Deep Particulate Matter Forecasting Model Using Correntropy-Induced Loss

Jongsu Kim [1] and Changhoon Lee[2*]

[1,2] *Yonsei University, Seoul 03722, Republic of Korea*

Corresponding author (Electronic mail: clee@yonsei.ac.kr)

Forecasting the particulate matter (PM) concentration in South Korea has become urgently necessary owing to its strong negative impact on human life. In most statistical or machine learning methods, independent and identically distributed data, for example, a Gaussian distribution, are assumed; however, time series such as air pollution and weather data do not meet this assumption. In this study, the detrended fluctuation analysis and power-law analysis are used in an analysis of the statistical characteristics of air pollution and weather data. Rigorous seasonality adjustment of the air pollution and weather data was performed because of their complex seasonality patterns and the heavy-tailed distribution of data even after deseasonalization. The maximum correntropy criterion for regression (MCCR) loss was applied to multiple models including conventional statistical models and state-of-the-art machine learning models. The results show that the MCCR loss is more appropriate than the conventional mean squared error loss for forecasting extreme values.