Literature Review PM2.5 Models

In recent years, geographic and environmental data has become increasingly accessible, with new satellites, sensors, and open-data initiatives. This has created an explosion of ever-more accurate and higher resolution models depicting the variation of particulate matter with an aerodynamic diameter under 2.5 micrometers (PM2.5).

Ross et al. (2007) constructed a Land Use Regression (LUR) model from traffic, land use, population, and emissions data that was able to explain over 60% of PM2.5 variation in the New York area during the examined time. Despite outperforming kriging, the model was found to be susceptible to unusual values in variables such as traffic, thereby limiting its overall utility for potential intracity use.

An invaluable source of satellite data to assess PM2.5 across various spatial and temporal scales has been NASA's Moderate Resolution Imaging Spectroradiometer (MODIS). The Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm, put forth by Lyapustin et al. (2011), dramatically improved the spatial resolution of aerosol optical depth (AOD) data from 10 km to 1 km. This improvement has resulted in MAIAC's use in numerous subsequent attempts to model PM2.5 at fine spatial scale, with Kloog et al. (2014) proposing mixed models with MAIAC and sensor data able to generate maps at both 1km and finer 200m spatial resolutions. Hu et al. (2014) validated MAIAC's ability to predict PM2.5 at the 1km scale, finding that both MAIAC and MODIS predicted the particulate matter at similar levels of accuracy, but with the former in a much higher spatial resolution.

Since MAIAC's publication, many have moved towards building hybrid models, employing both satellite AOD data as well some combination of sensor data, land use data, meteorological data, and more. Chudnovsky et al. (2015) presented a mixed-effects model using land use and meteorological variables that successfully predicted elevated levels of PM2.5 along major roads and in urban areas. The paper highlighted the potential of this mixed-effects model in modelling intra-city PM2.5. A Chemical Transport Model (CTM) was used by Donkelaar et al. (2015) in conjunction with geographic-weighted regression in order to examine bias in PM2.5 predictions and prevent overfitting. Xiao et al. (2017) also used a CTM to develop a Multiple Imputation (MI) method of filling in missing MAIAC data, due to the non-random nature of the missing data. Though the method underpredicted PM2.5, it did present a new alternative to filling in the data in places that might have regular long gaps due to cloud or snow cover.

The last few years have seen the advent of neural networking and machine learning as a whole's use in PM2.5 modelling. Di et al. (2016) proposed a neural network-based hybrid model that used AOD, Absorbing Aerosol Index (AAI), CTM outputs, land use terms, and meteorological variables in conjunction. They found that the hybrid model was more accurate than any one model individually. Lin et al. (2017) proposed using OpenStreetMap (OSM) in conjunction with EPA monitoring stations in order to assign a geographic context to each point. This geographic context consisted of a matrix for each monitoring station containing information on every OSM feature within buffers of different radii around the station. They used random forest ensemble learning in order to determine the importance of various features in their impact on PM2.5. The OSM model had a comparable prediction ability to Inverse Distance Weighting (IDW) but gave predictions at a substantially finer scale and, in doing so, was able to predict various areas of heightened PM2.5 that IDW could not.

While much progress has been made with PM2.5 modelling, particularly with the improvement of available satellite data, there is still work to be done. Existing methods largely neglect addressing intra-city PM2.5 variation. Many studies are examining the spatial variation on the scale of counties, states, or even whole countries. Sparse sensor data severely hinders the ability to further improve the spatial or temporal resolution of existing models. Nonetheless, the advancements that have been made are quite promising in the continued improvement of modelling spatial and temporal variation in PM2.5 at an ever finer level.

Chudnovsky, Alexandra A., Petros Koutrakis, Itai Kloog, Steven Melly, Francesco Nordio, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. “Fine Particulate Matter Predictions Using High Resolution Aerosol Optical Depth (AOD) Retrievals.” *Atmospheric Environment* 89 (June 2014): 189–98. https://doi.org/10.1016/j.atmosenv.2014.02.019.

Clougherty, Jane E, Iyad Kheirbek, Holger M Eisl, Zev Ross, Grant Pezeshki, John E Gorczynski, Sarah Johnson, Steven Markowitz, Daniel Kass, and Thomas Matte. “Intra-Urban Spatial Variability in Wintertime Street-Level Concentrations of Multiple Combustion-Related Air Pollutants: The New York City Community Air Survey (NYCCAS).” *Journal of Exposure Science & Environmental Epidemiology* 23, no. 3 (May 2013): 232–40. https://doi.org/10.1038/jes.2012.125.

Di, Qian, Itai Kloog, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. “Assessing PM $_{2.5}$ Exposures with High Spatiotemporal Resolution across the Continental United States.” *Environmental Science & Technology* 50, no. 9 (May 3, 2016): 4712–21. https://doi.org/10.1021/acs.est.5b06121.

Donkelaar, Aaron van, Randall V. Martin, Robert J. D. Spurr, and Richard T. Burnett. “High-Resolution Satellite-Derived PM $_{2.5}$ from Optimal Estimation and Geographically Weighted Regression over North America.” *Environmental Science & Technology* 49, no. 17 (September 2015): 10482–91. https://doi.org/10.1021/acs.est.5b02076.

Hu, Xuefei, Lance A. Waller, Alexei Lyapustin, Yujie Wang, Mohammad Z. Al-Hamdan, William L. Crosson, Maurice G. Estes, et al. “Estimating Ground-Level PM2.5 Concentrations in the Southeastern United States Using MAIAC AOD Retrievals and a Two-Stage Model.” *Remote Sensing of Environment* 140 (January 2014): 220–32. https://doi.org/10.1016/j.rse.2013.08.032.

Kheirbek, Iyad, Sarah Johnson, Zev Ross, Grant Pezeshki, Kazuhiko Ito, Holger Eisl, and Thomas Matte. “Spatial Variability in Levels of Benzene, Formaldehyde, and Total Benzene, Toluene, Ethylbenzene and Xylenes in New York City: A Land-Use Regression Study.” *Environmental Health* 11, no. 1 (December 2012). https://doi.org/10.1186/1476-069X-11-51.

Kloog, Itai, Alexandra A. Chudnovsky, Allan C. Just, Francesco Nordio, Petros Koutrakis, Brent A. Coull, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. “A New Hybrid Spatio-Temporal Model for Estimating Daily Multi-Year PM2.5 Concentrations across Northeastern USA Using High Resolution Aerosol Optical Depth Data.” *Atmospheric Environment* 95 (October 2014): 581–90. https://doi.org/10.1016/j.atmosenv.2014.07.014.

Lin, Yijun, Dimitrios Stripelis, Yao-Yi Chiang, José Luis Ambite, Rima Habre, Fan Pan, and Sandrah P. Eckel. “Mining Public Datasets for Modeling Intra-City PM2.5 Concentrations at a Fine Spatial Resolution.” *Proceedings of the ... ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems: ACM GIS. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 2017 (November 2017). https://doi.org/10.1145/3139958.3140013.

Lyapustin, A., Y. Wang, I. Laszlo, R. Kahn, S. Korkin, L. Remer, R. Levy, and J. S. Reid. “Multiangle Implementation of Atmospheric Correction (MAIAC): 2. Aerosol Algorithm.” *Journal of Geophysical Research* 116, no. D3 (February 11, 2011). https://doi.org/10.1029/2010JD014986.

Matte, Thomas D, Zev Ross, Iyad Kheirbek, Holger Eisl, Sarah Johnson, John E Gorczynski, Daniel Kass, Steven Markowitz, Grant Pezeshki, and Jane E Clougherty. "Monitoring Intraurban Spatial Patterns of Multiple Combustion Air Pollutants in New York City: Design and Implementation." *Journal of Exposure Science & Environmental Epidemiology*23, no. 3 (May 2013): 223–31. https://doi.org/10.1038/jes.2012.126.

Ross, Z, M Jerrett, K Ito, B Tempalski, and G Thurston. "A Land Use Regression for Predicting Fine Particulate Matter Concentrations in the New York City Region." *Atmospheric Environment*41, no. 11 (April 2007): 2255–69. https://doi.org/10.1016/j.atmosenv.2006.11.012.

Ross, Zev, Kazuhiko Ito, Sarah Johnson, Michelle Yee, Grant Pezeshki, Jane E Clougherty, David Savitz, and Thomas Matte. "Spatial and Temporal Estimation of Air Pollutants in New York City: Exposure Assignment for Use in a Birth Outcomes Study." *Environmental Health*12, no. 1 (December 2013). https://doi.org/10.1186/1476-069X-12-51.

Stafoggia, Massimo, Joel Schwartz, Chiara Badaloni, Tom Bellander, Ester Alessandrini, Giorgio Cattani, Francesca de' Donato, et al. "Estimation of Daily PM10 Concentrations in Italy (2006–2012) Using Finely Resolved Satellite Data, Land Use Variables and Meteorology." *Environment International*99 (February 2017): 234–44. https://doi.org/10.1016/j.envint.2016.11.024.

Xiao, Qingyang, Yujie Wang, Howard H. Chang, Xia Meng, Guannan Geng, Alexei Lyapustin, and Yang Liu. "Full-Coverage High-Resolution Daily PM 2.5 Estimation Using MAIAC AOD in the Yangtze River Delta of China." *Remote Sensing of Environment*199 (September 2017): 437–46. https://doi.org/10.1016/j.rse.2017.07.023.