Notes on the Supplementary Material

for the research paper

A Methodology for the Fast Identification and Monitoring of Microplastics in Environmental Samples using Random Decision Forest Classifiers

by

Benedikt Hufnagl[a,*], Dieter Steiner[a], Elisabeth Renner[a], Martin G. J. Löder[b], Christian Laforsch[b] and Hans Lohninger[a]

(*a*)
Institute of Chemical Technologies and Analytics
Vienna University of Technology

(*b*)
Department of Animal Ecology I and BayCEER
University of Bayreuth

(*) *Corresponding author:*
benedikt.hufnagl@tuwien.ac.at
https://orcid.org/0000-0002-1470-5787
Institute of Chemical Technologies and Analytics
Vienna University of Technology
Getreidemarkt 9
1060 Wien
Austria

This supplementary material contains the classification results of three different datasets (see table 1). As described in the corresponding research article the RDF model is made up out of six binary classifiers (PE, PP, PMMA, PAN, PS and NonPolymer). The output of each model can either be treated individually by setting a threshold or be combined with the other outputs using the highest value as the classification result.

For comparing the RDF outputs with other algorithms we have chosen MATLAB MAT File level 5 format [1] for publishing the classification results as it is an open format and can be read by many open source packages. We therefore also included a script which shows how the results can be accessed using python. Each `*.mat` file contains a struct named `classImg`. The field names correspond to the polymer types (e.g. `classImg.PE` refers to the binary classifier of polyethylene) and the OVA results are accessible through `classImg.OVA`. Further the OVA results are also given as PNG images which can be used for a visual comparison of the classification results. By intention the legend is not included so that pixel positions coincide with the associated spectra.

The datasets 'RefEnv1' and 'RefEnv2' were originally published by Primpke *et al.*[2] under a Creative Commons Attribution 4.0 International License. Please note that for the application of the RDF model we had to recalibrate the wavelengths of these datasets as we observed a minor shift when looking at peak positions. This was done by shifting the numeric data by one layer which corresponds to approximately 4 wavenumbers. The recalibrated data is available on request from the corresponding author.

The dataset 'Microplastic'[3] is accessible through https://doi.org/10.5281/zenodo.2555732.

Table 1: Supplementary files

File name	Description
ovaMicroplastic.png	OVA result of the 'Microplastic' dataset
ovaRefEnv1.png	OVA result of the 'RefEnv1' dataset
ovaRefEnv2.png	OVA result of the 'RefEnv2' dataset
ovaLegend.png	Legend for the OVA class colors
rdfMicroplastic.mat	Binary RDF class maps and OVA result of the 'Microplastic' dataset
rdfRefEnv1.mat	Binary RDF class maps and OVA result of the 'RefEnv1' dataset
rdfRefEnv2.mat	Binary RDF class maps and OVA result of the 'RefEnv2' dataset
pythonImportExample.py	an exemplary script for loading the `*.mat` files using python

[1] https://www.mathworks.com/help/pdf_doc/matlab/matfile_format.pdf

[2] Primpke, S., Wirth, M., Lorenz, C., Gerdts, G. *Reference database design for the automated analysis of microplastic samples based on Fourier transform infrared (FTIR) spectroscopy.* Anal Bioanal Chem (2018) 410: 5131. https://doi.org/10.1007/s00216-018-1156-x

[3] Hufnagl, B., Steiner, D., Renner, E., Löder, M. G. J., Laforsch, C., Lohninger, H. *Microplastic.* Zenodo (2019). https://doi.org/10.5281/zenodo.2555732