Developing and Evaluating an Integrated Machine Learning Approach to Estimating Near Real-Time Local Level Concentrations of Outdoor Ultrafine Particles using Street-Level Images and City Sounds

Scott Weichenthal*, Alessya Venuta,[1] Marshall Lloyd,[1] Shaainee Ramnawaz,[1] Marianne Hatzopoulou,[2] Doug Petican,[3] Blake Laufer[4]
[1] McGill University, Department of Epidemiology, Biostatistics, and Occupational Health, Montreal, Quebec, Canada
[2] University of Toronto, Department of Civil and Mineral Engineering, Toronto, Ontario, Canada
[3] Consultant, Hamilton, Ontario, Canada
[4] Mistall Insight Inc, Hamilton, Ontario, Canada
*scott.weichenthal@mcgill.ca

Existing epidemiological evidence suggests that outdoor ultrafine particles (UFPs, < 100 nm) have adverse impact on human health. However, most existing exposure modelling studies for outdoor UFPs have focused on predicting long-term exposures (e.g., annual average) and methods of predicting near real-time local level exposures are not readily available. To address this need, we developed a new platform including a solar-powered field monitoring device and machine learning algorithms that together provide near real-time predictions of outdoor UFPs and noise based on street-level images, audio data, and regional weather/air quality data. Machine learning models (convolutional neural networks and XGboost) were developed and evaluated using a large database of image/audio samples labelled with UFP/noise measurements (10-second samples) collected with research-grade instruments from 11 locations across Montreal, Canada, between 2021-2022.

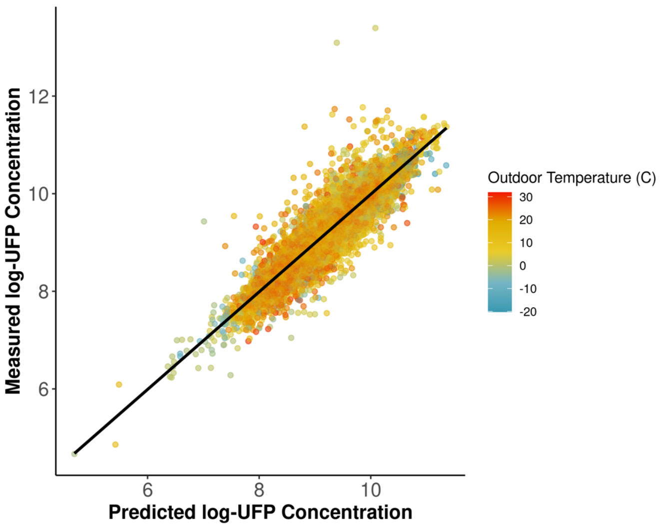

Figure 1. Measured vs. Predicted UFPs

In total, our new UFP model was based on approximately 120,000 samples and our noise model was based on approximately 99,000 samples. Outdoor UFP (median: 10,181 particles/cm^3; 1st: 1898, 99th: 51,390) and noise levels (median: 63.4 dB(A); 1st: 46.1, 99th: 80.3) varied substantially across our study period and reflect a wide range of weather conditions including very winter days (< -10 °C) warm summer days (>25°C) periods. When evaluated in random test sets (15,000 samples for UFPs and 12,000 samples for noise), our algorithms explained the majority of local-level variations in outdoor UFPs (R^2=0.85) and noise (R^2=0.86) with slopes close to one between measured and predicted values (see Figure 1 for UFP results).

Collectively, this new platform offers an efficient means of predicting local-level UFP and noise levels in near real-time and can be used to track outdoor concentrations over-time in locations of concern across cities. Ongoing work is focussed on the transportability of these models (i.e., developing models in one set of locations and testing them in completely different locations) as well as extending this method for use in occupational exposure environments.