
Predicting the Toxicity of Microplastic Particles Through Machine Learning Models

Ana Leticia Antonio Vital*^{†1}, Pedro M. Da Silva², Scott Coffin³, Luan De Souza Leite⁴, Win Cowger⁵, Maaïke Vercauteren⁶, Andrea Bonisoli-Alquati⁷, and Magdalena M. Mair^{8,9}

¹Statistical Ecotoxicology, University of Bayreuth (StatEcotox, UBT) – Bayreuth, Germany

²Potomac-Hudson Engineering (PHE) – Maryland, United States

³California Office of Environmental Health Hazard Assessment (OEHHA) – California, United States

⁴University of São Paulo (USP) – São Paulo, Brazil

⁵Moore Institute for Plastic Pollution Research (-) – California, United States

⁶Blue Growth Research Lab (BGRL) (UGent) – Gent, Belgium

⁷California State Polytechnic University (Cal Poly Pomona) – Pomona, United States

⁸Statistical Ecotoxicology, University of Bayreuth (StatEcotox, UBT) – Bayreuth, Germany

⁹Bayreuth Center of Ecology and Environmental Research (BayCEER) – Bayreuth, Germany

Abstract

Unlike chemicals, which can be identified based on their well-defined molecular structure and stable properties (e.g. CAS numbers, SMILES), micro- and nanoplastic particles (MNP) lack such straightforward classification. Each MNP has its own combination of traits, including characteristics such as polymer composition, particle dimensions (length and width), and shape, as well as physicochemical properties like surface charge, surface chemistry, and plastic-associated chemicals. Furthermore, these traits may change over time, particularly due to degradation processes when MNPs are exposed to natural environments. To achieve reliable hazard and risk assessments for MNPs, it is thus necessary to predict the toxicity of MNPs with trait combinations that have not been tested directly in the lab. Similar to Quantitative Structure-Activity Relationship (QSAR) models that link molecular structures of chemicals to toxic outcomes, models are needed to link MNPs traits to their toxicities. The recently compiled *Toxicity of Microplastics Explorer (ToMEx) 2.0* database consisting of 13,412 data points from 290 published studies of MNP effects on aquatic species offers a unique opportunity to approach this task. Using ToMEx 2.0 data, we trained machine learning models on tasks to predict the toxicity (presence/absence of effects, effect direction, effective concentrations) of untested MNP. We compare the predictive performance of two machine learning algorithms (*boosted regression trees* and *deep neural networks*) and use methods of explainable AI (average marginal effects) to gain insights into the relationships between toxic outcomes and MNP traits, experimental parameters, and species traits. We also compared whether the predictions change when filtering the dataset according to study quality based on assigned quality scores. Finally, we discuss how such models can be used to predict the toxicity of environmentally relevant mixtures of MNPs and how they contribute to the development of less toxic and more environmentally friendly plastic materials in the future.

*Speaker

†Corresponding author: ana.l.antonio-vital@uni-bayreuth.de

Keywords: boosted regression trees, ecotoxicity, effects prediction, neural networks