The Canadian Optimized Statistical Smoke Exposure Model (CanOSSEM): A machine learning approach to estimate national daily fine particulate matter ($PM_{2.5}$) exposure

Authors

Naman Paul [1, 2], Jiayun Yao [1], Kathleen E. McLean [1], David M. Stieb [3], Sarah B. Henderson [1, 2]

Affiliations

[1] Environmental Health Services, British Columbia Centre for Disease Control (BCCDC), Vancouver, Canada
[2] School of Population and Public Health, The University of British Columbia, Vancouver, Canada
[3] Population Studies Division, Health Canada, Vancouver, Canada

ABSTRACT

Biomass smoke exposure has been associated with a wide range of acute and chronic health outcomes. Over the past decades, the frequency and intensity of wildfires has increased in many regions, resulting in longer smoke episodes with higher concentrations of fine particulate matter ($PM_{2.5}$). There are also many communities where seasonal open burning and residential wood heating have short- and long-term impacts on ambient air quality. Understanding the acute and chronic health effects of biomass smoke exposure requires reliable estimates of $PM_{2.5}$ concentrations during the wildfire season and throughout the year, particularly in areas without regulatory air quality monitoring stations. We have developed a machine learning approach that estimates daily mean (24-hour) $PM_{2.5}$ concentrations across populated areas of Canada at a 5 km × 5 km spatial resolution between 2010 and 2019. The random forest model integrates $PM_{2.5}$ observations from the National Air Pollution Surveillance (NAPS) network with data from multiple sources including smoke plumes traced from remote sensing imagery, satellite estimates of meteorological parameters, measurements of fire radiative power and aerosol optical depth. The Root Mean Squared Error (RMSE) between predicted and observed $PM_{2.5}$ concentrations was 2.96 $\mu g/m^3$ for the entire prediction set, and more than 96% of the predictions were within 5 $\mu g/m^3$ of the NAPS $PM_{2.5}$ measurements. The model was evaluated using 10-fold, leave-one-region-out, and leave-one-year-out cross-validations. Overall, CanOSSEM performed well but performance was sensitive to removal of large wildfire events such as the Fort McMurray interface fire in May 2016 or the extreme 2017 and 2018 wildfire seasons in British Columbia. Exposure estimates from CanOSSEM will be useful for epidemiologic studies on the acute and chronic health effects associated with $PM_{2.5}$ exposure, especially for populations affected by biomass smoke where routine air quality measurements are not available.