Abstract

The aim of this doctoral dissertation was to develop new models for identifying the level of community exposure to selected air pollutants based on analysis of particulate matter fractions PM_{10}, pollution transport models and machine learning algorithms. Samples of PM_{10} were collected in Wadowice in 2017 and Krakow in 2020-2021.

The chemical composition of PM_{10} in terms of the content of polycyclic aromatic hydrocarbons (PAHs) was carried out using gas chromatography-mass spectrometry (GC-MS). A total of 252 samples of particulate matter PM_{10} were analyzed. The analyzes contained 16 basic PAHs identified by the US EPA as the most harmful: Acenaphtene (Acn), Acenaphthylene (Acy), Anthracene (Ant), Benzo[*b*]fluoranthene (B[*b*]F), Benzo[*a*]anthracene (B[*a*]A), Benzo[*a*]pyrene (B[*a*]P), Benzo[*ghi*]perylene (B[*ghi*]P), Benzo[*k*]fluoranthene (B[*k*]F), Chrysene (Chry), Dibenzo[*ah*]anthracene (D[*ah*]A), Fluoranthene (Flt), Fluorene (Flu), Indeno[*1,2,3-cd*]pyrene (IP), Naphthalene (Nap), and Phenanthrene (Phen) and Pyrene (Pyr). The obtained information on the concentrations of PAHs was used to determine the profiles of pollution sources, exposure profiles and the values of polycyclic aromatic hydrocarbons equivalent toxicity indicators recommended by the EPA: mutagenic equivalent to B[*a*]P (ang. *mutagenic equivalent, MEQ*), toxic equivalent to B[*a*]P (ang. *toxic equivalent, TEQ*) and carcinogenic equivalent to 2,3,7,8-tetrachlorodienzo-p-dioxin (ang. *carcinogenic equivalent, CEQ*).

In this work, air trajectory frequency analysis were performed in order to obtain information on the possibility of transporting pollutants from selected areas in the vicinity of the studied sites. The analyzes were performed using the NOAA Air Resources Laboratory's HYSPLIT model (*Hybrid Single-Particle Lagrangian Integrated Trajectory model*) developed by the NOAA Air Resources Laboratory (*National Oceanic and Atmospheric Administration*). The interpretation of trajectory results provided insights into the nature of air pollution sources.

Data of polycyclic aromatic hydrocarbons equivalent toxicity indicators (*MEQ, CEQ, TEQ*) were used to develop models for identifying the level of community exposure. The calculations were performed using Azure Machine Learning. Four machine learning algorithms were used in the models: support vectors machine, logistic regressions, decision forest and neural network.

Based on the analyses, the models of neural networks and decision forests demonstrated the highest accuracy. However, the models based on decision forest algorithms showed signs of overfitting, while the support vector machine and logistic regression algorithms had lower accuracy. The assessment of parameter significance on model accuracy indicated that the sampling site had the greatest influence. This research can serve as a basis for the development of early warning systems using neural network algorithms.