



Regional Air Quality Management: A Scalable, Data-Driven Airshed Framework using Low-Cost Sensors across the Indo-Gangetic Plain, India

Anandh P Chandrasekaran¹, Sachchida Nand Tripathi^{1,2}, Nimit Godhani¹, Malay Pandey³, Piyush Rai³, Navdeep Agrawal¹, Anil Kumar¹, and Snehaddeep Ballav¹

¹Centre for Excellence- Advanced Technologies for Monitoring Air-quality iNdicators (CoE – ATMAN), National Aerosol Facility, Indian Institute of Technology Kanpur, India (sivaanandh181@gmail.com)

²Department of Civil Engineering and Department of Sustainable Energy Engineering, Indian Institute of Technology Kanpur, India

³Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India

In India, ensuring clean air for all is vital and should not be limited to urbanites. However, the current air quality monitoring networks and clean air strategies are limited to cities. Notably, the air quality status across regions is yet to be measured, and comprehensive regional management plans are non-existent. To address these significant research gaps, the Ambient air quality Monitoring over Rural areas using Indigenous Technology (AMRIT) project was envisioned and implemented by the Centre of Excellence Advanced Technologies for Monitoring Air-quality iNdicators (CoE – ATMAN), Indian Institute of Technology Kanpur. For the first time, over 1,400 low-cost PM_{2.5} sensors were installed across the states of Uttar Pradesh and Bihar at the block level. The sensor locations encompass a diverse range of land-use and land-cover categories, demographics, and communities. By leveraging a dense network of low-cost sensors, we developed a data-driven machine learning framework to delineate airsheds for regional air quality management for the first time. We utilized a recurrent neural network-based long short-term memory (LSTM) and hierarchical clustering algorithms with PM_{2.5} and meteorological data to delineate airsheds. The LSTM embeddings learn latent representations from PM_{2.5}-ventilation coefficient (VC) time series, capturing spatiotemporal patterns and inter-variable relationships. These embeddings were then hierarchically clustered to delineate airsheds. Applying this framework, our results for Bihar show five distinct airsheds; three prevail in the north of the Ganges River, and two prevail in the south of the Ganges. Notably, Airshed 1, located in the northwest region, is highly polluted. However, during the post-monsoon and winter across the airsheds, PM_{2.5} levels were two to three times higher than the national standard (60 µg/m³), and on ~90% of days, people breathe unhealthy air. Similarly, we identified multiple distinct airsheds in Uttar Pradesh, as well as common airsheds that prevail across Bihar and Uttar Pradesh states. This emphasizes not only shared regional influences but also the need for an integrated approach to reduce PM_{2.5} pollution. Therefore, the identified airsheds would be instrumental in targeting the reduction of fine particulate pollution across the Bihar and Uttar Pradesh states. Furthermore, the scalable, data-driven airshed delineation framework using low cost sensors could be implemented

across India and potentially globally. Thus, this will facilitate airshed-based air quality management plans and integrated policy interventions to ensure “clean air for all.”

Keuwords: Air Quality; Low-Cost Sensors; Airshed; Indo-Gangetic Plain; Machine Learning