지표를 선정하여, 각 노선배정 결과와 실제 관측 통행량을 비교·분석하고, 현실 여건에 따라 각 모형의 적용성을 평가한다.⁶⁾



<그림 2> Sioux Falls의 가로망

2. 입력자료의 설정

1) 완전한 링크 교통량 자료

첫 번째 경우는 Sioux Falls의 총 36개의 링크에 대한 관측 교통량이 모두 갖춰져 있는 경우이다.

2) 불완전한 링크 교통량 자료

본 연구에서는 일반적인 O-D표 산출 모형에서 제기되고 있는 '불일치' 문제와 '불충분한 조건식' 문제를 유발시키기 위해 다음과 같은 가상 환경을 설정한다.

우선 '불충분한 조건식' 문제는 Sioux Falls의 36개 링크중 임의⁷⁾의 5개(약 14%), 10개(약 28%), 15개(약 45%) 및 20개(약 56%) 링크의 관측 통행량이 없는 경우들로 설정하고 있다.

또한 '불일치' 문제는 36개 링크중 임의의 5개, 10개, 15개 및 20개 링크의 통행량 대신에 임의의 값⁸⁾으로 대체하는 방법으로 가상환경을 설정한다.

3) 추가정보

여기에서는 일부 추가정보가 있는 경우에 각 모형의 예측력이 어떻게 달라지는지 분석한다.

우선, 총 통행량이 있는 경우엔 총통행량의 성장률(F)을 이용한 균일성장율법 적용하여 O-D표의 각 셀을 다시 조정한다.

$$t'_{ij} = t_{ij} \times F \tag{9}$$

또한 존별 통행 유출량에 대한 정보가 있는 경우엔 기점존의 성장률(F_i)를 적용하여 구하고자 하는 O-D표의 해당 셀값들을 조정한다.

$$t'_{ij} = t_{ij} \times F_i \tag{10}$$

아울러 각 존별 통행유출량과 유입량에 대한 정보

5) O/D표 산출모형의 적용성을 평가하기 위하여 동일한 노선배정법과 모형내 동일한 파라미터를 적용한다.
 6) 각 모형을 통해 산출된 O/D표와 실제 O/D표간의 비교도 동일한 평가지표를 사용하여 분석하였으나, 링크통행량의 비교와 차이를 보이지 않으므로 생략한다.
 7) 링크수 보다 작은 난수를 발생시켜, 그 수에 해당하는 링크의 자료를 삭제한다.
 8) 자료의 입력 방법은 우선 링크수(36)보다 작은 난수를 발생시켜, 그 수에 해당하는 링크를 선정하고, 다시 난수(여기서 난수는 실제 통행량 중 최대치 이하의 난수를 발생시킴)를 발생시켜 해당 링크에 대입하는 것이다.

가 있는 경우엔 프라타(T. J. Frata)에 의해 개발된 프라타법을 적용하여 O-D표를 조정한다.

$$t_{ij}' = t_{ij} \cdot E_i \cdot F_j \cdot \frac{L_i + L_j}{2}$$

$$L_i = \frac{\sum_{j=1}^n t_{ij}}{\sum_{j=1}^n t_{ij} \cdot F_j}, \quad L_j = \frac{\sum_{i=1}^n t_{ij}}{\sum_{i=1}^n t_{ij} \cdot E_i} \quad (11)$$

여기서, E_i, F_j 는 준별 유출·유입량의 성장율을 의미하며, L_i, L_j 는 보정식이다.

3. 평가방법

본 연구에 적용된 통계분석 방법은 평균제곱근 오차(RMSE), 퍼센트 평균제곱근 오차(%RMSE), 카이제곱(CHI-square: χ^2), 상관계수 R 그리고 모수적 검증방법인 정보이득(information gain)의 다섯 가지이다.

우선, 실제 관측링크 통행량과 모형에 의해 산출된 O-D의 배정 통행량간의 오차분포 특성을 분석할 수 있는 평균제곱근 오차(RMSE)와 퍼센트 평균제곱근 오차(%RMSE)의 계산식은 식(12)와 같다.

$$RMSE = \sqrt{\frac{\sum_i (t_i - T_i)^2}{N}}$$

$$\%RMSE = \left(\frac{RMSE}{T_e}\right) \times 100 \quad (12)$$

- 여기서, t_i : 링크 i의 배정된 통행량,
- T_i : 링크 i의 실제 관측 통행량,
- N : 링크수,
- T_e : 실제 관측 링크통행량의 평균

그리고 실제 관측 통행량과 배정된 링크별 통행량간의 동일성을 검증할 수 있는 카이제곱(CHI-square: χ^2)과 두 변량간의 비례적 관계를 나타내는 척도로서 상관계수 R은 식(13)과 식(14)에 의해 산출된다.

$$\chi^2 = \sum_i \frac{(t_i - T_i)^2}{T_i} \quad (13)$$

$$R = \frac{N \sum_i t_i T_i - \sum_i t_i \sum_i T_i}{\sqrt{N \sum_i t_i^2 - (\sum_i t_i)^2} \cdot \sqrt{N \sum_i T_i^2 - (\sum_i T_i)^2}} \quad (14)$$

또한 모수적 검증방법중 정보이득(information gain) 검증법을 이용하여 모형들의 예측력을 분석한다.

$$IG = \sum_{i=1}^N \frac{T_i \ln\left(\frac{T_i T_A}{t_i T_0}\right)}{T_0} \quad (15)$$

여기서, T_A 는 가로망에 배정된 링크통행량의 총합, 그리고 T_0 는 실제 관측링크통행량의 총합을 나타낸다.

IV. 모형의 적용 및 평가

1. 완전한 링크 교통량 자료

관측링크 통행량이 모두 갖추어져 있는 경우로써, 전술한 방법에 의해 대상가로망에 노선배정한 결과는 <표 2>, <그림 3>~<그림 8>과 같다. 또한 링크별로 배정된 교통량과 관측교통량과의 비교분석을 위한 모형별 통계검증 결과는 <표 1>과 같다.

<표 1>에 나타나듯이 베이시안 추정모형(Bayesian)과 제약일반화 최소자승법(CGLS)과 같은 통계적 모형이 타 모형에 비해 우수한 예측력을 보인다. RMSE, χ^2 등 모든 통계검증 지표에 있어서 통계적 모형이 가장 우수하다. 다음으로 선형계수 산출법(Linear), THE/1, THE/2 순으로 분석된다.

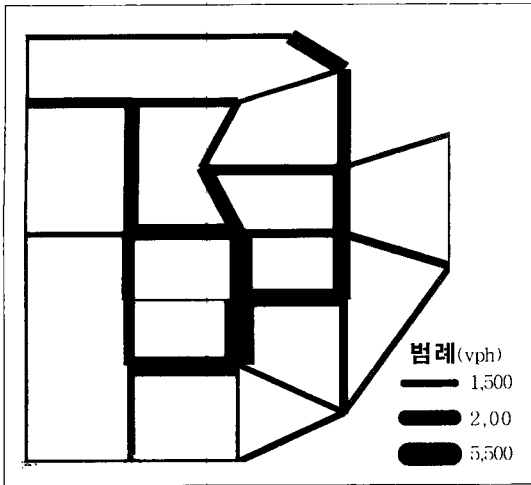
<표 1> 완전한 정보의 경우 모형별 통계검증

구분	Linear	Bayesian	CGLS	THE/1	THE/2
RMSE	807.50	558.47	558.47	641.32	599.16
%RMSE	61.01	42.19	42.19	48.45	45.27
CHI	26.171	13,800	13,800	17,704	16,481
R	.78	.96	.96	.70	.76
IG	.30	.17	.17	.41	.37

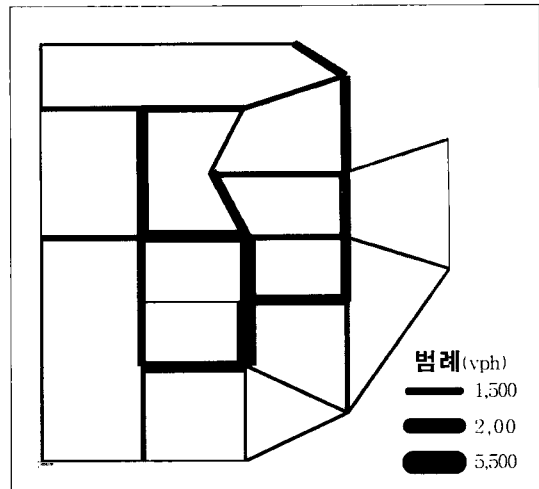
<표 2> 완전한 정보의 경우 모형별 링크 교통량

A	B	Real	Linear	Bayesian	CGLS	THE/1	THE/2	A	B	Real	Linear	Bayesian	CGLS	THE/1	THE/2
1	11	710	296	314	314	441	265	13	14	1,005	296	393	393	874	937
1	12	850	358	599	599	352	334	14	13	930	358	459	459	1,098	1,560
2	12	1,335	478	686	686	1,147	1,153	14	15	1,955	478	1,279	1,279	1,203	657
2	13	1,270	558	929	929	1,587	1,121	14	17	1,365	558	712	712	1,643	1,895
2	16	1,885	969	1,494	1,494	2,130	2,080	15	3	1,725	969	1,260	1,260	1,663	1,346
3	11	2,010	630	1,100	1,100	1,882	1,690	15	4	705	630	364	364	554	639
3	13	640	339	591	591	887	981	15	14	870	339	323	323	1,259	564
3	15	1,650	669	1,325	1,325	1,887	1,970	15	18	2,770	669	1,963	1,963	2,923	2,443
4	8	410	287	154	154	145	151	16	2	1,885	287	1,494	1,494	2,130	2,080
4	15	705	483	364	364	553	639	16	5	630	483	787	787	572	848
5	6	790	211	654	654	376	375	16	7	2,240	211	1,415	1,415	1,225	1,405
5	12	530	176	316	316	113	93	16	17	1,535	176	1,352	1,352	1,225	1,555
5	16	630	570	787	787	572	848	17	14	2,375	570	1,733	1,733	1,812	2,611
6	5	790	211	654	654	376	375	17	16	2,725	211	2,090	2,090	1,440	1,971
6	22	800	380	554	554	506	287	17	18	1,110	380	752	752	1,149	1,459
7	16	1,050	738	677	677	1,010	989	17	20	1,790	738	1,192	1,192	2,080	2,800
7	20	400	270	197	197	31	372	18	8	1,040	270	388	388	696	1,125
7	21	1,920	1,267	1,325	1,325	1,776	946	18	15	1,760	1,267	942	942	2,754	1,727
8	4	410	287	154	154	146	151	18	17	1,265	287	897	897	1,387	2,240
8	9	760	517	345	345	313	268	18	19	2,605	517	1,751	1,751	2,989	1,768
8	18	1,040	538	388	388	696	1,124	19	9	1,150	538	600	600	2,004	1,308
9	8	760	517	345	345	313	269	19	18	1,750	517	875	875	3,059	1,834
9	10	780	1,213	382	382	407	161	19	20	2,740	1,213	1,927	1,927	2,387	2,184
9	19	1,090	517	599	599	2,002	1,308	20	7	110	517	116	116	14	214
9	23	855	319	472	472	1,672	735	20	10	3,160	319	2,076	2,076	1,360	1,382
10	9	720	1,194	380	380	405	161	20	17	3,835	1,194	2,806	2,806	2,224	3,150
10	20	4,120	2,432	2,734	2,734	1,560	1,640	20	19	1,945	2,432	1,052	1,052	2,459	2,250
10	21	2,200	2,211	1,294	1,294	1,768	1,680	21	7	1,020	2,211	668	668	1,578	687
10	23	395	865	284	284	579	568	21	10	3,100	865	1,952	1,952	1,966	1,938
11	1	710	296	314	314	441	265	21	22	890	296	623	623	185	535
11	3	2,010	630	1,100	1,100	1,882	1,690	22	6	800	630	554	554	506	287
12	1	850	358	599	599	352	334	22	21	890	358	623	623	185	534
12	2	1,335	478	686	686	1,147	1,153	22	23	800	478	378	378	1,350	342
12	5	530	176	316	316	113	94	23	9	855	176	472	472	1,672	735
13	2	1,270	558	929	929	1,587	1,121	23	10	395	558	284	284	579	568
13	3	565	322	657	657	1,111	1,604	23	22	800	322	378	378	1,349	342

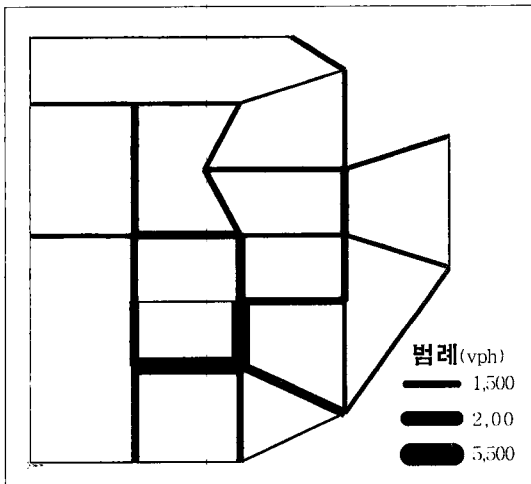
주) 여기서 Real은 관측 링크통행량, Linear은 선형계수 산출모형, Bayesian은 베이시안 모형, CGLS는 통계적 모형중 제약일반화 최소자승법, THE/1은 관측링크통행량만을 고려한 엔트로피 모형 그리고 THE/2는 관측링크통행량과 관측 회전교통량을 모두 고려한 엔트로피 모형을 나타낸다.



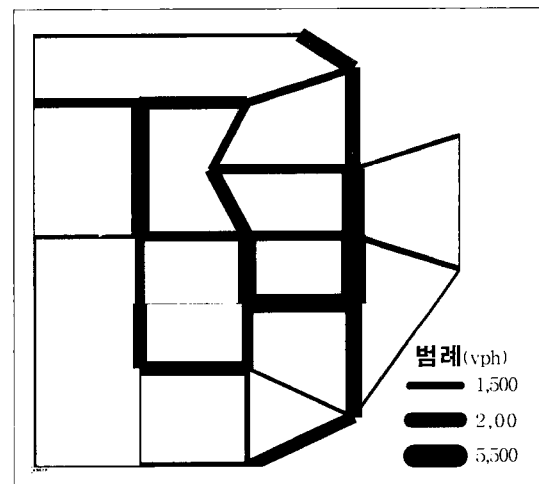
〈그림 3〉 관측 링크통행량



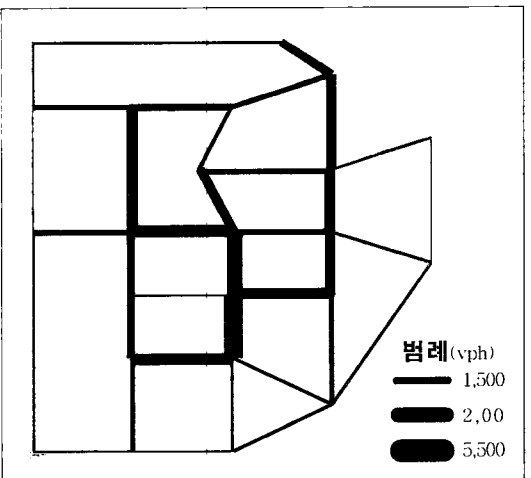
〈그림 6〉 CGLS의 경우 링크통행량



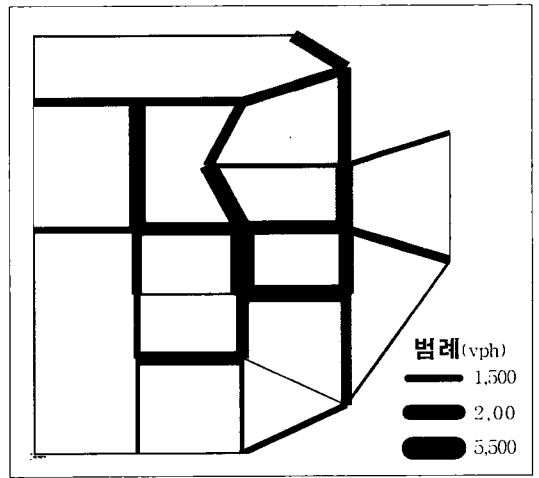
〈그림 4〉 선형모형의 경우 링크통행량



〈그림 7〉 THE/1의 경우 링크통행량



〈그림 5〉 베이저안 모형의 경우 링크통행량



〈그림 8〉 THE/2의 경우 링크통행량

2. 불완전한 링크 교통량 자료

1) 불충분한 자료

전체 36개의 링크 중 일부 관측통행량이 없는 경우의 분석결과는 <표 3>~<표 4>와 같다. 본 연구에서는 지면의 제약으로 불완전 링크통행량 자료와 추가정보의 경우에는 <표 2>와 같은 모형별·링크별 교통량과 <그림 3>에서 <그림 8>까지의 모형별로 배정된 링크통행량에 대한 분석결과는 생략한다.⁹⁾

관측링크 자료의 부족에 가장 민감한 모형은 통계적모형인 반면, 영향을 받지 않는 모형은 선형계수 산출모형과 THE/2 모형이다. 통계적 모형은 자료의 부족정도에 비례하여 RMSE값 및 카이제곱값이 증가한다.

<표 3> 불충분한 정보의 경우 각 모형의 RMSE값

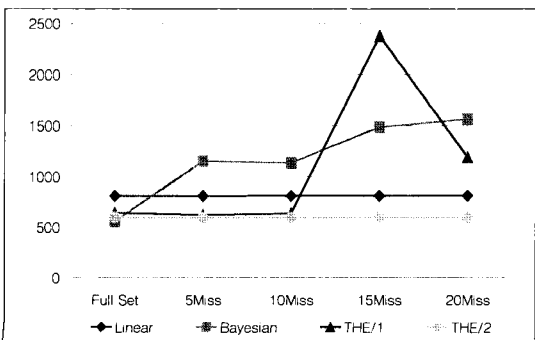
구분	Full Set	5Miss	10Miss	15Miss	20Miss
Linear	807.50	807.84	808.67	809.52	810.29
Bayesian	558.47	1,152.13	1,134.56	1,486.61	1,565.25
THE 1	641.32	621.41	635.37	2,376.95	1,189.56
THE 2	599.16	593.04	594.53	598.40	597.30

주 : 1) CGLS모형은 베이시안 모형과 유사한 결과를 보이므로 생략함(이하 <표 6>까지 동일)
 2) 5Miss는 총 36개의 링크중 5개 링크의 자료가 없는 경우를 나타냄(이하 <표 4>까지 동일)

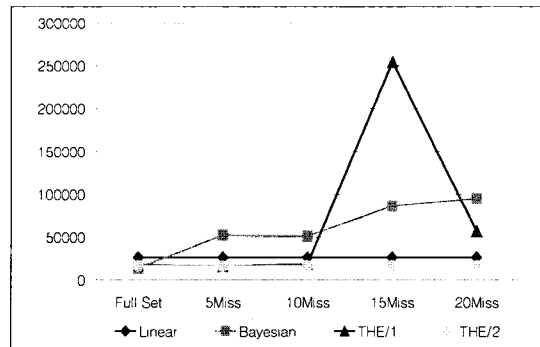
<표 4> 불충분한 정보의 경우 각 모형의 χ^2 값

구분	Full Set	5Miss	10Miss	15Miss	20Miss
Linear	26.171	26.222	26.278	26.352	26.438
Bayesian	13,800	52,336	51,105	86,290	95,246
THE 1	17,704	16,082	18,680	254,542	57,284
THE/2	16,481	16,108	16,412	16,564	16,161

주 : <표 3> 참조



<그림 9> 불충분한 정보의 경우의 RMSE값 비교



<그림 10> 불충분한 정보의 χ^2 값 비교

2) 불확실한 자료

일부 링크에 대한 관측통행량이 실측치와는 다른 임의의 값이 입력된 경우의 분석결과는 <표 5>~<표 6>과 같다.

<표 5> 불확실한 정보의 경우 각 모형의 RMSE값

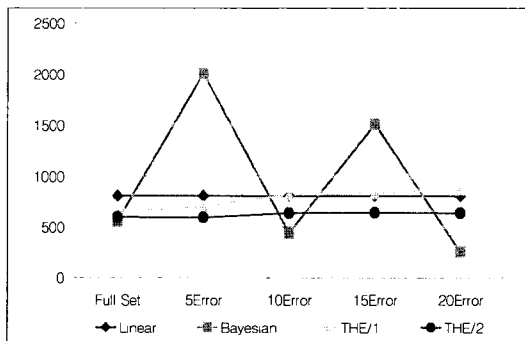
구분	Full Set	5Error	10Error	15Error	20Error
Linear	807.50	806.96	805.90	805.99	805.67
Bayesian	558.47	2,009.94	439.09	1,512.57	257.54
THE/1	641.32	704.04	818.85	831.79	873.08
THE/2	599.16	594.34	637.85	640.93	638.62

주) 5Error는 임의의 5개 링크의 자료가 불일치한 경우를 나타냄

<표 6> 불확실한 정보의 경우 각 모형의 χ^2 값

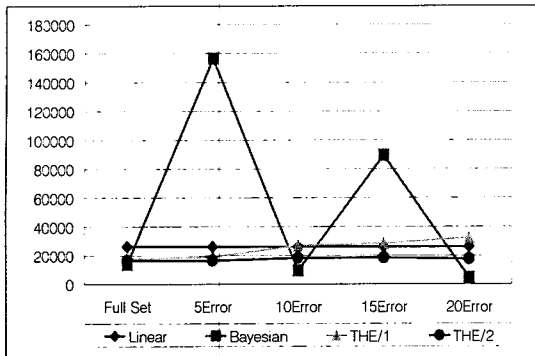
구분	Full Set	5Error	10Error	15Error	20Error
Linear	26.171	26.130	26.062	26.060	26.029
Bayesian	13,800	156,645	9,511	89,539	4,576
THE/1	17,704	19,025	26,844	28,459	32,337
THE/2	16,481	16,568	18,301	18,427	17,573

주) <표 5> 참조



<그림 11> 불확실한 정보의 경우의 RMSE값 비교

9) 아울러 여러 평가지표중 대표적인 값인 RMSE와 χ^2 값을 제시하되, 그밖에 %RMSE, R 및 IG는 지면의 제약상 생략한다.



(그림 12) 불확실한 경우의 χ^2 값 비교

불확실한 관측링크 통행량에 가장 민감한 모형이 통계적 모형인 반면, 영향을 받지 않는 모형은 선형계수 산출모형과 THE/2 모형이다.

통계적 모형의 경우 자료의 정확성이 부족한 경우 RMSE 값 및 카이제곱값이 크게 증가한다. 또한 THE/1의 경우, 링크 통행량 자료의 일관성이 부족하면 다소 영향을 받지만, THE/2의 경우엔 대체로 일정한 통계량을 유지하는 것을 알 수 있다.

본 연구에서는 입력자료의 불충분함과 불확실성에 각 모형의 반응을 분석하기 위해, 난수를 발생시켜 링크를 선정하고 교통량을 삭제하거나 혹은 다시 난수를 발생시켜 임의의 값을 입력시키는 방법을 이용하고 있다. 불완전한 링크 교통량 자료를 모형별로 분석한 결과를 요약하면 다음과 같다.

선형계수산출법의 경우 입력자료의 불완전 정도에 비교적 적게 영향을 받으나 전체적으로 타 모형에 비해 예측력이 낮은 것으로 나타나고 있다.

관측링크 통행량만을 고려한 엔트로피 모형(THE/1)의 경우는 자료가 불완전하면 할수록 신뢰하기 어려운 결과를 나타낸다(15개의 링크 통행량이 없는 경우 많은 오차를 나타냄). 이것은 입력데이터의 임의성에 따른 모형자체의 한계로 판단된다.

그러나 관측링크통행량과 관측 회전교통량을 모두 고려한 엔트로피 모형(THE/2)의 경우 입력자료의 불완전성에 거의 영향을 받지 않는다.

통계적 모형의 경우 타 모형에 비해 예측력은 우수하나 입력자료에 대한 초기정보(가 O/D)를 요구하는 것이 특징이다.

3. 추가정보

1) 총 통행량

총통행량을 알고 있는 경우의 통계분석은 <표 7>과 같다. <표 7>의 결과를 <표 1>과 비교 분석하면, 선형계수 산출모형과 THE/2는 거의 변화가 없는 반면에 통계적 모형의 경우엔 예측력이 개선된 것으로 판단된다.¹⁰⁾

<표 7> 총교통량 정보의 경우 모형별 통계검증

구 분	Linear	Bayesian	CGLS	THE/1	THE/2
RMSE	806.98	253.25	253.25	730.03	611.24
%RMSE	60.97	19.13	19.13	55.15	46.18
CHI	26,137	4,415	4,414	27,287	16,597
R	.78	.96	.96	.70	.76
IG	.30	.17	.17	.41	.37

2) 준별 통행발생량

준별 통행 발생량을 알고 있는 경우의 통계분석 결과는 <표 8>과 같다.

<표 8>의 결과를 <표 7>과 비교 분석하면, 대체로 양호한 결과를 산출한다. 통계적 모형의 경우 RMSE값과 카이제곱값이 매우 낮아지며, 상관계수 값도 1에 접근하게 된다. 정보이득 값도 0에 보다 접근한다.

선형 모형의 경우 카이제곱값은 총통행량 제약시보다 더 낮아지며 상관계수도 보다 1에 접근하는 것을 볼 수 있다. 정보이득값도 약간 더 낮아져 0에 접근하는 것을 볼 수 있다. 그리고 RMSE값의 경우는 오히려 약간 증가하나 의미있는 차이는 아니다.

THE/1의 경우는 RMSE와 카이제곱값이 더욱 낮아지며, 상관계수와 정보이득값도 각각 1과 0에 더욱

<표 8> 준별발생량 정보의 경우 모형별 통계검증

구 분	Linear	Bayesian	CGLS	THE/1	THE/2
RMSE	850.02	197.18	197.09	619.17	1,041.61
%RMSE	64.22	14.90	14.89	46.78	78.69
CHI	22,117	2,410	2,409	20,484	43,331
R	.88	.98	.98	.78	.71
IG	.27	.13	.13	.33	.44

10) THE/1의 경우엔 RMSE값과 카이제곱값이 오히려 약간 높아지는 반면, 상관계수와 정보이득 값은 일정한 수준을 유지하고 있다. 그러나 THE/2의 경우엔 모든 통계량이 일정한 값을 유지하지만, 관측 링크통행량만 적용할 때 보다 예측력은 더 높다는 것을 볼 수 있다.

접근한다. 그러나 THE/2의 경우엔 오히려 RMSE값과 카이제곱값이 큰 쪽으로 상승하고 있다.

3) 존별 유·출입 통행량

존별 통행 유·출입량을 알고 있는 경우의 통계분석 결과는 <표 9>와 같다.

표에 나타난 결과를 <표 7>과 <표 8>과 비교하면, 전체적으로 매우 좋은 결과를 보이고 있다. 통계적 모형의 경우 상관계수(R)는 0.99로 거의 1에 근사한 결과를 나타내며, 정보이득(IG)값도 0.1로 0에 거의 근사한 값을 나타낸다. 그리고 RMSE값도 제약하지 않는 경우 보다 많이 감소된다.

선형계수산출 모형의 경우도 모든 결과 값들이 좋은 결과를 보인다. RMSE값과 카이제곱값은 더욱 수치가 낮아지고 상관계수 및 정보이득 값은 약간 변동하지만 큰 차이를 보이지 않는다.

또한 THE/1과 THE/2의 경우도 RMSE값과 카이제곱값이 큰 쪽으로 감소하고 상관계수값도 1에 접근하는 것으로 나타난다. 그리고 정보이득값도 매우 낮아져 0에 접근한다.

<표 9> 존별유출입량 정보의 경우 모형별 통계검증

구 분	Linear	Bayesian	CGLS	THE/1	THE/2
RMSE	770.66	172.41	172.35	485.06	470.00
%RMSE	58.22	13.03	13.02	36.65	35.51
CHI	18,782	1,828	1,826	13,396	10,201
R	.81	.99	.99	.88	.85
IG	.28	.10	.10	.27	.29

V. 결론 및 제언

본 연구에서는 국내외 문헌조사를 통하여 도로의 관측 통행량 및 추가자료를 이용하여 기종점 통행표를 추정·보정하는 이론들을 검토한다. 또한 본 연구는 해당 모형들의 알고리즘을 분석하여 프로그래밍한 후, 시험 가로망 자료를 이용하여 모형별 특성을 분석하며, 다양한 현실 여건에 따른 적정 모형을 제시하고 있다.

이러한 분석에는 Sioux Falls의 가로망 자료가 이용되었으며, 교통계획 프로그램인 MINUTP를 이용하여 모형에 의해 산출된 O-D통행량을 비교분석하였다. 여기에서 실제 O-D의 배정결과를 관측 통행량으

로 가정하고 있다. 비교대상 모형으로는 선형계수 산출모형, 통계적 모형의 베이시안법과 제약일반화 최소화승법, 그리고 엔트로피 모형이다.

각 모형의 구현에는 컴퓨터 언어인 Fortran Powerstation 4.0이 이용되었으며, 각 모형의 산출결과를 분석하는 통계분석도 같은 툴(tool)을 이용하였다. 또한 통계분석에는 RMSE, %RMSE, 카이제곱(CHI-square), 상관계수(R), 정보이득(IG)값이 이용되었다. 본 연구에서 도출된 주요 분석결과는 다음과 같다.

첫째, 링크 통행량 자료가 모두 있는 경우에는 통계적 모형의 신뢰도가 높게 나타나고 있다. 따라서 링크 통행량이 모두 갖춰져 있는 경우에는 통계적 모형을 적용하여 O-D표를 추정하는 것이 보다 정확한 결과를 도출할 것으로 기대된다.

둘째, 관측 통행량 자료가 부족한 경우, 즉 불충분한 조건식 문제와 불일치 문제에 있어서 가장 민감한 모형은 통계적 모형이며 가장 둔감한 모형은 선형계수 산출모형이다. 아울러 엔트로피 모형은 관측자료의 확보나 일치성 여부에 상대적으로 적게 영향을 받고 있다.

셋째, 추가정보, 즉 총통행량 자료와 존별 유출량 혹은 유입·유출량 자료는 모형의 예측력을 높이는 데 도움을 준다.

본 연구는 기존연구와는 달리 관측링크 통행량자료가 부족하거나 불확실한 경우의 실제 적용성을 평가할 수 있다는 점에서 그 의의가 있다.

본 연구는 자료의 확보 정도에 따라 적정 기종점 산출모형을 선택하는데 도움을 줄 뿐만 아니라, 모형 자체의 특성에 대한 정보를 제공하여 모형개발에 도움을 줄 것으로 기대된다.

본 연구에서 다루지 못한 여러 모형들, 즉 비선형계수 산출모형, 균형모형 등에 대해서도 자료획득 여부라는 현실여건을 감안한 모형의 적용성이 평가되어야 할 것이다. 아울러 기종점 통행표 혹은 네트워크의 규모나 특성에 따라 예측력이 달라질 수도 있으므로 이에 대한 다양한 연구도 필요할 것이다.

참고문헌

1. 김동선, "O-D표 산출모형의 비교분석을 통한 모형 개발 방향에 관한 연구", 서울대학교 석사학위논문, 1989.

2. 김인태, "엔트로피 극대화방법을 이용한 링크별 O-D패턴 추정에 관한 연구", 서울대학교 석사학위논문, 1994.
3. 정필현, "도로교통량을 이용한 기종점자료 추정모형의 예측력 비교", 연세대학교 석사학위논문, 1991.
4. 이현홍, "Gradient Approach 방법을 이용한 기종점 통행표 보정에 관한 연구", 서울대학교 석사학위논문, 1996.
5. 최재훈·박병호, "통행배정기법 적용성 평가방법의 개발 및 적용, 『대한교통학회지』, 제16권 제2호, 1998. 6.
6. Hai Yang, "The Equilibrium-Based Origin-Destination Matrix Estimation Problem", TRB, 1994, pp.24~32.
7. Hai Yang, Tsuna Sasaki, and Yasunori Iida, "Estimation of Origin-Destination Matrices from Link Traffic Counts on Congested Network", TRB, 1992, pp.419~422.
8. Edward J. Bromage, *The Highway Emulator Model Manual*, Department of Urban Engineering, 1991.
9. Henk J. Van Zuylen, Luis G. Willumsen, "The Most Likely Trip Matrix Estimated from Traffic Counts", TRB, 1980, pp.282~288.
10. J. de D. Ortuzar, L. G. Willumsen, *Modelling Transport Second Edition*, Wiley, 1994.
11. Kurt Jörnsten, "Overcoming the (apparent) problem of inconsistency in origin-destination matrix estimation", *Transportation Science*, 1993.
12. M. Easa, "Urban Trip Distribution in Practice II", *Journal of Transportation Engineering*, Vol. 119, 1993, pp.816~830.
13. Maher, M. J., "Inference on Trip Matrices from Observations on Link Volumes : A Bayesian Statistical Approach", *Transportation Research*, Vol.16B, pp.235~448.
14. Sven Erlander, Kurt Jörnsten and Jan T. Lundgren, "On the estimation of trip matrices in the case of missing and uncertain data", TRB, 1985.