

# 지하역사의 미세먼지 농도 예측을 위한 소프트센서(soft sensor) 모델 개발

## Soft Sensing Model for Estimating Particulate Matters at Underground Subway Station

김민정<sup>\*†</sup>, 박덕신<sup>\*</sup>, 문대섭<sup>\*\*</sup>

Minjeong Kim<sup>\*†</sup>, Duckshin Park<sup>\*</sup>, Dae Seop Moon<sup>\*\*</sup>

**Abstract** An objective of this research is to develop a soft sensor for estimating particulate matter (PM) concentration at underground subway stations. The soft sensor infers difficult-to-measure variable based on information of easy-to-measure variables. To develop the soft sensor, an independent component regression (ICR) is used, where the ICR is an optimal technique to construct a linear regression model by extracting essential information from multivariate data. Experimental results in S-subway station show that the proposed method can improve the performance for estimating PM<sub>10</sub> concentration at subway station as compared to the conventional soft sensor which is developed using principal component regression (PCR).

**Keywords :** PM<sub>10</sub> estimation, Subway, Indoor air quality, Soft sensor, Independent component regression

**초 록** 본 연구는 지하역사 내 미세먼지(PM<sub>10</sub>)의 농도 예측을 위한 소프트센서 개발을 목적으로 하였다. 소프트센서란 측정이 용이한 변수들의 정보를 이용하여 측정이 어려운 변수를 추론하는 방법이다. 지하역사 내 PM<sub>10</sub> 소프트센서의 개발을 위해 본 연구는 독립성분회귀(independent component regression)를 사용하였는데 이는 다변량 데이터의 차원 축소를 통해 주요한 정보를 추출하며 이를 바탕으로 선형 회귀모형을 구축하는 통계기법이다. 개발된 소프트센서를 서울 S-역사 공기질 자료에 적용한 결과 주성분회귀(principal component regression)를 사용한 기존 소프트센서에 비해 지하역사 내 PM<sub>10</sub> 농도를 정확도 높게 예측함을 확인할 수 있었다.

**주요어 :** 미세먼지 예측, 지하역사, 실내공기질, 소프트센서, 독립성분회귀

### 1. 서론

지하철은 급격한 도시화에 따른 인구 과밀, 차량 증가 등에 의한 지상개발의 한계를 보완하기 위해 개발된 대중교통 중 하나이다. 현재 서울시 지하철의 교통수단 분담률은 서울시 전체 교통수단의 약 39%로 많은 유동인구가 지하철을 이용하고 있으나 지하공간의 폐쇄성, 내부 오염물질 발생으로 인해 입자상 물질(particulate matters, PM)의 오염도가 높은 것으로

† 교신저자: 한국철도기술연구원 교통환경연구팀(mjkim88@krri.re.kr)

\* 한국철도기술연구원 교통환경연구팀

\*\* 한국철도기술연구원 녹색교통물류시스템공학연구소

보고되고 있다 [1,2]. 입자상 물질은 호흡기계 질병, 심혈관계 질병을 야기하므로 지하철 이용객과 작업자의 건강 보호를 위해 지하철 내 입자상 물질의 정확한 모니터링 및 예측이 매우 중요하다.

측정이 용이한 변수들의 정보를 이용하여 측정이 어려운 변수를 추론하는 소프트센서(soft sensor)는 지하공간 내 공기질을 예측하기 위해 널리 사용되는 방법 중 하나이다 [3]. 입자상 물질의 경우 입자 포집 시간이 필요하여 실시간 측정이 어려우므로 소프트센서는 지하철 내 입자상 물질의 실시간 농도를 예측하는 대안이 될 수 있다. 이에 본 연구는 독립성분회귀(independent component regression, ICR)를 이용하여 지하역사 내 미세먼지( $PM_{10}$ ) 농도를 예측하는 소프트센서를 개발하였다.

## 2. 연구 방법

### 2.1 독립성분회귀

독립성분회귀(ICR)란 비정규 분포를 따르는 다변량 데이터의 차원 축소를 통해 주요한 정보(independent component, IC)를 추출하여 선형 회귀모형을 구축하는 통계기법이다 [4]. ICR 알고리즘은 입력 데이터 행렬  $\mathbf{X}$ 를 식(1)과 같이 분해한다.

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

여기서  $\mathbf{A}$ 는 혼합 행렬(mixing matrix)을,  $\mathbf{S}$ 는 독립성분 행렬(independent component matrix)을 의미한다. 출력 변수  $\mathbf{Y}$ 와 독립성분 행렬  $\mathbf{S}$  간의 ICR 모델은 식(2)와 같이 선형 모델로써 표현될 수 있다 [4].

$$\mathbf{Y} = \mathbf{S} \cdot \boldsymbol{\beta} + \mathbf{E} \quad (2)$$

여기서  $\boldsymbol{\beta}$ 는 회귀계수를,  $\mathbf{E}$ 는 잔차 행렬을 의미한다. 회귀계수  $\boldsymbol{\beta}$ 는 최소자승법(least square method)을 통해 계산되는데, 최소자승법이란 실제 값과 예측 값의 오차 제곱의 합이 최소가 되는 해를 구하는 방법이다 [4].

### 2.2 연구 방법

본 연구는 ICR 알고리즘 기반의 소프트센서 개발을 위해 서울시 S-역사의 데이터를 사용하였다. ICR 알고리즘을 데이터에 적용하기 앞서 이상치 제거, 결측치 채움 등 데이터의 전처리를 수행하였다. 그 후 전처리된 데이터에 ICR 알고리즘을 적용하여 소프트센서 모델을 개발하였다. 본 연구의 목적은 지하역사 내  $PM_{10}$  농도를 예측하는 것이므로 소프트센서 모델의 출력 변수( $\mathbf{Y}$ )로 지하역사의  $PM_{10}$  농도를 설정하였다. 소프트센서 모델의 입력 변수( $\mathbf{X}$ )로는 역사 내 공기질 측정자료(온도, 습도, 일산화질소, 이산화질소, 일산화탄소, 이산화탄소 농도)를 사용하였다. 개발된 ICR 기반 소프트센서의 타당성을 검증하기 위해 주성분회귀(principal component regression, PCR)를 사용한 기존 소프트센서와의 예측 정확도 비교를 수

행하였다. 예측 정확도의 정량적 비교를 위해서는 평균 제곱근 오차(root mean squared error, RMSE)를 사용하였다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{i,observed} - \hat{y}_{i,modeled})^2}{n-1}} \quad (3)$$

여기서  $y_{i,observed}$  는 실제 측정값을,  $\hat{y}_{i,modeled}$  는 소프트센서 모델을 통해 예측된 값을 의미한다.

### 3. 결과 및 고찰

Fig. 1은 소프트센서 모델을 통해 예측된 지하철역사 내  $PM_{10}$  농도를 보여준다. 여기서 O는 실제 지하철역사 내  $PM_{10}$  농도를, 실선은 ICR 알고리즘 기반 소프트센서의 예측값을, 점선은 PCR 알고리즘 기반 소프트센서의 예측값을 나타낸다. 본 그림을 통해서 두 소프트센서의 예측 정확도를 비교하기 어려움으로 RMSE를 이용하여 예측 정확도의 정량적 비교를 수행하였다. RMSE란 실제 측정값과 예측값의 오차를 나타내는 지표로써 그 값이 작을수록 높은 정확도를 의미한다. ICR 기반 소프트센서와 PCR 기반 소프트센서의 RMSE 값은 23.5, 25.4로 ICR 기반 소프트센서가 지하철역사 내  $PM_{10}$  농도를 예측함에 있어 신뢰성 높음을 확인할 수 있다. 이는 PCR 기반 소프트센서가 지하철역사 공기질 자료를 정규분포라 가정하여 예측 모델을 구축한 반면, ICR 기반 소프트센서는 지하철역사 공기질 자료의 비정규적 분포 특성을 고려하여 예측 모델을 구축했기 때문으로 사료된다.

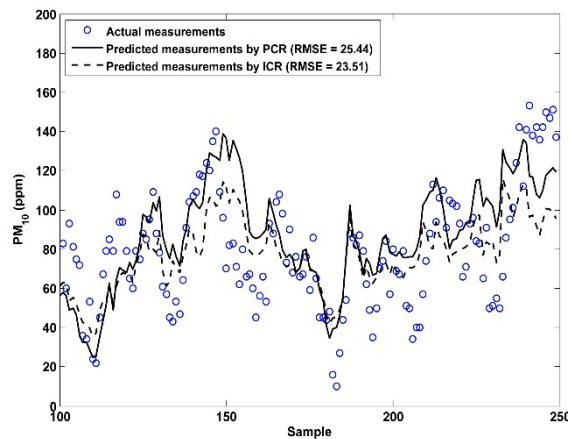


Fig. 1 Estimation of the  $PM_{10}$  concentration at subway station using two soft sensing models

### 사 사

본 연구는 국토교통부의 재원으로 미래철도기술연구사업(17RTRP-B082486-04)의 연구비 지원을 받아 수행되었으며, 이에 감사 드립니다.

## 참고문헌

- [1] T.-J. Lee, H. Lim, S.-D. Kim, D.-S. Park, D.-S. Kim (2015) Concentration and properties of particulate matters ( $PM_{10}$  and  $PM_{2.5}$ ) in the Seoul metropolitan, *J. Korean Soc. Atmos. Environ.*, 31(2), pp. 164-172.
- [2] E.-S. Lee, M.-B. Park, T.-J. Lee, S.-D. Kim, D.-S. Park, D.-S. Kim (2016) Characterizing particle matter on the main section of the Seoul subway line-2 and developing fine particle pollution map, *J. Korean Soc. Atmos. Environ.*, 32(2), pp. 216-232.
- [3] P. Kadlec, R. Grbic, B. Gabrys (2011) Review of adaptation mechanisms for data-driven soft sensor, *Computers and Chem. Eng.*, 35(1), pp. 1-24.
- [4] X. Shao, W. Wang, Z. Hou, W. Cai (2006) A new regression method based on independent component analysis, *Talanta*, 69(3), pp. 676-680.