Air Quality Forecasting Using Big Data and Machine Learning Algorithms

Youn-Seo Koo[1] · Yunsoo Choi[2] · Chang-Hoi Ho[3,4]

Traditionally, air quality forecasting has primarily relied on physiochemical models, specifically the chemical transport model. These numerical models, however, encounter challenges stemming from structural constraints, variations in meteorological data, emission inventories, and intrinsic model limitations. Machine learning provides a promising alternative to mitigate the limitations of traditional physiochemical models. However, their effectiveness is contingent on their ability to accurately capture the complex dynamics of pollutant transport, including long-range transport of pollutants and the secondary formation of air pollutants through intricate chemical reactions in the atmosphere (Koo et al. 2023). Recognizing this, the special issue compiles a wide range of research that capitalizes on both machine learning and big data for air quality forecasting. The evidence suggests that machine learning can improve the accuracy of forecasts by reducing the biases present in physiochemical models. With an objective to underscore the potential of big data in shedding light on the relationship between emission sources and receptors, and to emphasize the progress made in machine learning algorithms, this special issue features six significant contributions that portray a wide application of big data and machine learning algorithms.

1. Kim and Park (2023) highlight the pivotal importance of selecting the most effective input variables for artificial intelligence (AI) models to improve the forecasting of high $PM_{2.5}$ concentration events. Their study encompassed Seoul and four Chinese cities and identified specific factors including the east–west geopotential index, the Korean region blocking index, and concentration-wind, underscoring their significant role in enhancing the accuracy of predictions in AI-based air quality models.

2. Choi et al. (2023) point out an interesting pattern of declining wintertime air quality on weekends in Seoul, attributing this phenomenon to the formation of secondary particles and the presence of external pollutants. Their work emphasizes the necessity for accurate regional emission calculations and pinpointing unknown emission sources.

3. Ejurothu et al. (2023) introduce a cluster-based local hybrid-graph neural network for predicting $PM_{2.5}$ concentrations across India. By incorporating local weather variations into their approach, their methodology demonstrates a substantial enhancement in performance and computational efficiency compared to preceding approaches.

4. Ho et al. (2023) assess the performance of the long short-term memory (LSTM) algorithm in predicting $PM_{2.5}$ grades across 19 districts in the Republic of Korea. Their results indicate that LSTM forecasts are on par with the current AirKorea system's capabilities, with the added benefit of minimizing false alarms relative to the Community Multiscale Air Quality-only approach.

5. Koo et al. (2023) present an innovative $PM_{2.5}$ forecasting system for the Republic of Korea, utilizing advanced machine learning models. This system is specifically designed to account for long-range transport phenomena and provides forecasts for the 19 districts, predicting $PM_{2.5}$ concentrations for the next 48 h. Four algorithms were integrated into the system: a deep neural network, a recurrent neural network, a convolutional neural network, and an ensemble approach. Real-time evaluations demonstrate improved prediction accuracy and better agreement with observations.

✉ Youn-Seo Koo
koo@anyang.ac.kr

[1] Department of Environmental & Energy Engineering, Anyang University, Gyeonggi-do, Republic of Korea

[2] Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA

[3] School of Earth and Environmental Sciences, Seoul National University, Seoul, Republic of Korea

[4] Department of Climate and Energy Systems Engineering, Ewha Womans University, Seoul 03760, Republic of Korea

6. Lops et al. (2023) introduce a pioneering method for forecasting the El Niño-Southern Oscillation (ENSO) employing an ensemble of deep convolutional neural networks. Their ensemble model outperformed individual models, furnishing accurate ENSO predictions up to three years in advance, signifying a considerable breakthrough in climate forecasting.

Within this continually evolving research field, where methodologies are constantly being refined to create big data and machine learning algorithms that accurately represent environmental dynamics, insights from this special edition have the potential to ignite further investigations. The editorial team is excited to see how the contributions within this edition may spur the application of cutting-edge techniques in air quality forecasting, thereby further enriching our understanding of this crucial environmental issue.

References

Choi, W., Ho, C.-H., Heo, J.-W., Kim, K.-Y., Kim, and S.-W., Kim. J.: Recent air quality deterioration on weekends in Seoul, South Korea: A focus on external contribution. Asia-Pac. J. Atmos. Sci. (2023). https://doi.org/10.1007/s13143-022-00287-0

Ejurothu, P.S.S., Mandal, S., and Thakur, M.: Forecasting $PM_{2.5}$ concentration in India using a cluster based hybrid graph neural network approach. Asia-Pac J Atmos Sci (2023). https://doi.org/10.1007/s13143-022-00291-4

Ho, C.-H., Park, I, Kim, J., and Lee, J.-B.: $PM_{2.5}$ forecast in Korea using the long short-term memory (LSTM) model. Asia-Pac. J. Atmos. Sci. (2023). https://doi.org/10.1007/s13143-022-00293-2

Koo, Y.-S., Kwon, H.-Y., Bae, H., Yun, H.-Y., Choi, D.-R., Yu, S., Wang, K.-H., Koo, J.-S., Lee, Jae-Bum, Choi, M.-H., and Lee, Jeong-Beom: A development of $PM_{2.5}$ forecasting system in South Korea using chemical transport modeling and machine learning. Asia-Pac. J. Atmos. Sci. (2023). https://doi.org/10.1007/s13143-023-00314-8

Lops, Y., Choi, Y., Mousavinezhad, S., Salman, A.K., Nelson, D.L., Singh, D.: Development of deep convolutional neural network ensemble models for 36-month ENSO forecasts. Asia-Pac. J. Atmos. Sci. (2023). https://doi.org/10.1007/s13143-023-00319-3

Kim, S.-H. and Park, M.-S.: Determination of input variables for artificial intelligence models to predict the high $PM_{2.5}$ concentration events in Seoul, Korea. Asia-Pac. J. Atmos. Sci. (2023). https://doi.org/10.1007/s13143-023-00333-5