

EGU2020-2693

<https://doi.org/10.5194/egusphere-egu2020-2693>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## Input-adaptive proxy of air quality parameters: A case study for black carbon in Helsinki, Finland

**Pak L Fung**<sup>1</sup>, Martha A Zaidan<sup>1</sup>, Salla Sillanpää<sup>1</sup>, Anu Kousa<sup>2</sup>, Jarkko V Niemi<sup>2</sup>, Hilikka Timonen<sup>3</sup>, Joel Kuula<sup>3</sup>, Erkka Saukko<sup>4</sup>, Krista Luoma<sup>1</sup>, Tuukka Petäjä<sup>1</sup>, Sasu Tarkoma<sup>5</sup>, Markku Kulmala<sup>1</sup>, and Tareq Hussein<sup>1,6</sup>

<sup>1</sup>Institute for Atmospheric and Earth System Research (INAR)/Physics, University of Helsinki, FI-00560 Helsinki, Finland (pak.fung@helsinki.fi; martha.zaidan@helsinki.fi; salla.sillanpaa@helsinki.fi; krista.q.luoma@helsinki.fi;

tuukka.petaja@helsinki.fi; ma

<sup>2</sup>Helsinki Region Environmental Services Authority (HSY), PO Box 100, FI-00066 HSY, Finland (anu.kousa@hsy.fi;

jarkko.niemi@hsy.fi)

<sup>3</sup>Atmospheric Composition Research, Finnish Meteorological Institute, FI-00560 Helsinki, Finland (hilikka.timonen@fmi.fi;

joel.kuula@fmi.fi)

<sup>4</sup>Pegasor Oy, FI-33100 Tampere, Finland (erkka.saukko@pegasor.fi)

<sup>5</sup>Department of Computer Science, University of Helsinki, FI-00560 Helsinki, Finland (sasu.tarkoma@helsinki.fi)

<sup>6</sup>Department of Physics, The University of Jordan, Amman 11942, Jordan

Urban air pollution has been a global challenge, and continuous air quality measurement is important to understand the nature of the problem. However, missing data has often been an issue in air quality measurement. In this study, we presented a modified method to impute missing data by input-adaptive proxy. We used black carbon (BC) concentration data in Mäkelänkatu traffic site (TR) and Kumpula urban background site (BG) in Helsinki, Finland in 2017–2018 as training sets. The input-adaptive proxy selected input variables of other air quality variables based on their Pearson correlation coefficients with BC. In order to avoid overfitting, this proxy used the algorithm of least squares model with a bisquare weighting function and allowed a maximum of three input variables. The generated models were then evaluated and ranked by adjusted coefficient of determination ( $\text{adj}R^2$ ), mean absolute error and root mean square error. BC concentration was first estimated by the best model. In case of missing data in the input variables in the best model, the input-adaptive proxy then used the second-best model until all the missing data gaps were filled up.

The input-adaptive proxy managed to fill up 100% of the missing voids while traditional proxy filled only 20–80% of missing BC data. Furthermore, the overall performance of the input-adaptive proxy is reliable both in TR ( $\text{adj}R^2=0.86\text{--}0.94$ ) and in BG ( $\text{adj}R^2=0.74\text{--}0.91$ ). TR has a generally better regression performance because the level of BC can be mostly explained by traffic count, nitrogen oxides and accumulation mode. On the contrary, the source of BC in BG is more heterogeneous, which includes traffic emission and residential combustion, and the concentration of BC is influenced by meteorological parameters; therefore, the rule of including maximum three input variables might lead to the lower  $\text{adj}R^2$ . The proxy works slightly better for workdays scenario than

in weekends in both sites. In TR, the proxy works similarly in all seasons, while in BG, the proxy performance is better in winter and autumn than in the other seasons. The simplicity, full coverage and high reliability of the input-adaptive proxy make it sound to further estimate other air quality parameters. Moreover, it can act as an air quality virtual sensor alongside with on-site instruments.