

# 주행거리를 이용한 연료소비량 산정방법: 몬테카를로 기법 중심으로<sup>†</sup>

박천건<sup>1</sup> · 소진영<sup>2</sup> · 이영섭<sup>3</sup>

<sup>1</sup>경기대학교 수학과 · <sup>2</sup>에너지경제연구원 · <sup>3</sup>동국대학교 통계학과  
접수 2012년 2월 3일, 수정 2012년 2월 24일, 게재확정 2012년 3월 2일

## 요약

최근 온실가스 배출량에 대한 국제적 이슈가 지속적으로 대두되어 왔다. 본 연구는 이동 배출원 중에서 도로교통부분에 대한 배출량을 산출하는데 필요한 연료소비량 산정에 관한 통계적인 기법들을 비교한 것이다. 지금까지 연료소비량을 산정하는데 단순히 연료공급량 또는 주행거리에 의한 방법으로 초점을 두어왔다. 그러나 연료공급량을 토대로 산출된 배출량은 다양한 차량 및 연식을 반영하지 못한다. 본 연구는 통계적 관점에서 주행거리와 연비를 조합한 각 차량의 연료소비량 및 모든 차량의 소비한 총연료소비량의 추정에 적용되는 방법들을 나열하고 비교한다. 또한 향후 연구로 제시된 일부 방법들에 대해서 차량 및 연식을 고려한 연료소비량의 추정에 관한 현실적인 적용방법을 제시한다.

주요용어: 몬테카를로, 연료소비량, 연비, 주행거리, 핵밀도추정

## 1. 서론

### 1.1. 연구의 목적

최근 온실가스 배출량에 대한 국제적 이슈가 지속적으로 대두되어 왔다. 우리나라의 경우에 온실가스의 원인으로 여러 요인들이 있지만, 자발적으로 온실가스를 줄일 수 있는 부분으로 에너지소비를 들 수 있다. 에너지소비 비중을 살펴보면 산업은 약 57%이고 수송이 약 21%인 것으로 나타났다. 특히 수송에서 도로교통부분이 대부분의 온실가스 배출량을 차지하고 있다. 향후 자동차 수요 증가로 인하여 해당 온실가스가 증가할 것이며 신뢰성 있는 배출량의 산정이 필요할 것이다. 이를 위해서는 도로교통부문에서 연료소비량의 추정이 우선되어야 한다.

도로교통부분의 온실가스 배출량을 산정하는데 필요한 요인들은 연료소비량, 배출계수, 운행도로의 상태, 배출량에 관한 차량의 정보 등이다. 지금까지 많은 연구들이 온실가스인 CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O의 배출량 및 배출계수에 관한 산정방법에 대한 비교 및 검토를 실행해 왔다 (김기동 등, 2011; 국립환경과학원, 2009). 또한 배출량을 산출하기 위해 실제로 각 차량의 연료소비량으로 합산한 총연료소비량이 사용되어야 하는데도 불구하고, 현실적인 제약으로 인해 정유사들이 제공하는 연료공급량의 기준으로 배출량이 산정되어왔다. 연료공급량의 자료는 에너지경제연구원에서 매년 발간되는 에너지통계연보 (에

<sup>†</sup> 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2010-0009273).

<sup>1</sup> (443-760) 경기도 수원시 영통구 이의동 산 94-6, 경기대학교 수학과, 조교수.

<sup>2</sup> (437-713) 경기도 의왕시 내손순환로 132, 에너지경제연구원, 연구위원.

<sup>3</sup> 교신저자: (100-715) 서울특별시 중구 필동 3가 26, 동국대학교 통계학과, 교수.

E-mail: yung@dongguk.edu

너지경제 연구원, 2011)에서 최종 에너지 소비 부분별 및 에너지원별로 제공되지만, 차량의 종류 및 연식에 대한 연료소비량은 반영하지 못하고 있다. 따라서 다양한 차량정보를 가진 연료소비량의 산출이 필요하기 때문에 이러한 정보가 제공되는 주행거리 및 연비를 기초로 한 연료소비량을 추정할 수 있는 방법론의 확립이 절실히 필요하다. 하지만 주행거리와 연비를 적용하여 연료소비량의 추정에 대한 연구는 드물며 통계적인 접근으로 실행된 연구는 현재까지 거의 없다.

따라서 본 연구에서는 도로교통부에서 연료소비량을 추정할 수 있는 다양한 방법론들을 제시하고 몬테카를로 (Monte Carlo) 기법을 이용하여 나온 결과들을 비교분석하여 현실적으로 적용 가능한 방법론을 도출하는데 목적이 있다.

## 2. 연료소비량 산정

### 2.1. 연료소비량 산정방법

각 차량의 연료소비량은 운행한 거리에 비례하지만, 차량의 종류 (차종), 운전자의 운행형태, 차량의 연식, 운행한 도로의 상태, 교통량 그리고 기상조건, 이외에도 많은 요인들에 의해 결정된다. 본 연구에서 제시하는 것은 앞에서 설명한 자료인 주행거리와 연비를 토대로 각 차량의 연료소비량 및 총연료소비량을 산정할 수 있는 수식을 검토하고 통계적인 기법들을 이용하여 총연료소비량의 추정방법을 비교한 것이다.

만약 각 차량의 연료소비량이 주어진다면 모든 차량의 연료소비량을 합산하면 총연료소비량의 값을 얻을 수 있다. 하지만 이 방법은 여러 가지 제약으로 인하여 현재 많이 사용하는 방법은 각 정유사들이 공급하는 연료의 공급량을 이용하는 것이다. 하지만 이것도 실제 연료소비량이 아니며, 또한 차종을 고려하지 않았기 때문에 객관적인 근거의 결여와 차종에 따른 총연료소비량의 계산이 불가능하다.

표 2.1 자료설명

내용		기호
모집단 (모든 차량)의 수		$N$
표본 (일부 차량)의 수		$n$
자료	$i$ 번째 차량 주행거리 (km)	$d_i$
	$i$ 번째 차량 연료소비량 ( $\ell$ )	$c_i = d_i/e_i$
	$i$ 번째 차량 연비 (km/ $\ell$ )= 연료 $\ell$ 당 주행거리 (km)	$e_i = d_i/c_i$

따라서 통계적으로 총연료소비량을 추정하고 그 신뢰구간을 제시하기 위해서는 표본추출방법을 이용하는 것이 바람직할 것이다. 그러나 모든 차종 및 연료에 따라 표본조사를 하는 것은 시간과 비용이 많이 소요되기 때문에 교통관련 기관에서도 부분적이고 한시적으로 시행해왔다. 그래서 대안적인 방법으로 제시할 수 있는 것은 교통안전공단에서 보유하고 있는 연간 주행거리와 연비의 정보를 이용해서 총연료소비량과 그 신뢰구간을 추정하는 것이다.

본 연구에서 총연료소비량을 추정하기 위해 사용한 자료와 추정하고자 하는 값들은 다음과 같다. 여기서  $i$ 는 편의상 각 차량의 번호이며  $i = 1, 2, \dots, N$ 이다.

본 연구에서 총연료소비량을 추정하는데 여러 가지 방법들을 제시하는 이유는 차량의 연료소비량이 주어지지 않고 다만 주행거리와 연비자료가 주어졌기 때문이다. 주어진 자료의 상황을 크게 두 가지로 나열해보면 (1) 모든 차량의 주행거리와 일부 차량의 연비가 존재하거나 (2) 두 자료 모두 표본으로 존재는 경우이다. 현실적으로 한 차량의 주행거리가 주어진다고 해서 반드시 그 차량의 연비가 존재하는 것은 아니다. 반대의 경우도 존재한다. 본 연구에서 연료소비량을 추정하는데 사용되는 방법들을 비교하기 위해 적용되는 자료는 첫 번째 상황 및 일부 차량의 주행거리와 연비가 서로 대응되는 것으로 가정

한다. 즉, 주행거리는  $d_i, i = 1, 2, \dots, N$ , 연비는  $e_j, j = 1, \dots, n (n < N)$  그리고 일부 차량에 대한 주행거리와 연비의 쌍은  $(s_j, e_j), j = 1, \dots, n$ 으로  $S = \{s_j, j = 1, \dots, n\} \subset D = \{d_i, i = 1, \dots, N\}$ 이다.

<방법 1>은 표본차량의 평균 연료소비량과 전체 차량의 수를 이용하여 총연료소비량 ( $T$ )을 추정한다. 제시된 <방법 1>은 가장 손쉽게 계산할 수 있으며 구체적인 수식은 다음과 같다.

$$\hat{T}_1 = N \cdot \bar{c} \tag{2.1}$$

여기서 표본으로 추출된 각 차량의 연료소비량은  $c_j = d_j/e_j, j = 1, \dots, n$ 이고 연료소비량의 표본평균은  $\bar{c} = \sum_{j=1}^n c_j/n$ 이다.

<방법 2>은 모든 차량의 주행거리 ( $d_i, i = 1, 2, \dots, N$ )와 표본차량에서 얻은 연비 ( $e_j, j = 1, \dots, n$ )를 이용하여 총연료소비량을 산정하는 것이다. 총연료소비량을 추정하기 위해서는 표본에서 제외된 차량의 주행거리에 대응되는 연비가 필요하다. 따라서 표본이 아닌 차량들의 연비는 표본차량에서 계산된 가중조화 평균연비 ( $\hat{c}$ )로 대체할 수 있다. 가중조화 평균연비는 표본차량의 총주행거리 ( $\sum_{j=1}^n d_j$ )에서 총연료소비량 ( $\sum_{j=1}^n c_j$ )을 나눈 것이다.

$$\hat{c} = \frac{\sum_{j=1}^n d_j}{\sum_{j=1}^n c_j} \tag{2.2}$$

그리고 총연료소비량 ( $\hat{T}_1$ )은 모든 차량의 주행거리의 합에 가중조화 평균연비로 나누어 얻어진다.

$$\hat{T}_2 = \frac{1}{\hat{c}} \sum_{j=1}^N d_j \tag{2.3}$$

<방법 3>은 표본차량의 연비와 주행거리로부터 각 차량의 연료소비량을 구하고 그 확률밀도함수 (Probability density function)를 추정한다. 총연료소비량을 추정하기 위해 몬테카를로 (Monte Carlo)기법을 이용한다 (권세혁, 2010). 구체적인 과정은 다음과 같다.

<방법 3>은 표본차량의 연비와 주행거리로부터 각 차량의 연료소비량을 구하고 그 확률밀도함수 (Probability density function)를 추정한다. 총연료소비량을 추정하기 위해 몬테카를로 (Monte Carlo)기법을 이용한다 (권세혁, 2010). 구체적인 과정은 다음과 같다.

(단계 1) 표본 차량  $n$ 개의 주행거리와 연비로부터 연료소비량 ( $c_j$ ) 계산:

$$c_j = d_j/e_j, j = 1, \dots, n$$

(단계 2) 연료소비량의 확률밀도함수를 추정:  $f(c)$

반복 (단계 3) ~ (단계 4):  $k = 1, \dots, m\{$

(단계 3) 연료소비량의 확률밀도함수로부터 모든 차량수 ( $N$ )대한 난수생성:

$$c_{(i,k)} \sim f(c), i = 1, \dots, N$$

(단계 4) 생성된 연료소비량의 합을 계산:  $\tilde{T}_k = \sum_{i=1}^N c_{(i,k)}$

(단계 5) 생성된 총연료소비량의 평균을 계산:  $\hat{T}_3 = \frac{\sum_{k=1}^m \tilde{T}_k}{m}$

<방법 4>는 표본차량의 연비로부터 확률밀도함수를 추정한다. 주행거리의 자료는 제공되지만 연비 자료가 없는 차량에 대해 추정된 연비의 확률밀도함수로부터 난수를 발생하여 임의로 각 차량에 할당한 후 총연료소비량을 추정하는 것이다 (최현석과 김태윤, 2010). 다음에 제시된 절차로 총연료소비량을 구한다.

(단계 1) 표본  $n$ 개 차량의 연비로부터 확률밀도함수를 추정:  $f(e)$

반복 (단계 2)~(단계 4):  $k = 1, \dots, m$

(단계 2) 연비의 확률밀도함수로부터 모든 차량수 ( $N$ )의 난수생성:

$$e_{(i,k)} \sim f(e), i = 1, \dots, N$$

(단계 3) 생성된 연비를 무작위로 각 차량의 주행거리에 할당:  $(d_i, e_{(i,k)}), i = 1, \dots, N$

(단계 4) 모든 차량수 ( $N$ )의 연료소비량을 계산:  $c_i = d_i/e_{(i,k)}, i = 1, \dots, N$

(단계 5) 총연료소비량을 계산:  $\tilde{T}_k = \sum_{i=1}^N c_{(i,k)}$

(단계 6) 생성된 총연료소비량의 평균을 계산:  $\hat{T}_4 = \frac{\sum_{k=1}^m \tilde{T}_k}{m}$

<방법 1>과 <방법 2>는 주어진 표본차량의 연료소비량 또는 연비에 전체 차량수 또는 전체 차량의 주행거리를 기반으로 총연료소비량을 추정한 단순한 방법이다. 반면 <방법 3>은 표본 연료소비량의 확률밀도함수를 추정하여 몬테카를로기법을 적용하여 총연료소비량을 추정하였다. <방법 4>도 <방법 3>과 유사한 것으로 표본 차량으로부터 추정된 연비의 확률밀도함수를 이용하여 총연료소비량을 추정한 것으로 차이점은 전체 차량의 주행거리가 이용되었다는 것이다.

<방법 4>의 장점은 각 차량의 연비가 여러 요인들에 영향을 받기 때문에 다양하게 나타날 수 있다. 따라서 주어진 연비의 분포로부터 무작위로 연비를 생성하여 각 주행거리에 할당해서 얻어진 총연료소비량의 변동을 관찰할 수 있다.

## 2.2. 확률밀도함수의 추정

본 연구에서 제안한 <방법 3>과 <방법 4>에서 연료소비량 및 연비의 확률밀도함수에 대한 추정이 있다. 확률밀도함수 추정의 여러 방법 중에 많은 분야에 적용되어온 핵밀도추정 (kernel density estimation)이 본 연구에 적용되었다 (Park, 2011; Bowman과 Azzalini, 1997; Silvean, 1986; Shim, 2011).

한 점  $x$ 에 대한 확률밀도함수  $f$ 의 핵밀도추정은

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

이다. 여기서 확률표본  $X_i \stackrel{iid}{\sim} f(x), i = 1, \dots, n$ 이고  $h$ 은 대역폭 (bandwidth) 그리고 핵함수인  $K$ 은  $\int K dx = 1$ 이다. 특히 평활모수 (smoothing parameter)인 대역폭과 핵함수의 선택에 대한 개략적인 내용은 Sheather (2004)에 언급되었다.

본 연구에서는 선택한 평활모수는 대역값 (global value)과 핵함수는 가우시안분포가 적용되었고 주어진 자료의 확률밀도함수는 Matlab 6.5에서 제공되는 ksdensity함수 (Martinez와 Martinez, 2002)를 사용하여 추정하였다.

## 3. 모의실험

### 3.1. 모의실험설계

총연료소비량을 추정하는데 제시된 네 가지의 방법들을 비교하기 위해 가상의 자료를 설정하여 모의 실험을 실행하였다. 모의실험에 설정된 자료는 주행거리, 연비 그리고 연료소비량이다. 연료소비량은 주행거리와 연비로부터 얻어지고 연비도 주행거리와 연료소비량이 주어지면 계산될 수 있다. 따라서 주

행거리와 연비에 대한 가상자료를 표 3.1에서 제공되는 정보를 토대로 생성하였다. 제공된 정보인 주행거리는 교통안전공단에서 매년 발간되는 “자동차 주행거리 실태조사”보고서와 연비는 소비자리포트(www.sobijareport.org)에서 일시적으로 조사된 것을 이용하였다. 특히, 주행거리와 연비자료는 생성 시기의 불일치 및 원시자료의 접근이 가능하지 않은 관계로 몇 가지의 가정들을 기초로 가상의 자료를 생성하였다.

주행거리는 휘발유를 사용하는 승용차의 일일주행거리 (km)와 연비의 자료를 생성한다. 표 3.1은 주행거리와 연비의 자료를 생성하기 위해 제공되는 값들이다. 구체적인 자료의 생성은 표 3.1에서 제공되는 평균 및 표준편차에 해당되는 정규분포로 주어진 최솟값 및 최댓값의 범위로 절단된 분포를 구성한다. 주행거리와 연비에 대한 기술통계값은 실제값들이지만 원시자료를 구하기가 어려운 관계로 모집단의 형태가 알려지지 않았기 때문에 임의로 모집단의 분포를 절단된 정규분포로 가정하였다. 자료의 생성은 주어진 비대칭으로 절단된 정규분포로부터 주행거리와 연비를 생성한다. 총연료소비량을 추정하기 위해 사용되는 전체 차량수는  $N = 10,000$ 이고 각 차량의 연비 (또는 연료소비량)는 표본에 의해 추출되며 표본수는  $n = 30, 100, 500, 1000$ 로 설정한다.

표 3.1 주행거리와 연비에 대한 정규분포의 모수 및 조건

	평균	표준편차	최솟값	최댓값
주행거리 ( $d_i$ , km)	31.5	15.5	7.6	327
연비 ( $e_i$ , km/ℓ)	9.5	2.2	4.7	15

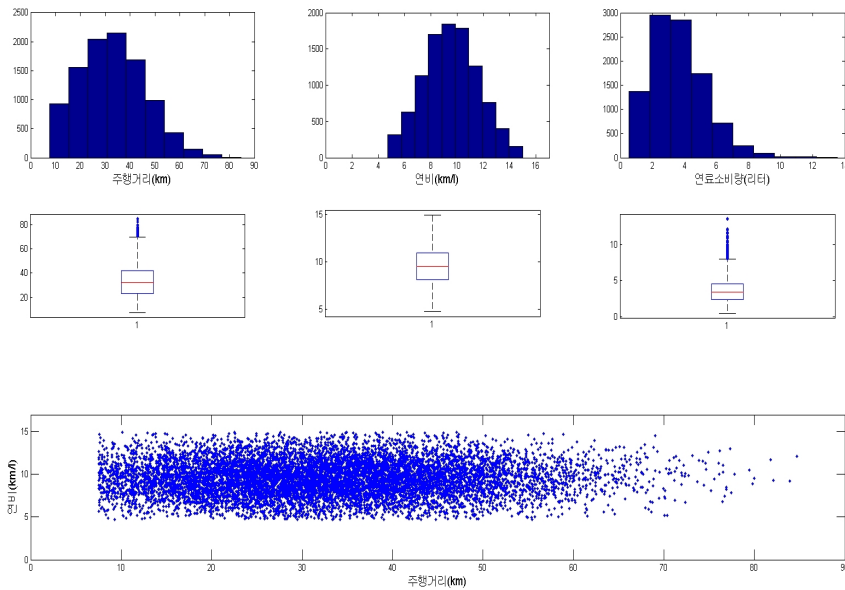


그림 3.1 모집단에 대한 주행거리, 연비 그리고 연료소비량의 형태

그림 3.1은 표 3.1에서 주어진 가정에 의해 설정된 비대칭으로 절단된 정규분포로부터 독립적으로 난수를 발생하여 얻어진  $N = 10000$ 의 주행거리와 연비의 자료 ( $d_i, e_i$ ),  $i = 1, \dots, N$ 와 주행거리와 연비에 의해 계산된 연료소비량의 히스토그램, 상자그림과 산점도이다. 주행거리와 연료소비량은 다소

양의로 비대칭인 분포이고 연비는 정규분포와 유사하다. 또한 주행거리와 연비는 상관성이 거의 없고, 주행거리가 70km 이상 차량들의 연비는 다른 주행거리의 연비에 비해 분산이 적은 편이다. 여기서 총 연료소비량 ( $T$ )은 각 차량의 주행거리와 연비의 역수를 곱해서 얻어진 것으로  $N = 10000$ 대 차량들의 연료소비량을 합한 것이다. 또한 앞에서 언급한 가정에 의해 생성된 주행거리와 연비에 의해 계산된 10,000대 차량들의 연료소비량을 계산하고 합한 총연료소비량 ( $T$ )은 36,146ℓ이다.

### 3.2. 모의실험결과

총연료소비량을 추정하기 위해 제안된 방법 4가지에 대한 모의실험을 통하여 비교분석한다. 표본의 수는 네 가지 ( $n = 30, 100, 500, 1000$ )로 구성했으며 각 표본의 수에서 100번의 반복을 통하여 주어진 방법들에 대해 비교하였다.

표 3.2는 표본 연비에 대한 기술통계량으로 표본수가 증가할 수로 100번 반복한 것의 평균과 표준편차의 범위가 줄어드는 것을 알 수 있다.

표 3.2 100번 반복에 대한 표본 연비

표본수 (표본비율)	평균					표준편차				
	최솟값	평균	중앙값	최댓값	표준편차	최솟값	평균	중앙값	최댓값	표준편차
30 (0.3%)	8.58	9.55	9.59	10.49	0.39	1.45	2.01	1.99	2.60	0.23
100 (1%)	9.06	9.55	9.56	9.94	0.19	1.79	2.06	2.05	2.34	0.12
500 (5%)	9.34	9.52	9.52	9.81	0.09	1.88	2.05	2.05	2.18	0.05
1,000 (10%)	9.38	9.53	9.52	9.69	0.06	1.95	2.05	2.05	2.18	0.04

표 3.3은 총연료소비량 추정에 적용된 방법들을 비교하기 위해 100번 반복을 이용하여 얻어진 결과를 요약한 것이다. 모든 표본들은 확률표본기법인 단순임의추출법으로 추출된 것이다.

<방법 1>은 총연료소비량 ( $\hat{T}_1$ )을 추정하는데 사용된 방법들 중 가장 간단한 방법인 1대당 표본평균 연료소비량에 전체 차량수를 곱한 것이다. 표본수에 따른 범위들이 다른 방법들보다 다소 넓은 편이다.

<방법 2>는 표본으로 추출된 연비로부터 얻은 가중조화 평균연비와 전체 차량의 주행거리의 곱으로 총연료소비량 ( $\hat{T}_2$ )을 추정하는 것이다. 결과를 살펴보면 표본수가 늘어남에 따라 실제 총연료소비량에 근접하고 있다.

<방법 3>은  $n$ 개의 표본 주행거리와 이에 대응되는 연비자료로부터 계산된 연료소비량의 확률밀도함수를 추정한다. 추정된 연료소비량의 확률밀도함수로부터 전체 차량수  $N$ 에 해당되는 연료소비량자료를 생성하여 이들의 총합으로 총연료소비량 ( $\hat{T}_3$ )을 추정하는 것이다. <방법 3>은 다른 방법들에 비해 표본수와 관계없이 그 범위가 다소 넓은 것으로 나왔다.

<방법 4>는 먼저 표본 연비자료의 확률분포를 추정한다. 이 확률분포로부터  $N$ 개의 연비자료를 생성하여 주어진  $N$ 개 차량의 주행거리에 각각 할당하여  $N$ 개의 쌍을 이루게 한다. 이  $N$ 개의 쌍에서 각각 주행거리를 연비로 나누어서  $N$ 개의 연료소비량을 계산한 후 이들의 총합으로 총연료소비량 ( $\hat{T}_4$ )을 추정한다. 표 3.4의 결과에서 주어진 네 번째 방법으로 얻어진 범위들이 다른 방법들보다 다소 폭이 적은 것으로 나왔다.

제시된 방법들을 평가하기 위해 표 3.4와 3.5에서 제시된 범위, 변동계수 그리고 평균제곱오차 (MSE: mean squared error)를 이용하였다. 여기서 평균제곱오차 (MSE)는

$$MSE = \frac{1}{100} \sum_{j=1}^{100} (\hat{T}_{i_j} - T)^2, \quad i \text{는 제시된 방법, } j \text{는 반복수}$$

표 3.3 100번 반복에서 총연료소비량의 추정에 대한 요약통계값

총연료소비량		표본수 (전체 차량의 수 10,000)			
		30	100	500	1,000
방법 1	최 소 값	29,025	31,597	34,117	35,088
	평 균	36,047	35,998	36,223	36,189
	중 앙 값	35,878	36,073	36,273	36,133
	최 대 값	45,084	40,016	37,864	38,082
	표준편차	3,475	1,670	733	589
방법 2	최 소 값	31,384	34,484	35,322	35,567
	평 균	36,064	36,049	36,193	36,185
	중 앙 값	35,995	36,056	36,183	36,167
	최 대 값	40,925	38,141	37,159	36,982
	표준편차	1,714	854	388	272
방법 3	최 소 값	30,146	32,953	35,413	36,284
	평 균	37,787	37,490	37,545	37,491
	중 앙 값	37,430	37,533	37,550	37,416
	최 대 값	47,903	41,270	39,572	39,460
	표준편차	3,727	1,756	794	613
방법 4	최 소 값	32,404	34,377	34,963	35,374
	평 균	35,869	35,908	35,989	35,959
	중 앙 값	35,783	35,935	35,979	35,970
	최 대 값	39,567	37,704	36,692	36,486
	표준편차	1,439	690	326	227

이다. <방법 4>가 주어진 표본수에 관계없이 변동계수가 다른 방법들에 비해 가장 적게 나왔고 범위도 가장 적은 결과를 얻었다. 또한 평균제곱오차는 표본수가 1,000의 경우에 <방법 2>가 가장 적고 나머지 표본수에 대해서는 <방법 4>가 총연료소비량에 가장 가까운 결과를 얻었다.

표 3.4 범위와 변동계수

		표본수 (전체 차량의 수 10,000)			
		30	100	500	1,000
범위	방법 1	16,059	8,419	3,747	2,994
	방법 2	9,541	3,657	1,837	1,415
	방법 3	17,757	8,317	4,159	3,176
	방법 4	7,163	3,327	1,729	1,112
변동계수	방법 1	0.0964	0.0464	0.0202	0.0163
	방법 2	0.0475	0.0237	0.0107	0.0075
	방법 3	0.0986	0.0468	0.0211	0.0164
	방법 4	0.0401	0.0192	0.0091	0.0063

표 3.5 평균제곱오차 (MSE)

		표본수 (전체 차량의 수 10,000)			
		30	100	500	1,000
MSE	방법 1	7.47e+6	3.00e+6	5.14e+5	2.48e+5
	방법 2	2.23e+6	9.23e+5	1.72e+5	6.50e+4
	방법 3	1.37e+7	5.36e+6	2.45e+6	2.04e+6
	방법 4	1.52e+6	6.67e+5	1.53e+5	8.00e+4

#### 4. 결론 및 향후과제

온실가스 국가감축에 관한 연구가 다양한 분야에서 활발히 진행되고 있다. 특히 이동배출원에서 도로수송 부문의 배출량 산정방법은 연료소비량 또는 주행거리를 이용하는 것으로 제시되어왔다 (IPCC, 2006). 지금까지 제시된 방법들은 연료소비량 또는 주행거리의 자료를 이용하여 단지 배출량 산정에만 초점을 맞추어 개발되어왔다. 또한 우리나라의 경우에는 연료판매량 또는 주행거리를 이용하여 국가 또는 지자체별 온실가스 배출량 산정에 대한 연구를 진행해 왔다.

본 연구는 도로수송 배출량을 산정하기 위해 적용되었던 연료판매량 대신에 주행거리 및 연비를 이용하여 연료소비량을 추정하는 방법들을 비교한 것이다. 일부 제시된 방법의 장점은 배출량 산정뿐만 아니라 변동성에 대한 정보도 제공할 수 있다는 것이다. 또한 연료판매량은 전국적인 배출량을 산정하는 것으로만 만족해야 하지만 주행거리와 연비를 이용해서 추정된 연료소비량은 차종, 연식, 지역 등의 다양한 분류로 배출량의 산정뿐만 아니라 변동성에 대한 것도 제공할 수 있다는 장점이 있다.

본 연구에서 제시한 네 가지의 방법들은 통계적인 기법을 기초로 한 것으로 다소 현실성이 떨어질 수 있다. <방법 1>은 표본 연료소비량과 전체 차량수를 이용하여 연료소비량뿐만 아니라 신뢰구간도 제시할 수 있으나 표본수에 따라 정확도가 결정된다. <방법 3>은 몬테카를로 기법을 이용한 것으로 <방법 1>과 마찬가지로 그 정확도는 차이가 없다. <방법 2>는 보편적으로 많이 사용하는 방법으로 총연료소비량의 추정은 가능하나 그 변동성을 나타내는 것은 불가능하다. <방법 4>는 모든 차량의 주행거리가 제공되고 연비의 분포에 따른 난수를 생성하여 주행거리에 할당해서 총연료소비량의 정보를 추출하는 것이다. 이 방법은 가능한 각 차량의 주행거리와 연비의 조합으로 해서 총연료소비량을 추정한 것이다. 이 방법의 장점은 연비의 표본수에 영향을 덜 받고 모의실험의 결과에서 총연료소비량의 변동성이 다른 방법에 비해 적다는 것이다. 하지만 추정된 연비분포가 실제연비분포에서 많이 벗어나는 경우에는 실제 총연료소비량이 포함되지 않는 경우가 발생했다.

결론적으로 <방법 4>가 다른 방법들에 비해 개선의 여지가 있으며, 표본수가 적을 경우에 실제 총연료소비량이 포함되지 않는 것은 여러 연비분포들을 설정하여 총연료소비량을 추정하면 해결될 것으로 판단된다. 즉 모의실험에서는 연비의 분포가 표본평균과 표본표준편차를 가지는 정규분포로 설정되어서 총연료소비량을 추정하였지만 향후에는 현실적인 실정을 기반으로 하여 연비분포의 현실적인 반영을 고려한다면 총연료소비량의 추정뿐만 아니라 변동성도 신뢰할 수 있는 추정이 가능할 것으로 보인다. 또한 연비분포에 대한 추정과 배출계수를 고려하여 배출량의 산정뿐만 아니라 변동성에 대한 연구가 이루어져야 할 것이다. 또한 차종, 연식, 배기량 등과 같은 세세한 분류에 따른 자료확보 및 배출량과 변동성에 대한 연구가 조속히 이루어져야 할 것이다. 본 연구는 여태까지 시도되지 않았던 몬테카를로 기법을 통하여 도로교통부문의 총연료소비량을 추정하는 방법들을 제시하고 비교분석하였다는 것에 의의가 있다. 이러한 연구를 바탕으로 향후에는 보다 정확하고 효율적인 총연료소비량 추정방법들을 개선되어야 하고 그 변동성에 대한 연구도 활발히 진행되길 기대한다.

#### 참고문헌

- 권세혁 (2010). 시뮬레이션 실험조건 이상 진단 연구. <한국데이터정보과학회지>, **21**, 853-861.
- 김기동, 고현기, 이태정, 김동술 (2011). 배출량 산정방법에 따른 지자체 도로수송부문의 온실가스 배출량 산정 비교. <한국대기환경학회지>, **27**, 405-415.
- 국립환경과학원 (2009). <수송부문 온실가스 기후변화대응 시스템 구축(II): 자동차 온실가스 Bottom-up 배출계수 개발>, 국립환경과학원, 인천.
- 에너지경제연구원 (2011). <에너지통계연보>, 에너지경제연구원, 의왕시.
- 최현석, 김태운 (2010). 엑셀 매크로 기능을 이용한 표본추출에 관한 연구. <한국데이터정보과학회지>, **21**, 481-491.



- Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *New S language*, Chapman and Hall/CRC, London.
- Bowman, A. W. and A. Azzalini. (1997). *Applied smoothing techniques for data analysis*, Oxford University Press, New York.
- IPCC (2006). *2006 IPCC guidelines for national greenhouse gas inventories*, IPCC, Switzerland.
- Martinez, W. L. and Martinez, A. L. (2002). *Computational statistics handbook with MATLAB*, Chapman & Hall/CRC, New York.
- Park, C. G. (2011). On statistical properties of some difference-based error variance estimators in nonparametric regression with a finite sample. *Journal of the Korean Data & Information Science Society*, **22**, 575-587.
- Scott, D. W. (1992). *Multivariate density estimation: Theory, practice and visualization*, Wiley, New York.
- Sheather, S. J. and Jones M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society B*, **53**, 683-690.
- Sheather S. J. (2004). Density estimation. *Statistical Science*, **19**, 588-597.
- Shim, J. (2011). Variable selection in the kernel Cox regression. *Journal of the Korean Data & Information Science Society*, **22**, 795-801.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall/CRC, London.
- Venables, W. N. and Ripley, B. D. (2002). *Modern applied statistics with S*, Springer, New York.

## Estimation methods of fuel consumption using distance traveled: Focused on Monte Carlo method<sup>†</sup>

Chun Gun Park<sup>1</sup> · Jinyoung Soh<sup>2</sup> · Yung-Seop Lee<sup>3</sup>

<sup>1</sup>Department of Mathematics, Kyonggi University

<sup>2</sup>Korea Energy Economics Institute

<sup>3</sup>Department of Statistics, Dongguk University

Received 3 February 2012, revised 24 February 2012, accepted 2 March 2012

### Abstract

Recently, estimation of greenhouse gas (GHG) emission has continuously emerged as an important global issue. This study compares various statistical methods for estimation of fuel consumption, which is necessary for calculation of GHG emission in road transportation sector. Existing methods have focused on using merely transportation fuel supply or distance traveled for calculation of fuel consumption. Estimates of GHG emission based on fuel supply, however, cannot reflect various vehicle types or model year. This study suggests and compares, from statistical point of view, several methods, which can be applied to estimate fuel consumption of each vehicle, by combining distance traveled and fuel efficiency (mileage), and total fuel consumption of all vehicles. It also suggests practical measures that can reflect vehicle types and model year to suggested methods for future research.

*Keywords:* Distance traveled, fuel consumption, fuel efficiency, kernel density estimation, Monte Carlo.

---

<sup>†</sup> This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0009273).

<sup>1</sup> Assistant professor, Department of Mathematics, Kyonggi University, Suwon, Gyeonggi-do 443-760, Korea.

<sup>2</sup> Researcher, Korea Energy Economics Institute, 665-1 Naeson-dong, Euiwang, Kyunggi-do 437-713, Korea.

<sup>3</sup> Corresponding author: Professor, Department of Statistics, Dongguk University, 26 Phil-dong 3-ga, Jung-gu, Seoul 100-715, Korea. Email: yung@dongguk.edu