

DOI: 10.1177/0003702820917760

**Paper type:** *Special Issue: Microplastics*

## **Toward the Systematic Identification of Microplastics in the Environment: Evaluation of a New Independent Software Tool (siMPle) for Spectroscopic Analysis**

Sebastian Primpke<sup>1,\*</sup>, Richard K. Cross<sup>2</sup>, Svenja M. Mintenig<sup>3</sup>, Marta Simon<sup>4</sup>, Alwise Vianello<sup>4</sup>, Gunnar Gerdts<sup>1</sup>, and Jes Vollertsen<sup>4,\*</sup>

<sup>1</sup> Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Biologische Anstalt Helgoland, Kurpromenade 201, 27498 Helgoland

<sup>2</sup> Pollution Science Area, UK Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK

<sup>3</sup> Copernicus Institute of Sustainable Development, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands,

<sup>4</sup> Aalborg University, Department of the Built Environment, Thomas Manns Vej 23, 9220 Aalborg Øst, Denmark

Corresponding authors emails: Sebastian.Primpke@awi.de; jv@civil.aau.dk

### **Abstract**

Microplastics (MP) are ubiquitous within the environment, but the analysis of this contaminant is currently quite diverse, and a number of analytical methods are available. The comparability of results is hindered as even for a single analytical method such as Fourier transform infrared spectroscopy (FT-IR) the different instruments currently available do not allow a harmonized analysis. To overcome this limitation, a new free of charge software tool, allowing the systematic identification of MP in the environment (siMPle) was developed. This software tool allows a rapid and harmonized analysis of MP across FT-IR systems from different manufacturers (Bruker Hyperion 3000, Agilent Cary 620/670, PerkinElmer Spotlight 400, Thermo Fischer Scientific Nicolet iN10). Using the same database and the automated analysis pipeline (AAP) in siMPle, MP were identified in samples that were analyzed with instruments with different detector systems and optical resolutions, the results of which are discussed.

**Keywords:** Microplastic, quantification, identification, Fourier transform infrared imaging, FT-IR imaging, automated analysis, Raman, freeware, big data, micro-FT-IR,  $\mu$ FT-IR, wastewater

## Introduction

Small plastic particles, called microplastics (MP) are currently recognized as a potential risk for environmental and human health.<sup>1,2</sup> The near-ubiquitous contamination of the environment, from terrestrial soils and air to the freshwater, and marine environments has raised the profile of this topic in recent years, leading to a wealth of methods and approaches for sampling and analyzing MPs in environmental matrices.<sup>3–11</sup> In general, these particles are defined as <5 mm in size while a lower size limit and a standard definition of MP has yet to be agreed on.<sup>7,12</sup> Subcategories distinguishing between large MP (5 mm–500 µm) and small MP (500–1 µm)<sup>13</sup> are often used, reflecting practical considerations during the full analytical procedure.

The analytical procedure to identify particles can be divided into three steps, starting with sampling for MP, followed by sample extraction and finally, identification and quantification. Each of these steps has its challenges (cf. Lusher et al.,<sup>3</sup> Brander et al.,<sup>4</sup> Primpke et al.,<sup>6</sup> Cowger et al.<sup>14</sup>). Additionally, there is an increasing awareness for quality assurance and quality control (QA/QC) to successfully and reliably identify MP in environmental samples.<sup>15,16</sup> Individual steps for QA/QC are currently discussed within this overarching special issue.<sup>4</sup> While QA/QC is important for the quality of the results, the intercomparison of studies is further hampered by a missing harmonization of the three steps sampling, sample extraction and identification. Especially for the chemical analysis, a plethora of different methods and software are used to interpret generated data.<sup>5,6,14</sup> Among the spectroscopy-based techniques, Fourier transform infrared spectroscopy (FT-IR) is considered most suitable to analyze MP from 5 mm to ~10 µm, even if it requires different acquisition modes according to the particle size.<sup>6</sup> Micro-FT-IR (µFT-IR) imaging<sup>17–20</sup> permits scanning of the whole sample, avoiding the manual sorting/point-and-shoot steps that are otherwise necessary, which are a source of bias as the analysis becomes operator dependent.<sup>21,22</sup> On the other hand, FT-IR imaging for MP analysis produces a huge amount of data (large areas are scanned at a very high spatial resolution) which are difficult to manage using the commercial software provided by the instrument manufacturers.<sup>4</sup> The difficult data handling, together with the lack of a suitable software tool, impedes not only a reliable workflow, but also makes difficult, if not impossible, comparison between studies performed using FT-IR Imaging.<sup>14</sup> While initial studies comparing various analytical methods have been conducted,<sup>23</sup> a comparison between studies using the same analytical technique but on different instruments is missing for MP research. One reason is the lack of a suitable software tool to

perform such a comparison. Such a study, however, is essential to reproduce and compare results within meta-studies. Moreover, it allows the harmonized analysis of MP in the future, as the advantages and disadvantages of the currently available high throughput analysis pipelines for individual systems can be determined.

To achieve this goal, a new software tool is presented and intercalibrated with existing studies.<sup>24</sup> Its performance is evaluated via the analysis against existing reference data sets and comparison of the achieved results. Furthermore, the tool allows the analysis of data generated by different FT-IR imaging techniques from various manufacturers. This is demonstrated through the application of the software to datasets from state of the art instruments from four major manufacturers, namely Agilent, Bruker, PerkinElmer, and ThermoFisher Scientific, with different types of detection modes ranging from focal plane array (FPA), linear arrays to single-element detectors. This analysis is followed by a short performance evaluation of the assigned spectra. The software and corresponding reference spectra databases<sup>24</sup> are available free of charge via the internet and can be used by everyone for the analysis of FT-IR data.

## **Materials and Methods**

### *Sample Extraction and Analysis Using the Agilent System*

Sample preparation started with an effluent sample (Ryaverket wastewater treatment plant, WWTP, Götheborg, Sweden) as collected using filtration with a custom filtration device containing a Ø100 mm stainless steel filter (10 µm mesh size).

The material collected on the steel filter mesh was treated to extract MP using a method derived by Löder et al.<sup>13</sup> and modified by Liu et al.<sup>22</sup> The filter was sonicated into filtered Milli-Q water containing 5% (w/v) sodium dodecyl sulphate (SDS) to detach the solids and left stirring (100 rpm) at 50 °C for 48 hours. The resulting suspension was then filtered onto a 47 mm steel filter (10 µm mesh). The particles trapped were re-suspended and first incubated with proteolytic enzymes (Alcalase, Novozymes, Denmark) in a Tris buffer (pH 8.2, stirring at 100 rpm, and 50 °C for 48 hours), filtered onto a 47 mm steel filter, and the solids were then removed from the filter surface. A second enzymatic treatment was performed using cellulolytic enzymes (Cellulase blend and Viscozyme, Sigma-Aldrich) in acetate buffer (pH 4.8, stirring at 100 rpm and 50 °C for 48 hours) to eliminate the majority of the organic fraction of the sample matrix. The remaining undigested matter was filtered onto a 47 mm steel filter, and the solids were again

removed from the filter surface. The gathered solids were oxidized using Fenton reaction (hydrogen peroxide catalyzed by Fe(II) at  $\sim 20$  °C for 24 hours. After a further filtration on a 47 mm steel filter, the solids were removed from the filter surface and recovered. MP were further separated from the inorganic particles in a zinc chloride solution ( $1.7 \text{ g cm}^{-3}$ ) using a glass separatory funnel. After discharging the settled material, the supernatant was filtered (47 mm steel filter) and the material was recovered following the same procedure described for the previous steps using 50% (v/v) ethanol. The extracted MP were transferred in 10 mL glass headspace vial, the liquid was evaporated at  $55^\circ\text{C}$ , and finally, 5 mL 50% (v/v) ethanol solution were added to obtain a known volume in the vial.

In order to minimize MP contamination deriving from the equipment used for sampling and sample preparation, all lab tools were flushed with filtered ( $1.2 \mu\text{m}$ ) Milli-Q water three times before use. Tools made of glass or metal or which were coated with PTFE were used instead of plastic whenever possible. Sample containers were covered with aluminum foil to reduce airborne contamination, and steel filters were muffled at  $500^\circ\text{C}$  before usage.

The  $\mu\text{FT-IR}$  analysis was performed using a focal point array (FPA)-based  $\mu\text{FT-IR}$  imaging technique provided by a Cary 620  $\mu\text{FT-IR}$  microscope from Agilent Technologies (USA) coupled with a Cary 670 FT-IR spectroscope. The instrument was equipped with a  $128 \times 128$  FPA/mercury–cadmium–telluride (MCT) imaging detector (FPA-MCT-imaging detector). The analysis was carried out in transmission mode, using a  $15\times$  Cassegrain (visible IR) objective-condenser system which produces  $5.5 \mu\text{m}$  pixel resolution on the FPA detector. An aliquot of the sample ( $600 \mu\text{L}$ ) was deposited onto a  $\varnothing 13 \text{ mm} \times 2 \text{ mm}$ -thick zinc selenide (ZnSe) transmission window. A background scan was collected before each sample scan on a clean window at  $8 \text{ cm}^{-1}$  spectral resolution, using 120 co-added scans in the spectral range of  $3750\text{--}850 \text{ cm}^{-1}$ . Subsequently, an area of  $14 \times 14$  tiles was scanned on the samples window, using 30 co-added scans, and the same settings as for the background scan. The analyzed area covered the entire active surface of the windows (diameter of 10 mm, area  $78.5 \text{ mm}^2$ ), recording the spectra of all deposited particles.

#### *Sample Extraction and Analysis Using the PerkinElmer System*

The sample represents a composite sample taken at 30-minute intervals across a 24-hour period, directly sampling from the effluent of a WWTP. The auto-sampler filtered this water through a

woven stainless steel cylindrical filter cartridge (27.8 cm long, nominal pore size 10  $\mu\text{m}$ , ~500  $\text{cm}^2$  filter area; Wolftechnik Germany). The concentrated sample was transferred from the filter into dispersion for processing in the lab. Processing in order to “clean” the sample in preparation for  $\mu\text{FT-IR}$  analysis consisted of two steps: a Fenton’s reaction to chemically degrade organic material and enzyme digestion to remove cellulosic and proteinaceous material. The Fenton’s reaction used a Fe(II) solution (0.05 M  $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$ , Fischer Scientific, USA, >98% purity) acidified with 0.2% sulfuric acid ( $\text{H}_2\text{SO}_4$ , AnaTaR, 98.07% purity) and 30% hydrogen peroxide ( $\text{H}_2\text{O}_2$ , Fisher Scientific). The reaction was allowed to exhaust itself overnight before the sample was filtered and re-dispersed for enzymatic digestion. The enzyme digestion steps utilized cellulase in a pH 5, phosphate-buffered saline solution (MP Biomedicals, USA) incubated at 50  $^\circ\text{C}$  for 48 hours and trypsin at 37  $^\circ\text{C}$  for 30 minutes (Sigma-Aldrich, Germany). The final concentrated sample was dispersed and stored in 50% ethanol (Sigma-Aldrich, Germany) before depositing onto 25 mm diameter 5.0  $\mu\text{m}$  pore size silver membrane filters (Sterlitech, USA) for  $\mu\text{FT-IR}$  analysis. All reagents were filtered through a 1.2  $\mu\text{m}$  glass-fiber filter before use, and all processing took place in the Microflow Biological Safety Cabinet, fitted with HEPA filters to control for airborne microplastic contamination during the processing of samples.

For the  $\mu\text{FT-IR}$  analysis, a PerkinElmer Spotlight 400 FT-IR microspectrometer with MCT detector was used for the analysis of the sample. The silver filter containing the processed sample was mounted on a glass slide and held in place with a clamped stainless steel O-ring. The spectrometer collected spectra in the range between 4000 and 700  $\text{cm}^{-1}$  in reflectance mode. A background spectrum was collected for each sample from a blank area of the silver filter at a spectral resolution of 8  $\text{cm}^{-1}$  and pixel size of 25  $\mu\text{m}$ . A total of 90 scans were taken per pixel with an interferometer speed of 2.2  $\text{cm s}^{-1}$ . An optical image was first collected by tiling single field of view images together to cover an area of approximately 13 mm x 13 mm. A smaller mapping area for the FT-IR spectrum of 11.6 mm x 11.6 mm was selected (92 % of the filter area), due to constraints on the file size that could be generated by the PerkinElmer SpectrumIMAGE software. The FT-IR mapping was performed with the same parameters as that of the background scan, but at four scans per pixel. Atmospheric correction was performed on the resulting .fsm file.

### *Sample Extraction and Measurement Using the ThermoFisher Scientific System*

The sample was taken by filtering water at the effluent of a WWTP over a stainless sieve ( $\varnothing 20$  cm, ThermoFisher Scientific) having a mesh size of 20  $\mu\text{m}$ . The concentrated sample was exposed successively to sodium dodecylsulphate (one day, 5%, Serva Electrophoresis GmbH, Germany), potassium hydroxide (five days, 10%, Carl Roth GmbH, Germany), and hydrogen peroxide (two days, 32%, Carl Roth GmbH, Germany). During all steps, the sample was incubated in an oven with a temperature set at 35 °C. In between steps the sample was filtered over a  $\varnothing 47$  mm stainless steel filter with a mesh size of 20  $\mu\text{m}$ . Inorganic particles were removed by performing a density separation using a zinc chloride ( $\text{ZnCl}_2$ , Carl Roth GmbH, Germany) solution with a density of 1.6  $\text{g cm}^{-3}$ . Subsequently, sample residues were filtered on an aluminum oxide filter (Anodisc 25 mm, Whatman, UK) which was then dried at 35 °C for several days.

In order to minimize MP contamination, all chemicals were filtered through stainless steel filters with a mesh size of 10  $\mu\text{m}$ . Additionally, all lab equipment was thoroughly rinsed before usage, and the lab surfaces cleaned with ethanol (30%, Carl Roth GmbH and Co. KG, Germany) and Milli-Q water. Whenever possible plastic equipment was reduced by tools made of glass or metal, and when finishing sample handling these were immediately covered with aluminum foil to reduce airborne contamination.

For the  $\mu\text{FT-IR}$  analysis, an FT-IR microscope equipped with a single MCT detector (Nicolet iN10, ThermoFisher Scientific, USA) was used. For the measurements, an Anodisc filter was placed on a calcium fluoride crystal (EdmundOptics, Germany). About one third of each filter was mapped in transmission mode, with one scan recorded per pixel, the aperture size set at 50 x 50  $\mu\text{m}$ , the spectral resolution as 16  $\text{cm}^{-1}$ , and the spatial resolution at 20  $\mu\text{m}$ . A background scan using the same settings was conducted on a blank area of the same filter.

### *Sample Extraction and Measurement Using the Bruker System*

For this comparison study, a dataset of a sample investigated in previous studies<sup>25,26</sup> was chosen. The sample was from the effluent of the WWTP Holdorf. Sample location and further sampling information are available within the previous studies.<sup>25,26</sup> Briefly summarized, the sample was directly taken from the effluent of the WWTP. The sample was processed by the enzymatic digestion protocol of Löder et al. 2017<sup>13</sup> and subsequently concentrated onto an Anodisc filter

(25 mm diameter, 0.2  $\mu\text{m}$  pore size, GE Whatman). The sample was placed and covered by a  $\text{BaF}_2$  window prior to measurement on a custom made sample holder as described in detail in Primpke et al. 2019.<sup>25</sup>

The  $\mu\text{FT-IR}$  measurement was performed using a Bruker TENSOR II spectrometer, which is connected to a Hyperion 3000  $\mu\text{FT-IR}$ -microscope (Bruker Optics GmbH, Ettlingen, Germany). The spectra were collected using a 64 x 64 FPA MCT detector as described in literature.<sup>19</sup> Prior to measurement, a visual image of the sample surface was recorded. The FT-IR measurements were performed using 15x Cassegrain objective, in the spectral range of 3600–1250  $\text{cm}^{-1}$  with 4x4 binning at a spectral resolution of 8  $\text{cm}^{-1}$  and six coadded scans. With this setup, a pixel size of 11.1  $\times$  11.1  $\mu\text{m}$  per spectra was achieved. All data were collected using Bruker software OPUS 7.5 (Bruker Optics GmbH, Germany).

#### *Sample Preparation of Algae Samples*

Ecotoxicity tests were carried out with the microalga *Raphidocelis subcapitata* with modifications according to "Water Quality–Freshwater Algal Growth Inhibition Test with Unicellular Green Algae, ISO Standard 8692, 2004". The test included five replicates of the control sample, in which algae were not exposed to the toxicant, and triplicates of algae exposed to five different concentrations of the tested toxicant. The replicates were combined after the toxicity test, and algae from all treatments were preserved with Lugol's iodine solution for their infrared analysis.

For the infrared analysis described in Kansiz et al.,<sup>27</sup> 2 mL of the preserved control sample was prepared. The cells were centrifuged, and the residues of the preservative and growth media were washed off to prevent interference with algae spectra. The purified cells were resuspended in 200  $\mu\text{L}$  deionized water, and the entire volume was deposited on a 13 mm diameter, 1 mm-thick  $\text{CaF}_2$  transmission window, and dried in a vacuum desiccator.

A Cary 620  $\mu\text{FT-IR}$  microscope from Agilent Technologies (USA) coupled with a Cary 670 FT-IR spectroscope was used for the FPA-based  $\mu\text{FT-IR}$  imaging analysis of the algae cells. The analysis was carried out in transmission mode using the 15X Cassegrain objective in high magnification mode to create a mosaic with 1.1  $\mu\text{m}$  pixel resolution on the FPA. A background scan was collected on a clean window in the range of 3750–850  $\text{cm}^{-1}$  with an 8  $\text{cm}^{-1}$  spectral resolution applying 256 co-added scans. An area of 6 x 6 tiles was scanned on the sample

window following the background scan with the same parameters applying 240 co-added scans per pixel.

## Software and Database

The software siMPle is the combination of the software MPhunter presented in Liu et al.<sup>22</sup> and the automated analysis of Primpke et al.<sup>21</sup> It is written in Delphi using Windows 10 as the operating system, which is available at [www.simple-plastics.eu](http://www.simple-plastics.eu), where reference databases for FT-IR and Raman spectroscopy are also provided (Figure 1).

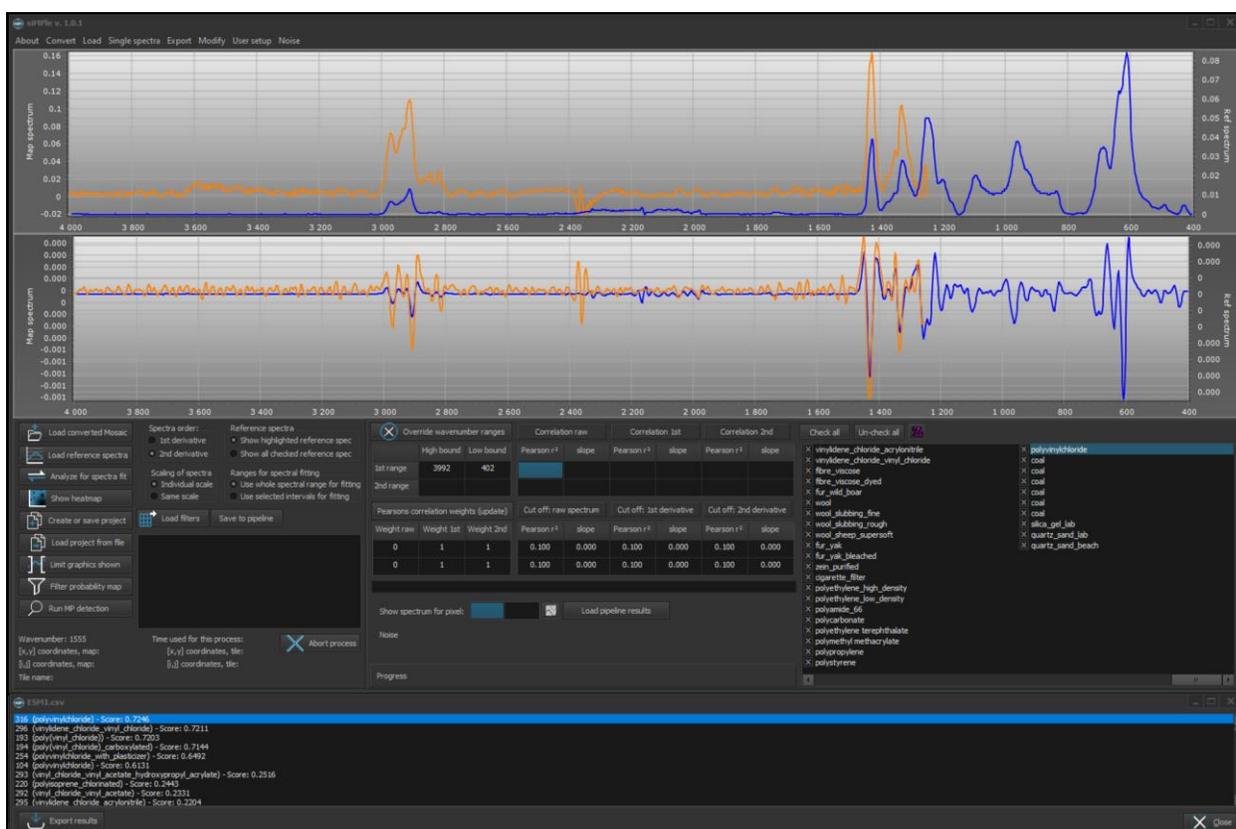


Figure 1. The software graphical user interface with a loaded reference (blue) and a sample spectrum (orange) using the Match single spectrum function of siMPle on a spectrum assigned as polyvinylchloride using ESM1.csv. In this work, the database for automated analysis<sup>24</sup> and the release version (1.0.0) of siMPle were used to analyze all datasets. Prior to analysis, all spectra were converted from the original file format (Agilent: .dmd and PerkinElmer: .fsm) or JCAMP-DX files (Bruker and Thermo Fisher Scientific) into two siMPle file formats, namely .spe and .wno files, allowing fast data access (see How to use siMPle.pdf Supplemental Material file for

further instructions). These file formats are accompanied by an extra file, the MaschineData.ini, which contains all necessary information for size calculations and data handling. In all cases the mosaic structure is either kept or newly generated, allowing faster data handling. After loading the data and reference spectra, the spectral fit between the two was calculated by Pearson correlation for the untreated data, the first derivative and the second derivative, resulting in their correlation factors  $r_0$ ,  $r_1$ ,  $r_2$ , respectively. If not further specified the following settings were used: Omit CO<sub>2</sub> peak (upper wavenumber: 2420 cm<sup>-1</sup>, lower wavenumber: 2200 cm<sup>-1</sup>), suppress negative correlations and include second order derivatives.

To investigate the performance between Bruker OPUS and siMPle, the calculation times were measured using a HP KP719AV computer (Intel Core 2 Duo Processor, 8 GB RAM, AMD Radeon HD 5450 graphic card, extra USB 3.0 Controller card and SANDISK Extreme 64 GB USB-Stick) which is the same as used in previous studies.<sup>21,24-26,28-34</sup> Further calculations were performed on a HP Z440 workstation (Intel Xeon E5-2630 v.3 CPU, 64 GB RAM, NVidia Quadro M2000 graphics card) for all other purposes.

The automated analysis pipeline (AAP) pipeline<sup>21</sup> identifies the recorded spectra based on the results of the Pearson correlation factors ( $r$ ) calculated for the respective untreated spectra  $r_0$  and the first derivatives  $r_1$ . Only if maximum values of  $r_0$  and  $r_1$  are assigned to the same polymer entry then the spectra are counted as identified and the polymer type added to the list of analyzed pixels together with the  $x,y$  coordinates and the summarized hit quality index (HQI, Eq. 1):

$$HQI = (r_1 + r_2) \times 1000 \quad (1)$$

This type of data represents a false colour image which was then *in silico* treated by Image Analysis as described in Primpke et al.<sup>21</sup> by a pixel hole closing mechanism prior to the size determination and particle quantification. For further calculations, the data thresholds described in Lorenz et al.<sup>32</sup> were used. To avoid confusion for the reader, we did not apply the second analysis pipeline<sup>1,22,35</sup> available in siMPle for the scope of this study.

The siMPle software allows rapid quality control and quality assurance (QA/QC) for the assigned polymer hits. During image analysis, a designated file is generated named “\_forqc.csv”, which contains  $x,y$  coordinates of the measured spectra together with the hit quality, the assigned

polymer type, and a reference spectrum identifier. Via the button “Load Pipeline Results” the QA/QC process will be started (Figure 2).

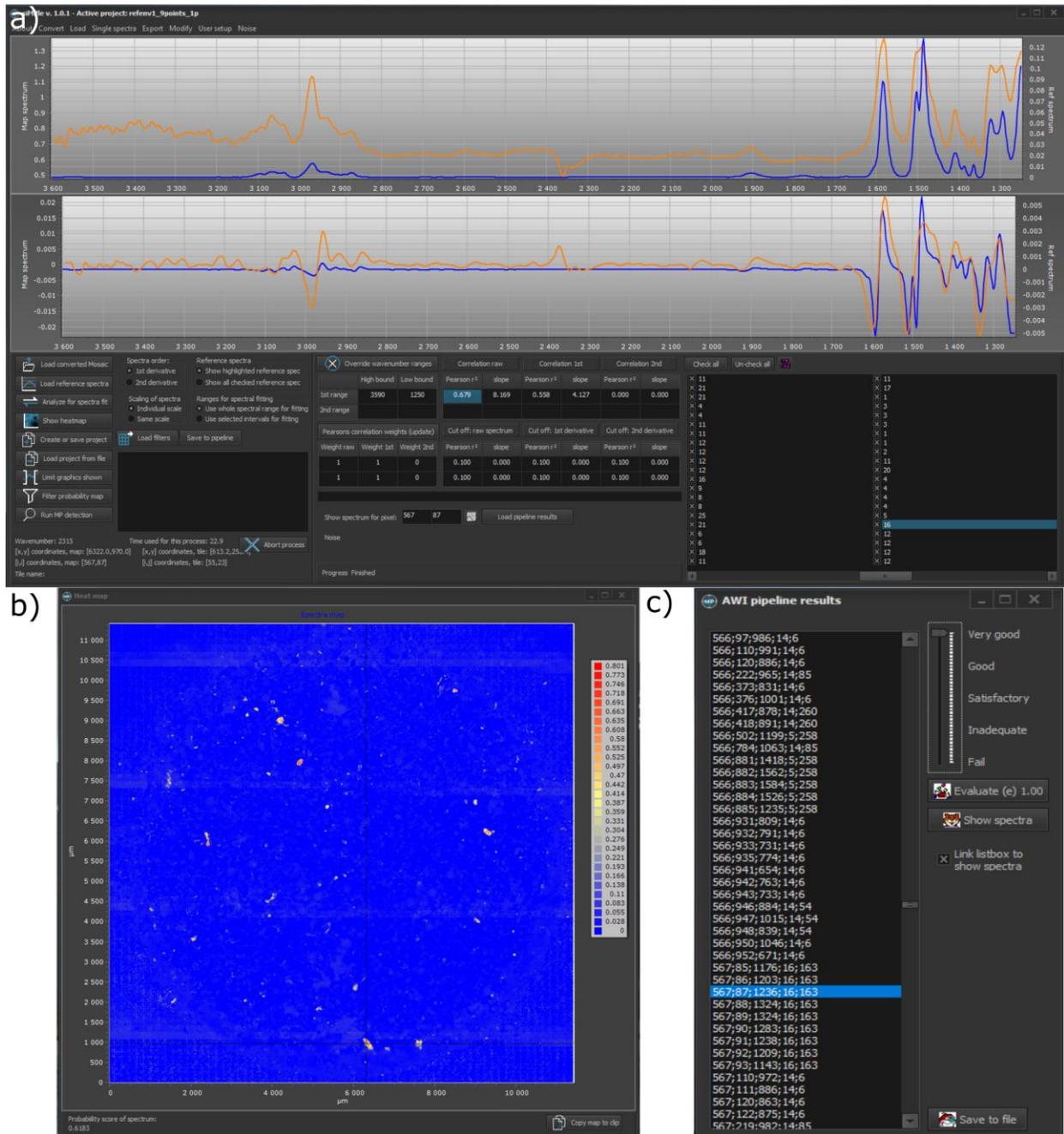


Figure 2. Quality assurance/quality control for siMPLE for results of the image analysis of sample RefEnv1<sup>24</sup> via the AWI pipeline allowing a direct comparison (a) between sample spectrum (orange) and database spectrum (blue) by clicking on the determined pixel (c). The process allows checking other spectra by clicking on another database entry as well as the individual

Pearson correlation factors. The heatmap (b) allows the user to locate the assigned spectra on the map. Using a designated window, each spectrum can be individually assessed and rated from a perfect assignment down to a full misassignment with values ranging from 1 to 0.01. Following the approach of Pimpke et al.,<sup>21</sup> the spectra were rated with five different values (1.00, 0.75, 0.5, 0.25, and 0.01) ranging from best match to a full misassignment, respectively. To screen the data of the different instruments and evaluate the successful assignment of polymer identify to particles, a randomly selected number ( $n = 100$ ) of spectra were manually reanalyzed for each instrument dataset.

For matching single spectra, the score used within siMPle was first described in Liu et al.<sup>22</sup> In short, prior to the quantification of particles the score ( $S_d$ ) for the identification of the polymer type was calculated using the following Eq. 2:<sup>22</sup>

$$S_d = \frac{k_0 r_0^2 + k_1 r_1^2 + k_2 r_2^2}{k_0 + k_1 + k_2} \quad (2)$$

Each result ( $r_0, r_1, r_2$ ) was squared and multiplied by a weight ( $k_0, k_1, k_2$ , respectively) for the respective correlation factor, which can be assigned by the user.

## Results and Discussion

The software allows two types of data analyses, first the matching of single spectra and secondly the analysis of large filter areas. To interpret a single spectrum, the reference database must first be loaded and then a single spectrum in a defined file format (Paragraph S1; ESM1.csv, Supplemental Material) must be loaded (Figure 1). In this example, the spectrum was assigned to polyvinylchloride (PVC) with a  $S_d$  of 0.7246 using the default options of siMPle ( $k_0=0$ ,  $k_1=1$ ,  $k_2=1$ ). By assigning weights, the user can decide which correlation method should be represented by  $S_d$ , e.g., for comparison with studies using Bruker OPUS<sup>28,30-33</sup> where only the first derivative was used for single spectrum analysis. For the chosen example, values of  $k_0=k_2=0$ ,  $k_1=1$  (first derivative only) were used. In this case the  $S_d$  decreased to 0.6689. If the user decides to include all correlation results ( $k_0=k_1=k_2=1$ ) the  $S_d$  was further decreased to 0.6435. Therefore, it is mandatory to state the weights  $k_0$ ,  $k_1$ ,  $k_2$  within the material and methods section for comparison of studies if siMPle is used. In general, this analysis is independent from the instrumental source of the data. All types of data in the described file format (see Paragraph S1) can be processed

independently from manufacturer and measurement method. An example of this using pre-processed Raman spectra is shown in Figure S1. Future data processing steps for single spectra or lists of spectra will be introduced in future releases. All releases will be accompanied via a change log and a living manual on the website.

During this process, QA/QC is easily available, because the spectrum with the highest hit is indicated at the end of the analysis together with a hit list for all other database entries (Figure 1). Together with this, the hit result can be exported for further use.

### **Analysis of Chemical Imaging Files**

For larger datasets, the siMPle software allows a time-efficient loading of data using a harmonized file system for data storage which is independent from the original file format. The file formats introduced were optimized for the fast use within the software. The commonly shared file format by the International Union of Pure and Applied Chemistry (IUPAC) J-CAMPdx file format was limited to the import of data, due to long loading times of such text type based files. Currently, siMPle is able to convert native data from Agilent and PerkinElmer systems using the file import function while for Bruker and ThermoFisher Scientific systems extra steps are necessary (see How to Use siMPle.pdf, Supplemental Material).

To validate the performance of siMPle, it was tested against existing reference datasets from literature. These reference sets consist of materials of known origin (Ref7P) or from environmental samples (RefEnv1 and RefEnv2). These three samples were analyzed using siMPle and the results compared to the automated analysis via Bruker OPUS. Starting with Ref7P we started to analyze the performance on an artificial sample only containing MP (polyethylene (PE), polyester (PEST), polyamide (PA) and polyurethane (PU)), cellulose, quartz, and diatomaceous earth. Out of these seven particle types, only diatomaceous earth could not be detected within the wavenumber range of Anodisc. In comparison to OPUS, siMPle identified almost twice the number of spectra on the specific polymer types (Figure 3). Especially cellulose (plant fiber in the database), and polycaprolactone (PCL, not originally introduced as a material) were affected by factors larger than four (Figure 3, Ref7P 9Points and OPUS).

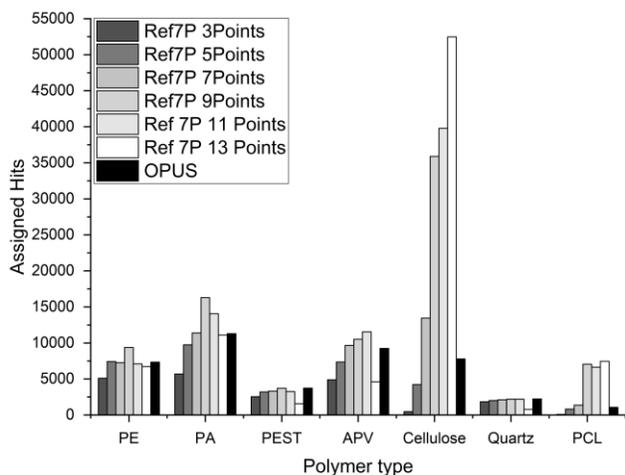


Figure 3. Assigned polymer types for sample Ref7P<sup>24</sup> using different smoothing factors by siMPLe (3 to 13) versus OPUS with an unknown fixed value for polyethylene (PE), polyamide (PA), polyester (PEST), acrylates/polyurethanes/varnish (APV), cellulose, quartz and polycaprolactone (PCL). This difference was rather striking and the main difference between both kinds of software was found in the data handling for the calculation of the first derivative. In the default settings siMPLe adds a nine data points smoothing to reduce the noise. For Bruker OPUS it is not documented if smoothing is applied. To test for a better comparison of the results, a range of this value from three to 13 data points was investigated for siMPLe (Figure 3).

By screening the number of data points for smoothing, it was found that an optimal hit was reached with nine data points for most polymer type assignment (Figure 3). Only cellulose kept an increase in assigned polymer hits while PCL reached a constant level. The data determined by OPUS could not be assigned to a smoothing factor applied by siMPLe. Still, for this particular sample, the high number of assignments to PCL started using this number of data points for smoothing. To avoid any misassignment issues a manual reanalysis on the assigned spectra to PCL was performed.

Through quality assurance, it was found that the spectra assigned to PCL were caused by a misinterpretation of the measured PEST spectrum. This spectrum has a high similarity with PCL in transmission, which was not visible by using Bruker OPUS during cluster performance analysis.<sup>24</sup> In Figure S2, one of the assigned spectra from Ref7P is plotted against the assigned reference spectrum and the spectra of the original material. The original database states this

material as a pure PEST, but via an extended material research, it was found that the material was meanwhile relabeled to copolyester by the manufacturer. Due to these differences, the material could not be assigned to the original PEST cluster, as no pure PEST spectrum was yielded. This issue will be addressed in the future by a database update including more materials and using siMPle for cluster performance analysis. All samples in the following were analyzed using the default nine data points smoothing.

The datasets RefEnv1 and RefEnv2 were also analyzed using siMPle and OPUS. The siMPle analysis required only two hours for RefEnv1 and three hours for RefEnv2, which is twelve times faster than the analysis with OPUS using spectral correlation only for raw and first derivative data. With a look at the polymer composition (see Table S1 for details) it was found that siMPle was more sensitive and identified higher numbers of polymers and also larger sized polymer particles (Figure 4) in comparison to OPUS.

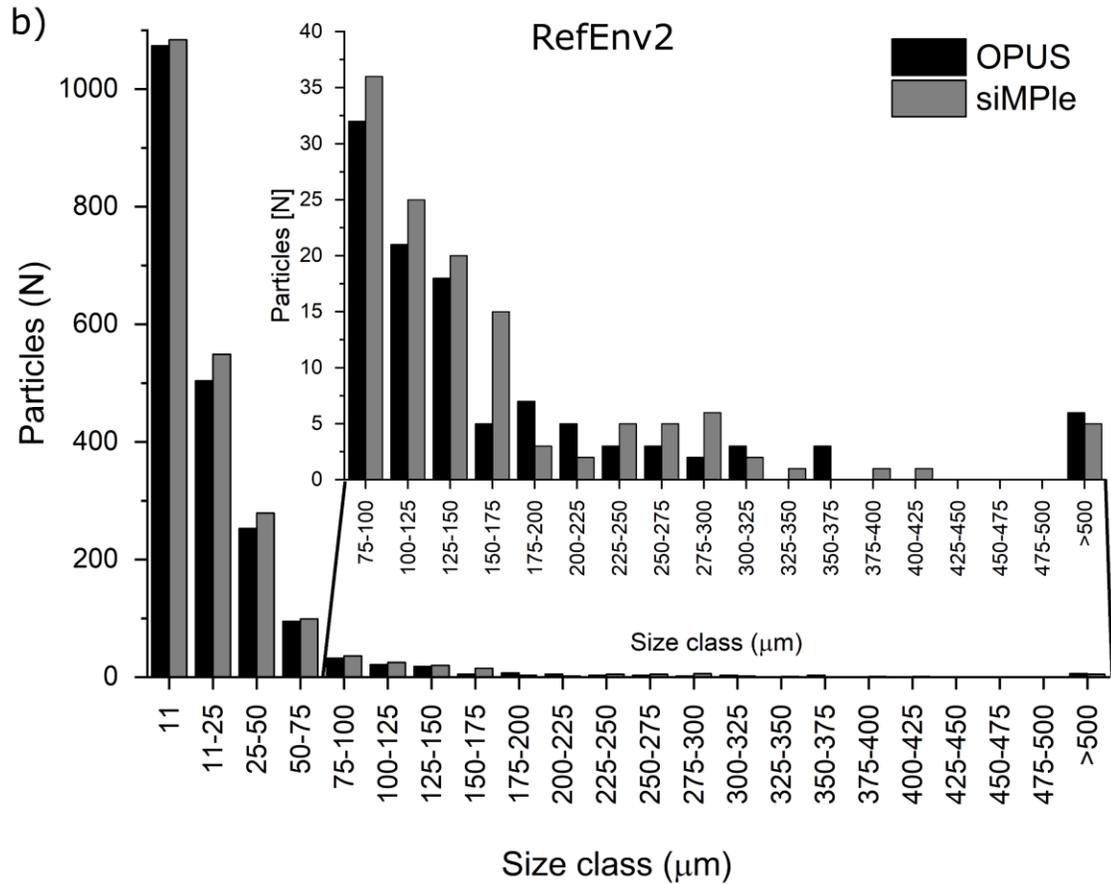
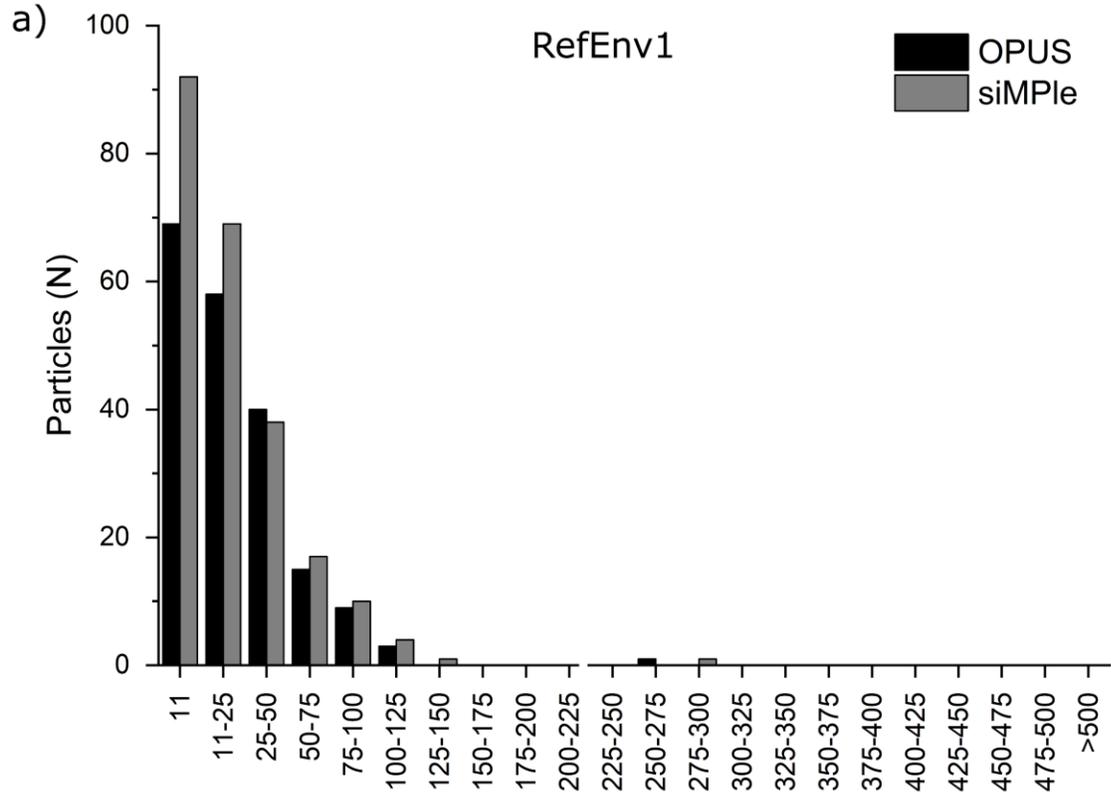


Figure 4. MP size classes derived from the automated image analysis<sup>21</sup> for the reference datasets (a) RefEnv1<sup>24</sup> and (b) RefEnv2<sup>24</sup> analyzed using Bruker OPUS and siMPle. Both analyses found a strong trend toward smaller MP sizes. Especially striking was the higher identification rate for cellulose (plant fiber) during the analysis with siMPle (see RefEnv2, Table S1). Sample RefEnv2 also showed the largest differences in the size distribution and showed a better particle assignment compared to the data derived via OPUS (Figure S3).

Here, it was found that larger particles were identified more accurately with siMPle in comparison to OPUS, which missed areas of larger particles yielding in a separation into two particles. Furthermore, the analysis using siMPle improved closing holes, which is important for morphological analysis of the particles, see for example the large PP particle on the rightmost edge of the filter. In summary, these results show that data determination with siMPle is better suited for the analysis of imaging data due to transparent data handling and easy data validation. To test the ability for a harmonized analysis of MP, the performance of siMPle was assessed for the FT-IR imaging data from instruments from the four mentioned manufacturers. In this case, the computation of a full spectral analysis based on Pearson correlation for the untreated data, the first derivative and the second derivative were performed.<sup>22</sup> The calculation time was determined on the same computer systems applied for the automated analysis via the OPUS software of the Bruker dataset. This allows a comparison of calculation times with existing studies using OPUS.<sup>26,28</sup> The determined calculation times are summarized in Table I together with further information on the datasets, such as pixel size on the filter area, the size of the analyzed area and the number of spectra recorded.

As mentioned previously, the calculation time on the Bruker datasets was reduced considerably when applying siMPle (five hours, in comparison to 48 hours using OPUS) which also included the second derivative (it was omitted during OPUS analysis). When the calculation time was normalized to the number of analyzed spectra, the spectra from Thermo Fisher Scientific were correlated twice as fast compared to the other data sets. The reason is the spectral resolution of  $16\text{ cm}^{-1}$  instead of  $8\text{ cm}^{-1}$ . Still, one has to keep in mind that the Bruker system and the Thermo Fisher Scientific system need an additional transformation step within their respective software which increases the overall handling and calculation times independent from siMPle. In Figure 5 the false colour images of the analyzed samples are shown.

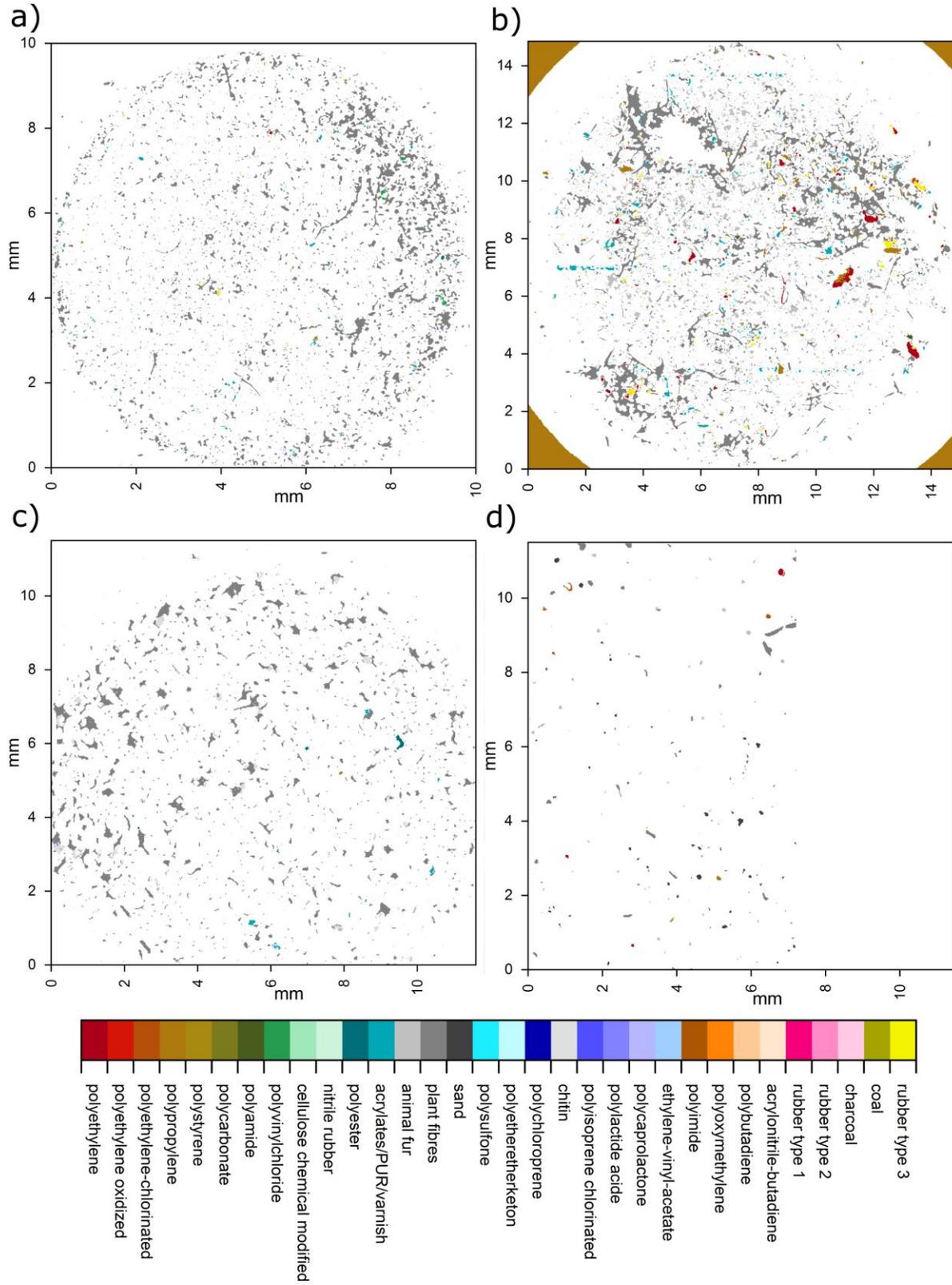


Figure 5. Overview images of the measured filters using the automated analysis pipeline for the (a) Agilent, (b) Bruker, (c) PerkinElmer, and (d) Thermo Fisher Scientific samples. Sample (d) was measured in a rectangular shape and the area of on the right side was left blank to avoid irritations. Qualitatively, it can be observed from the images that the samples are similar in nature, containing a large proportion of natural materials (Figure 6, gray colors), amongst which a number of artificial polymers are successfully identified, irrespective of the manufacturer of the instrument or the various sampling and extraction methods employed prior to analysis (see ESM2.xlsx, Supplemental Material, for details). Due to the varying nature of the WWTP sampled from, and the variety of sampling and extraction methods utilized between samples, commentary on any differences in enumeration of MP between the samples is beyond the scope of this study. However, application of the software on these real-world example datasets demonstrates promising consistency in the proportion of particles identified which are of synthetic origin (3–25%) and of the major polymer types which are identified across the samples (see ESM2.xlsx for details). Compared to existing commercial software solutions the harmonized analysis via siMPle saves working time and computational costs. Current computers can run several instances of the software unattended, allowing the data analysis of up to 16 samples per day compared to OPUS (2 days) for 1.8 million spectra per file. Still, it is possible also to use low-cost office computers which normally can calculate up to three samples per day containing 1.8 million spectra. As a minimum requirement, a processor speed of 3 GHz is advised with 8 of RAM. To assess the performance of siMPle on these datasets, a QA/QC analysis on the overall result was performed (Figure 6).

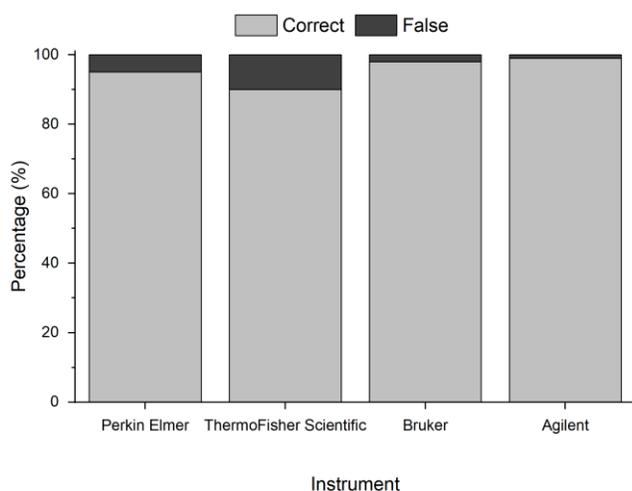


Figure 6. Assignment rates of correct and misidentified spectra for the different instruments based on manual reanalysis similar to Primpke et al.<sup>21</sup>

In all cases, correct assignment rates >90% were reached, for three systems (PerkinElmer, Bruker, and Agilent) these were even >95% (Figure 6). Those correct assignment rates were exceedingly high, independently from instrument and sample preparation, proving the high potential of siMPle as a harmonized tool for MP analysis. Still, it is suggested and recommended that each study conducts an own QA/QC analysis for each sample series for each polymer type identified as demonstrated, e.g., in Lorenz et al.<sup>32</sup> Further questions, such as a comparison between existing analytical pipelines, their harmonization and a full QA/QC analysis will be addressed in a later detailed study.

To conclude it is noteworthy that the siMPle software is not limited to MP analysis, it also allows the analysis of other types of data like the spectral comparison of nano-FT-IR data<sup>36</sup> or the analysis of single algae species (Figure S4).

Using siMPle single cells can be selected, or specific characteristics can be highlighted (Figure S4). Here, the data shows a strong Halo effect (Figure S4b) mainly caused by interference between the sample and the surface of the CaF<sub>2</sub> window, which is not visible using a heatmap (Figure S4a). In the future, Heatmaps based on the integration of specific regions will be introduced to allow even more control over the data. Further, additional functions are currently planned to be introduced, and new possibilities can be determined by contacting the authors to explore its application in a broader scope for future research.

## **Conclusion**

With siMPle we present a freeware data analysis tool for the harmonized and systematic analysis of spectroscopic data, with application for example, in the identification of MP in the environment. It allows data determination and interpretation in a transparent manner and reproducible manner. In addition, it provides a simpler quality QA/QC compared to existing commercial software tools like Bruker OPUS and shows an increased identification rate. Furthermore, it allows the analysis independently from the instrument manufacturer for a single spectrum but also for large fields generated by imaging techniques. In particular, the field of FT-IR imaging benefits greatly, as the data calculation time is reduced from several days to five hours using this software tool. Compared to other techniques, all spectra are correlated via three

different data treatments with the database yielding high-quality results for all investigated instrument systems. This new tool improves the application of FT-IR imaging in monitoring studies for MP, as it is accessible for most types of spectrometers, free of charge and reduces the human bias during manual data analysis.

### **Funding**

This work was supported by the German Federal Ministry of Education and Research (Project BASEMAN - Defining the baselines and standards for microplastics analyses in European waters; BMBF grant 03F0734A). S.M. Mintenig was funded by the Dutch Technology Foundation TTW (project 13940).

### **Conflict of Interest**

The authors declare that they have no conflict of interest.

### **Supplemental Material**

All supplemental material mentioned in the text is available in the online version of the journal.

### **References**

1. A. Vianello, R.L. Jensen, L. Liu, J. Vollertsen. "Simulating Human Exposure to Indoor Airborne Microplastics Using a Breathing Thermal Manikin". *Sci. Rep.* 2019. 9 (1): 8670.
2. C.M. Rochman, C. Brookson, J. Bikker, N. Djuric, et al. "Rethinking Microplastics as a Diverse Contaminant Suite". *Environmental Toxicology and Chemistry.* 2019. 38 (4): 703–711.
3. A.L. Lusher, K. Munno, L. Hermabessiere, S. Carr. "Isolation and Extraction of Microplastics from Environmental Samples: An Evaluation of Practical Approaches and Recommendations for Further Harmonisation". *Appl. Spectrosc.* 2020. Paper Submitted.
4. S. Brander, V. Renick, M. Foley, A. Lusher, et al. "Sampling and QA/QC, or How Many Blanks Do I Need?: A Guide for Scientists Investigating the Occurrence of Microplastics Across Different Matrices". *Appl. Spectrosc.* 2020. Paper Submitted.

5. C. Zarfl. "Promising Techniques and Open Challenges for Microplastic Identification and Quantification in Environmental Matrices". *Anal. Bioanal. Chem.* 2019. 411 (17): 3743–3756.
6. S. Primpke, S.H. Christiansen, W. Cowger, H. De Frond, et al. "Critical Assessment of Analytical Methods for the Harmonized and Cost Efficient Analysis of Microplastics". *Appl. Spectrosc.* 2020. Paper Under Revision.
7. C. Schwaferts, R. Niessner, M. Elsner, N.P. Ivleva. "Methods for the Analysis of Submicrometer- and Nanoplastic Particles in the Environment". *Trends Anal. Chem.* 2019. 112: 52–65.
8. Y. Yu, D. Zhou, Z. Li, C. Zhu. "Advancement and Challenges of Microplastic Pollution in the Aquatic Environment: A Review". *Water, Air, Soil Pollut.* 2018. 229 (5): 140.
9. J.Y. Li, H.H. Liu, J.P. Chen. "Microplastics in Freshwater Systems: A Review on Occurrence, Environmental Effects, and Methods for Microplastics Detection". *Water Res.* 2018. 137: 362–374.
10. F. Stock, C. Kochleus, B. Bansch-Baltruschat, N. Brennholt, G. Reifferscheid. "Sampling Techniques and Preparation Methods for Microplastic Analyses in the Aquatic Environment - a Review". *Trends Anal. Chem.* 2019. 113: 84–92.
11. N.P. Ivleva, A.C. Wiesheu, R. Niessner. "Microplastic in Aquatic Ecosystems". *Angew Chem. Int. Ed.* 2017. 56 (7): 1720–1739.
12. N.B. Hartmann, T. Huffer, R.C. Thompson, M. Hasselov, et al. "Are We Speaking the Same Language? Recommendations for a Definition and Categorization Framework for Plastic Debris". *Environ Sci Technol.* 2019. 53 (3): 1039–1047.
13. M.G.J. Löder, H.K. Imhof, M. Ladehoff, L.A. Loschel, et al. "Enzymatic Purification of Microplastics in Environmental Samples". *Environ. Sci. Technol.* 2017. 51 (24): 14283–14292.
14. W. Cowger, A. Gray, H. De Frond, A. Deshpande, et al. "Review of Data Processing and Interpretation Techniques for Microplastic with Future Outlook". *Appl. Spectrosc.* 2020. Paper Submitted.
15. E. Hermsen, S.M. Mintenig, E. Besseling, A.A. Koelmans. "Quality Criteria for the Analysis of Microplastic in Biota Samples: A Critical Review". *Environ. Sci. Technol.* 2018. 52(18): 10230–10240.

16. A.A. Koelmans, N.H.M. Nor, E. Hermsen, M. Kooi, et al. "Microplastics in Freshwaters and Drinking Water: Critical Review and Assessment of Data Quality". *Water Res.* 2019. 155: 410–422.
17. A. Kappler, D. Fischer, S. Oberbeckmann, G. Schernewski, et al. "Analysis of Environmental Microplastics by Vibrational Microspectroscopy: FT-IR, Raman or Both?". *Anal. Bioanal. Chem.* 2016. 408 (29): 8377–8391.
18. A.S. Tagg, M. Sapp, J.P. Harrison, J.J. Ojeda. "Identification and Quantification of Microplastics in Wastewater Using Focal Plane Array-Based Reflectance Micro-FT-IR Imaging". *Anal. Chem.* 2015. 87 (12): 6032–6040.
19. M.G.J. Loder, M. Kuczera, S. Mintenig, C. Lorenz, G. Gerdt. "Focal Plane Array Detector-Based Micro-Fourier-Transform Infrared Imaging for the Analysis of Microplastics in Environmental Samples". *Environ. Chem.* 2015. 12 (5): 563–581.
20. J.P. Harrison, J.J. Ojeda, M.E. Romero-Gonzalez. "The Applicability of Reflectance Micro-Fourier-Transform Infrared Spectroscopy for the Detection of Synthetic Microplastics in Marine Sediments". *Sci. Total Environ.* 2012. 416: 455–463.
21. S. Primpke, C. Lorenz, R. Rascher-Friesenhausen, G. Gerdt. "An Automated Approach for Microplastics Analysis Using Focal Plane Array (FPA) FT-IR Microscopy and Image Analysis". *Anal. Methods.* 2017. 9 (9): 1499–1511.
22. F. Liu, K.B. Olesen, A.R. Borregaard, J. Vollertsen. "Microplastics in Urban and Highway Stormwater Retention Ponds". *Sci. Total Environ.* 2019. 671: 992–1000.
23. G. Renner, A. Nellessen, A. Schwiers, M. Wenzel, et al. "Data Preprocessing and Evaluation Used in the Microplastics Identification Process: A Critical Review and Practical Guide". *Trends Anal. Chem.* 2019. 111: 229–238.
24. S. Primpke, M. Wirth, C. Lorenz, G. Gerdt. "Reference Database Design for the Automated Analysis of Microplastic Samples Based on Fourier Transform Infrared (FT-IR) Spectroscopy". *Anal. Bioanal. Chem.* 2018. 410 (21): 5131–5141.
25. S. Primpke, P.A. Dias, G. Gerdt. "Automated Identification and Quantification of Microfibres and Microplastics". *Anal. Methods.* 2019. 11 (16): 2138–2147.
26. S. Primpke, H. Imhof, S. Piehl, C. Lorenz, et al. "Environmental Chemistry Microplastic in the Environment". *Chem. Unserer Zeit.* 2017. 51 (6): 402–412.

27. M. Kansiz, P. Heraud, B. Wood, F. Burden, et al. "Fourier Transform Infrared Microspectroscopy and Chemometrics as a Tool for the Discrimination of Cyanobacterial Strains". *Phytochemistry*. 1999. 52 (3): 407–417.
28. L. Cabernard, L. Roscher, C. Lorenz, G. Gerdt, S. Primpke. "Comparison of Raman and Fourier Transform Infrared Spectroscopy for the Quantification of Microplastics in the Aquatic Environment". *Environ. Sci. Technol.* 2018. 52 (22): 13279–13288.
29. M. Bergmann, S. Mützel, S. Primpke, M.B. Tekman, et al. "White and Wonderful? Microplastics Prevail in Snow from the Alps to the Arctic". *Sci. Adv.* 2019. 5 (8): Eaax1157.
30. M. Bergmann, V. Wirzberger, T. Krumpfen, C. Lorenz, et al. "High Quantities of Microplastic in Arctic Deep-Sea Sediments from the HAUSGARTEN Observatory". *Environ. Sci. Technol.* 2017. 51 (19): 11000–11010.
31. M. Haave, C. Lorenz, S. Primpke, G. Gerdt. "Different Stories Told by Small and Large Microplastics in Sediment—First Report of Microplastic Concentrations in an Urban Recipient in Norway". *Mar. Pollut. Bull.* 2019. 141: 501–513.
32. C. Lorenz, L. Roscher, M.S. Meyer, L. Hildebrandt, et al. "Spatial Distribution of Microplastics in Sediments and Surface Waters of the Southern North Sea". *Environ. Pollut.* 2019. 252: 1719–1729.
33. T. Mani, S. Primpke, C. Lorenz, G. Gerdt, P. Burkhardt-Holm. "Microplastic Pollution in Benthic Midstream Sediments of the Rhine River". *Environ Sci Technol.* 2019. 53 (10): 6053–6062.
34. I. Peeken, S. Primpke, B. Beyer, J. Gutermann, et al. "Arctic Sea Ice is an Important Temporal Sink and Means of Transport for Microplastic". *Nature Commun.* 2018. 9 (1): 1505.
35. K.B. Olesen, D.A. Stephansen, N. Van Alst, J. Vollertsen. "Microplastics in a Stormwater Pond". *Water*. 2019. 11 (7): 10.3390/W11071466.
36. M. Meyns, S. Primpke, G. Gerdt. "Library Based Identification and Characterisation of Polymers with Nano-FT-IR and IR-sSNOM Imaging". *Anal. Methods*. 2019. 11 (40): 5195–5202.

**Table I. Calculation times of the different datasets measured on systems of four different manufacturers using siMPle.**

Data Set	Size .spe	Spectra	Pixel Size	Calculation Time	Calculation performance	Filter area measured
	GB	N	$\mu\text{m}$	s	spectra/s	mm
Agilent	9.01	3211264	5.5	29979	107	$10 \times 10$
Bruker	4.14	1806336	11.05	16464	110	$14.9 \times 14.9$
PerkinElmer	0.66	215296	25	2877	75	$11.6 \times 11.6$
ThermoFisher Scientific	0.25	221184	20	1129	195	$11.5 \times 7.2$

## Supplemental Material

### **Toward the Systematic Identification of Microplastics in the Environment: Evaluation of a New Independent Software Tool (siMPle) for Spectroscopic Analysis**

DOI: 10.1177/0003702820917760

Sebastian Primpke<sup>1,\*</sup>, Richard K. Cross<sup>2</sup>, Svenja M. Mintenig<sup>3</sup>, Marta Simon<sup>4</sup>, Alvis Vianello<sup>4</sup>, Gunnar Gerdt<sup>1</sup>, and Jes Vollertsen<sup>4,\*</sup>

<sup>1</sup> Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Biologische Anstalt Helgoland, Kurpromenade 201, 27498 Helgoland

<sup>2</sup> Pollution Science Area, UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK

<sup>3</sup> Copernicus Institute of Sustainable Development, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands,

<sup>4</sup> Aalborg University, Department of the Built Environment, Thomas Manns Vej 23, 9220 Aalborg Øst, Denmark

Corresponding authors emails: Sebastian.Primpke@awi.de; jv@civil.aau.dk

**Paragraph S1:** File format for Single Spectra Correlation

The file must be saved as .csv file containing the entries WaveNumber (in  $\text{cm}^{-1}$ ) and Spectrum (absorbance value at the assigned WaveNumber) separated by a “,”. Please keep in mind that the format is currently case sensitive. Afterwards follows the spectrum starting from the lowest wavenumber and the measured absorbance for (FTIR) or counts (for Raman).

Example:

WaveNumber,Spectrum

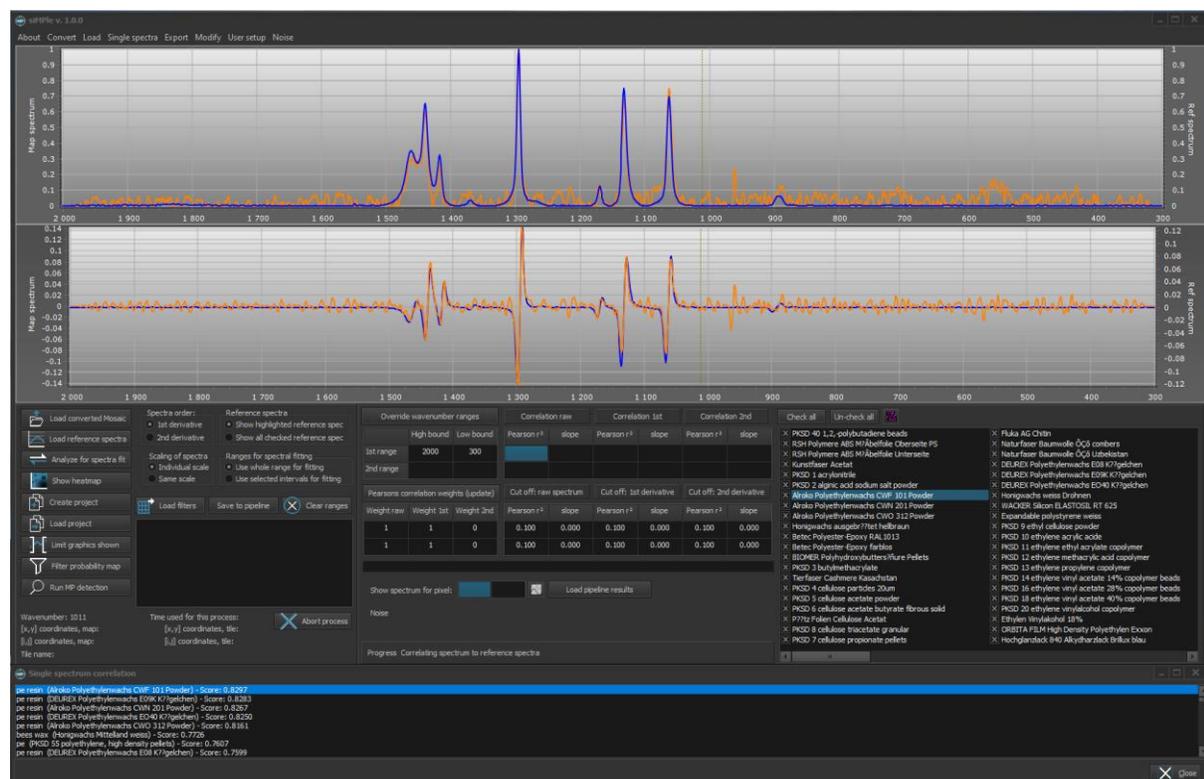
1249.5058,0

1251.43405,0.00269

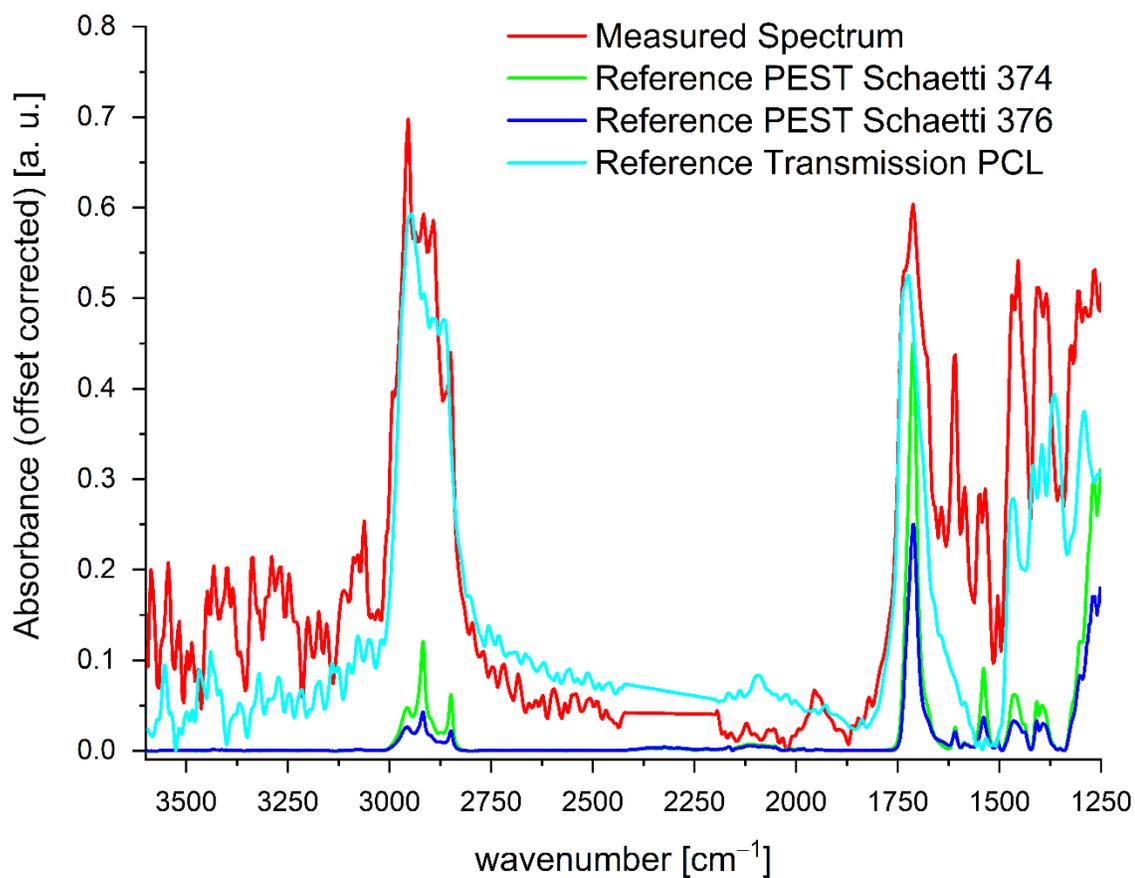
1253.3623,0.02026

1255.29055,0.03502

etc.



**Figure S1. Exemplary processed Raman spectrum identified using siMPLS  $S_d = 0.8292$  for polyethylene.**



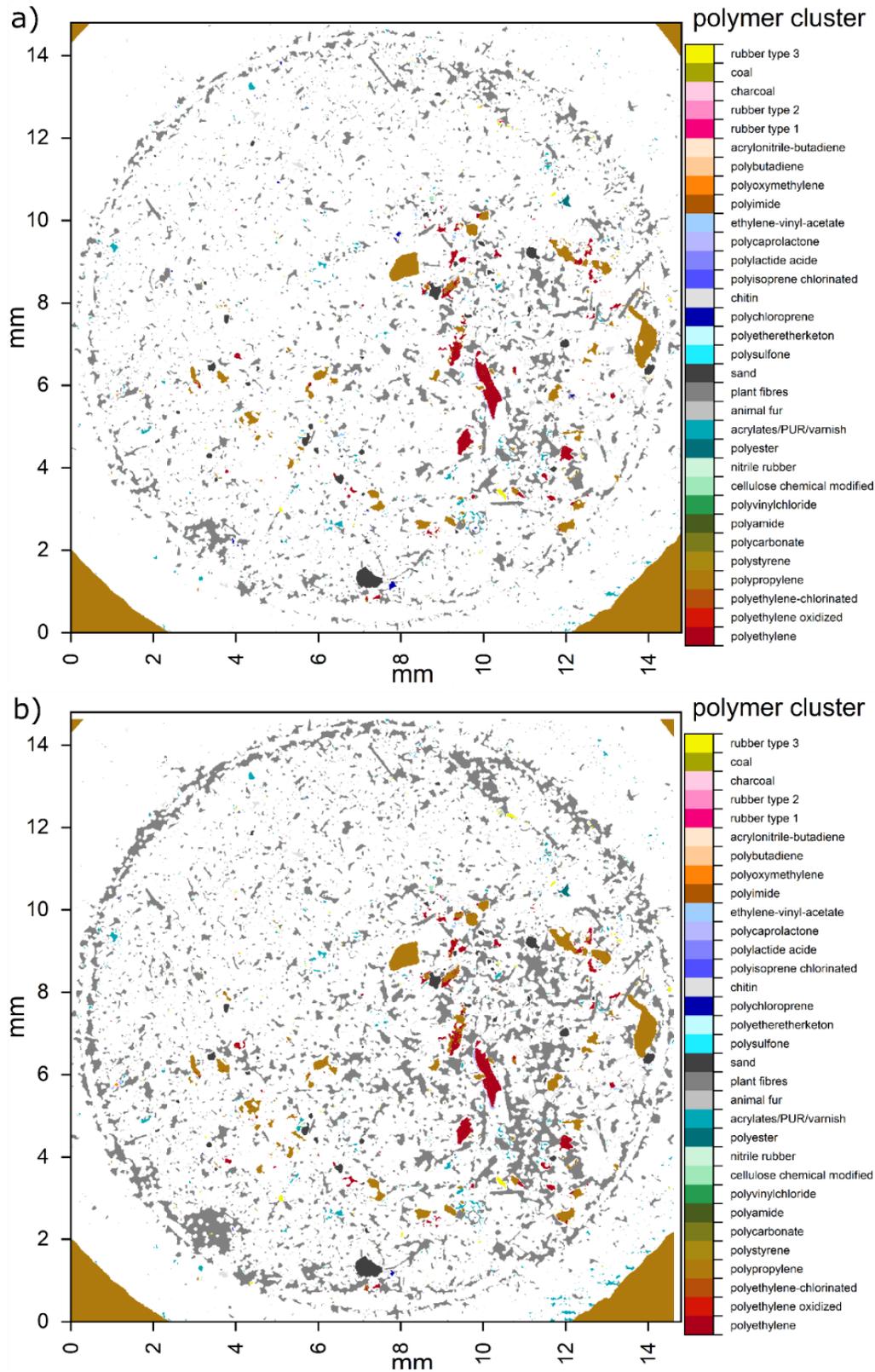
**Figure S2.** Comparison between a measured spectrum assigned to polycaprolactone (PCL), the polyester (PEST) added to the sample from the manufacturer Schaetti and the spectrum of PCL causing the assignment measured in transmission.

**Table S1.** Spectrum based polymer type identification rates using OPUS and siMPLe on sample RefEnv1.

	RefEnv1	RefEnv1	RefEnv2	RefEnv2
Polymer type	OPUS	siMPLe	OPUS	siMPLe*

Polyethylene	45	46	7114	7579
Polyethylene oxidized	0	0	7	0
Polyethylene-chlorinated	0	0	47	93
Polypropylene	753	794	54934	54965
Polystyrene	152	170	4	3
Polycarbonate	1	2	2	0
Polyamide	33	33	65	80
Polyvinylchloride	30	32	1	6
Cellulose chemical modified	0	0	163	163
Polyester	4	4	179	198
Acrylates/polyurethanes/varnish	430	450	3001	3860
Animal fur/natural polyamides	126	180	685	1151
Plant fibers/ cellulose	4520	13750	82563	141833
Sand	10	1	7348	6808
Polysulfone	495	508	2	1
Polychloroprene	1	0	401	105
Chitin	10	7	3756	4382
Polycaprolactone	0	0	180	285
Ethylene-Vinyl-Acetate	1	2	522	334
Polyoxymethylene	3	1	12	20
Rubber type 1	0	1	26	0
Charcoal	0	1	1	1
Coal	0	11	1	6
Rubber type 3	112	302	683	1067

\*Please note that the export for siMPle could only be performed for a fieldsize of 85\*85 FPA fields by combing 2 x 2 FPA fields to a larger one, an export of each field was not possible because the Marco stop working due to unknown reasons. The dataset misses 42910 spectra and mainly PP assignments as it was part of the outer region also visualized in Figure XY.



**Figure S3. False color overview images from the analysis of the sample RefEnv2 via OPUS (a) and siMPLe (b). Please note that the export for siMPLe could only be performed for**

RefEnv2 (field size of 85\*85 FPA fields) by combing 2 x 2 FPA fields to a larger one, an export of each field was not possible because the Marco stop working due to unknown reasons. The dataset misses 42910 spectra visible as white area at the top and right side.

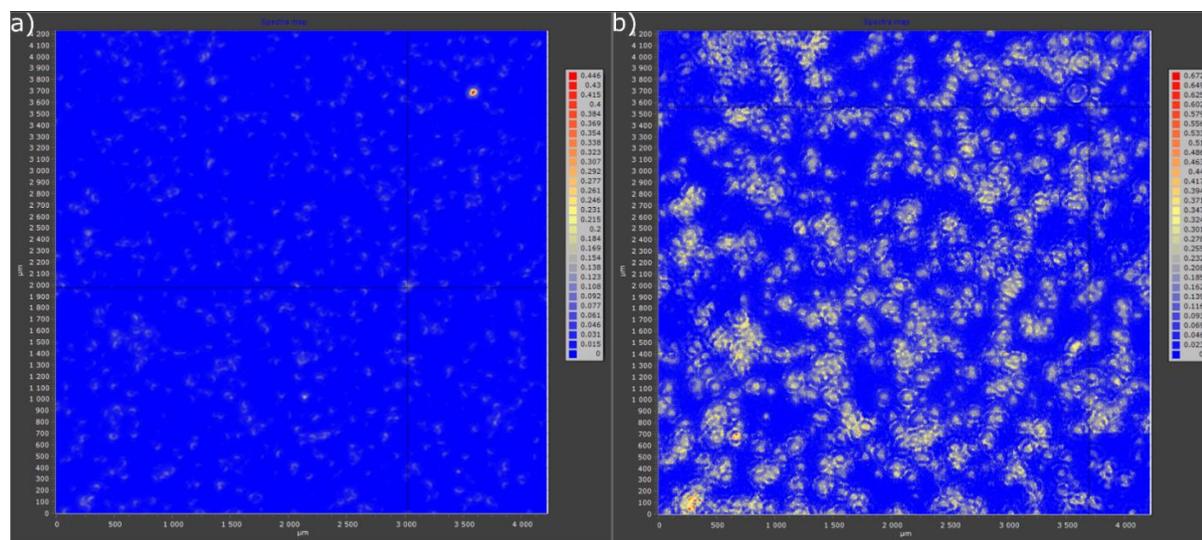


Figure S4. Algae cells (*Raphidocelis subcapitata*) placed on  $\text{CaF}_2$  measured using the Agilent system. (a) General Heatmap based on the highest Peak in the spectrum. (b) Visualization of the particles using spectral comparison against the database entry algae *fucus serratus*.

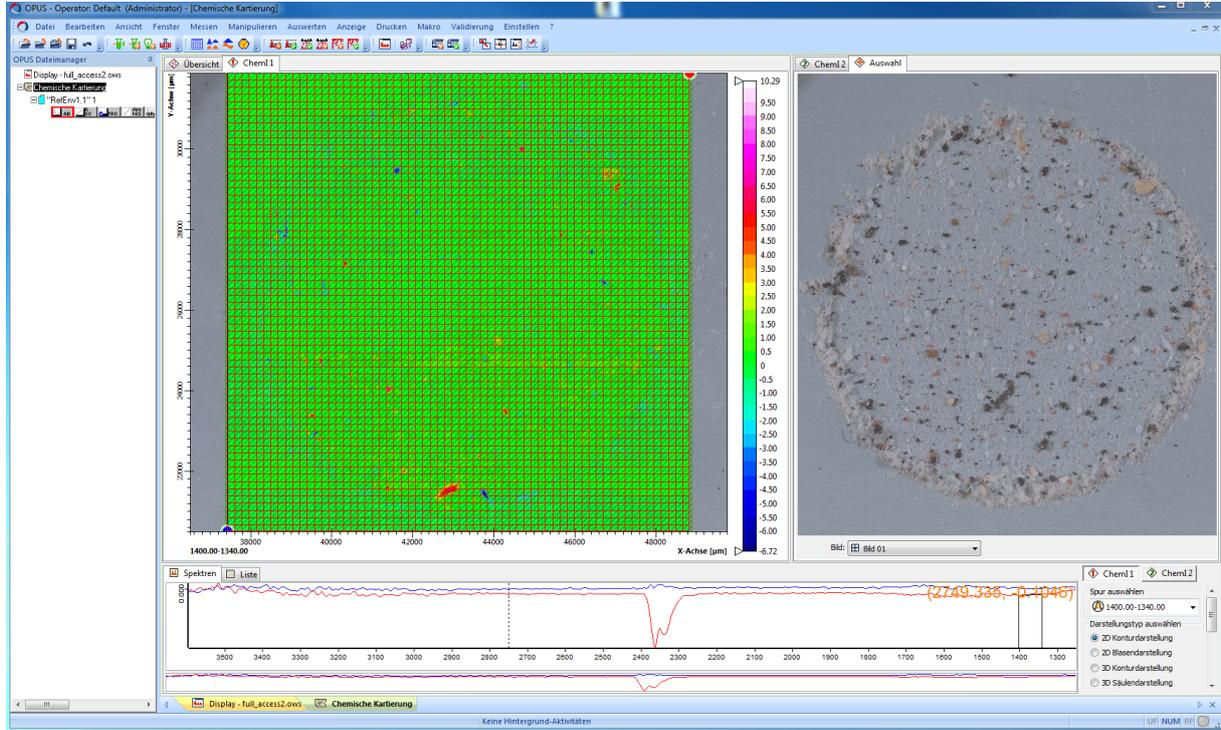


## Converting Bruker and ThermoFisher Scientific Data into files suitable for siMPLe.

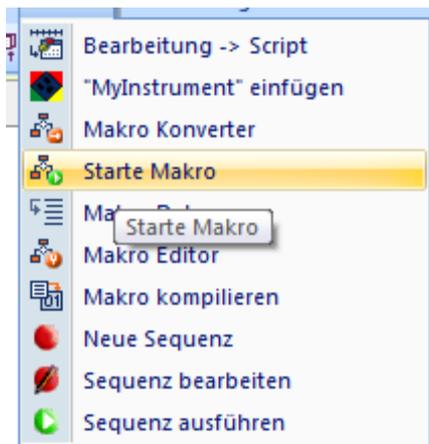
### Bruker OPUS

Conversion of get OPUS Data to siMPLe

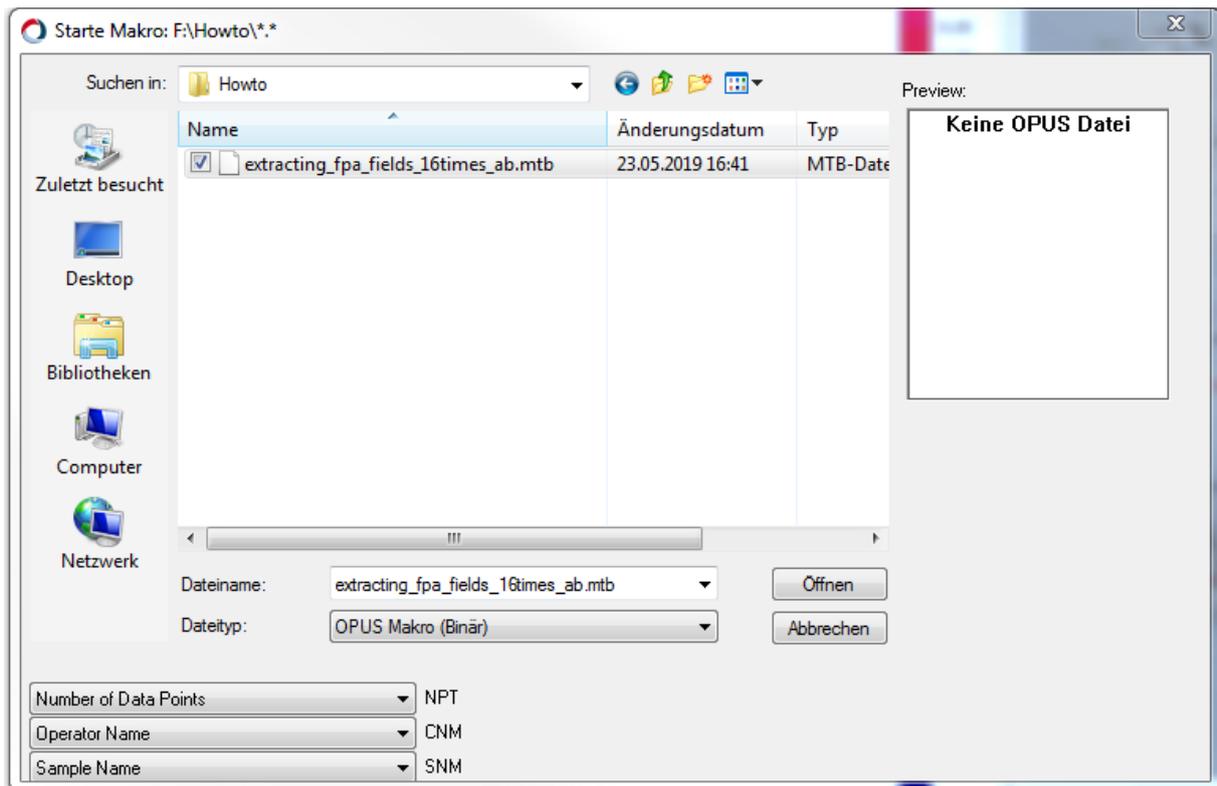
1. Load the respective file into OPUS: (This description is using the fullAccess.ows)



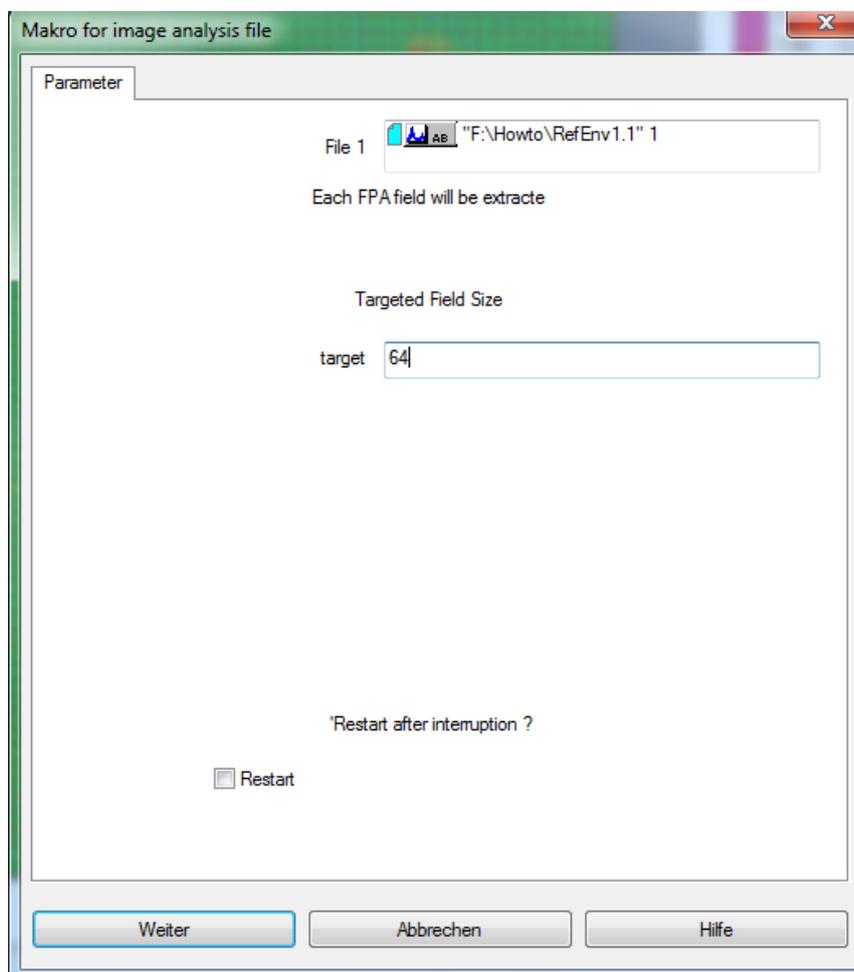
2. Go to Macro/Makro:



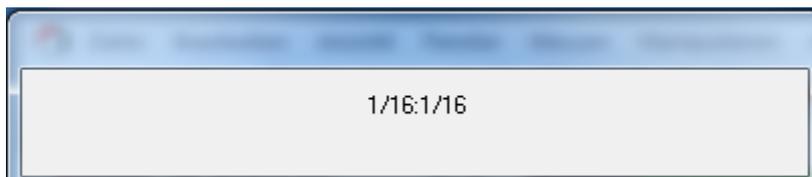
3. Choose the macro from the GoogleDrive (<https://drive.google.com/open?id=1O3vtsb963KoGwsTTGvDZgo8KUqjGjNAD>) (Choose as file type OPUS Macro (binary)) and open it.



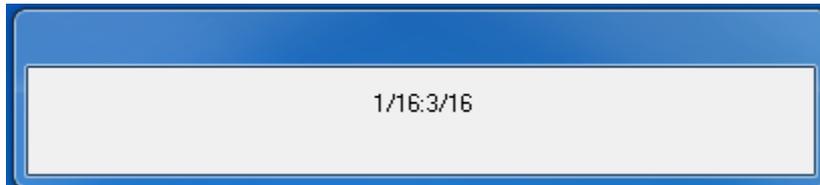
4. Choose your target field size (simple support 16,32,64,128 **datapoints** (will read in as 16 x 16, etc.) and click Weiter/Proceed. This case uses a value of 64.



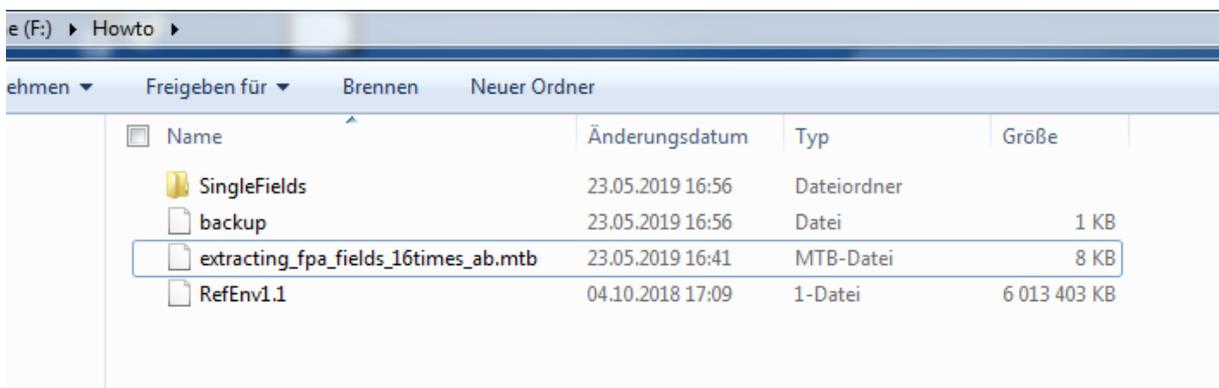
5. Data extraction starts, it is beneficial to minimize OPUS for performance:



6. These numbers should now increase:



7. Depending on the hard drive this process is rather slow or fast and one will find a new file and a folder generated:



The folder Single Fields contains the generated .dx files. Depending on the file size this process may be instable. If problems occur please inform the corresponding authors.

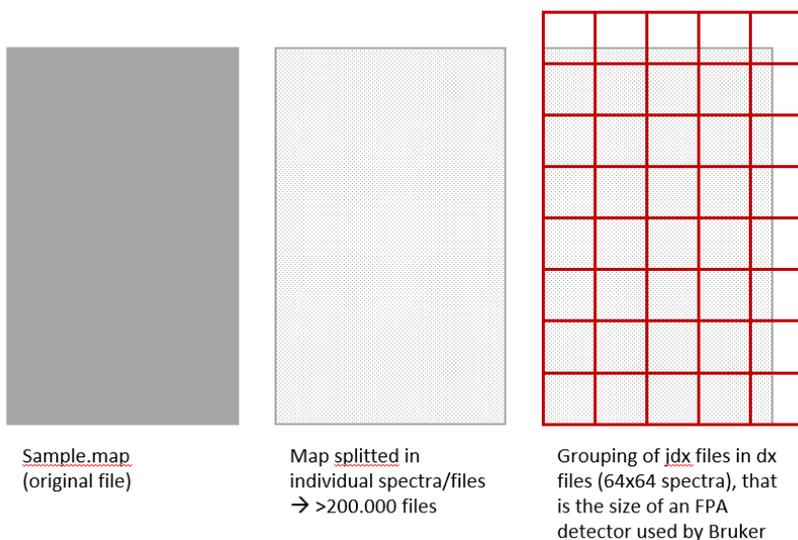
In this case the extraction of 1 million spectra in 256 single fields took 121 minutes. These values may change for different field sizes. Currently .dx files larger 128 datapoints are currently not supported by siMPle.

The files are ready to be converted via simple via BRUKER .DX INTO .SPE FILES by selecting all dx files.

## THERMO $\mu$ -FTIR data

### OMNIC PICTA

Preparation of map files that have been generated using the Thermo (single MCT) FTIR. Using Thermo software Picta to transform maps in a file type that can be read by the MPHunter (single jdx files).



1. Open the .map file in Omnic Picta
2. Note the dimensions of the mapped area (in measuring points). As can be seen in above scheme the dx files might be larger than the original map. Rows will be 'filled up' by repeating the result of the last pixel, this needs to be corrected (see below \*).
3. Create a .txt file with the bold text below. In Omnic, click ANALYZE (left bar) – SHOW MAP INFO, copy following information (bottom) in the .txt file and save this in a folder as "Map info.txt".

#### MAP DESCRIPTION

Tile name: MyTiles

First map location: -40.0, -5474.5

Last map location: 6880.0, 4845.5

Mapping stage X step size: 20.0 micrometers

Mapping stage Y step size: 20.0 micrometers

4. In Omnic, click ANALYZE – SPLIT MAP: Select path (created folder), choose jdx file and then OK. Splitting the data can take > 45 min (depending on the map size).

## siMPle

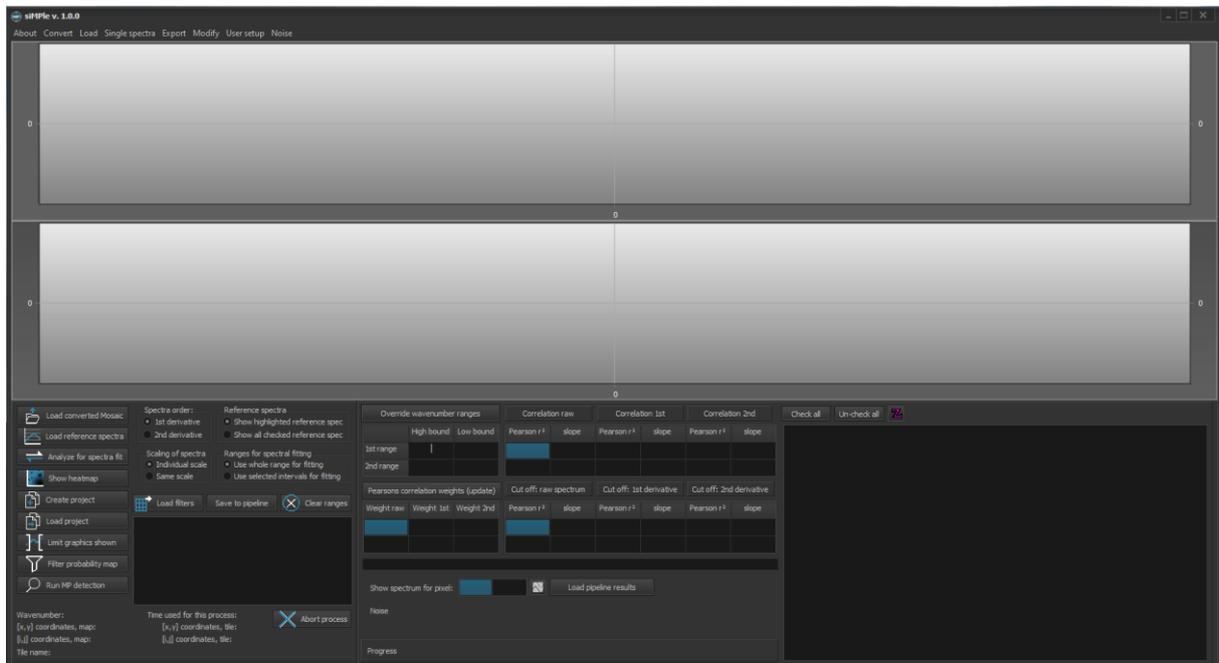
1. Click CONVERT – CONVERT NICOLET SPECTRA .JDX INTO .DX FILES: A window will open, then choose the created folder and click on the generated .txt file (map info). This will take a few minutes.

NOTE: the .jdx files must be in the same directory as the map info file, the names/ numbering done during the splitting cannot be changed!

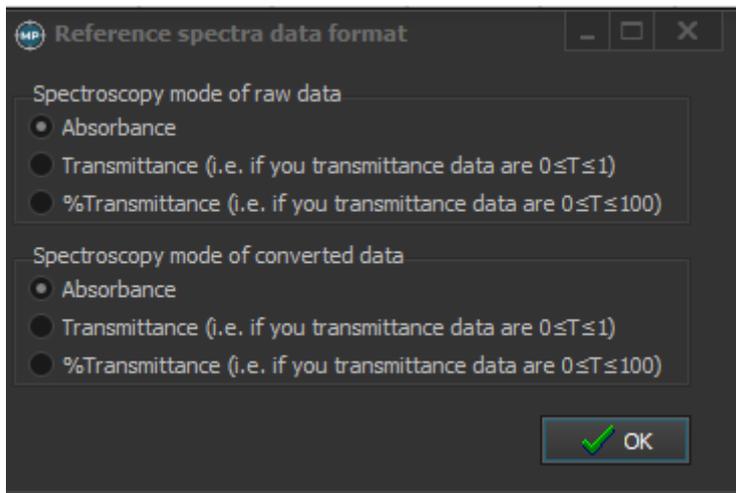
2. Click CONVERT – CONVERT BRUKER .DX INTO .SPE FILES by selecting all dx files (This will take also few minutes)
3. LOAD .SPE FILES

### How to use the AWI pipeline:

1. Load .spe files or convert (.dx, .dmd, .jdx or .fsm) files into .spe, files.

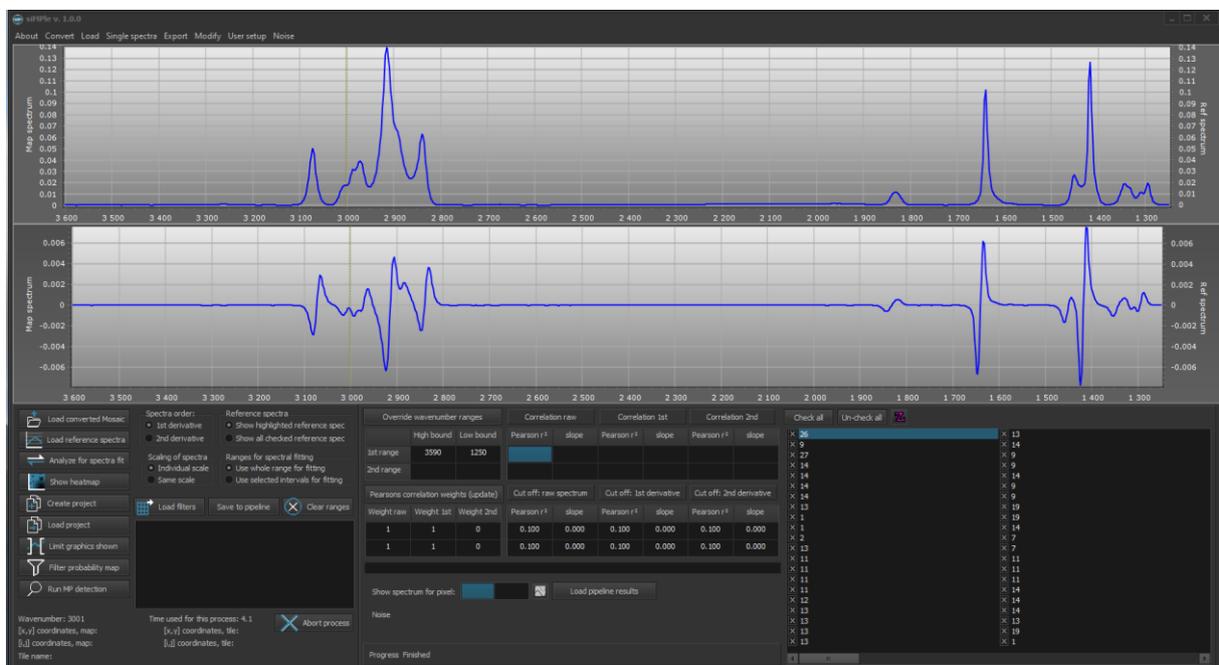


2. Load the file containing reference spectra, for AWI analysis choose: siMPle\_database\_Version1.0.txt

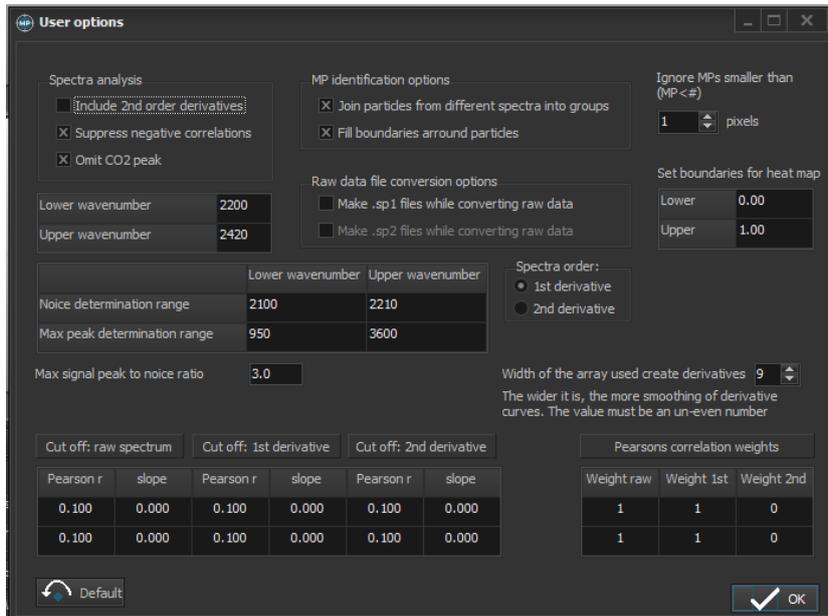


Please keep in mind that the reference spectra were measured in absorbance, press OK.

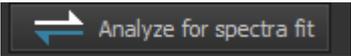
3. siMPle should now look like this:



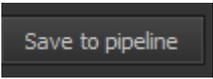
- Under User setup -> Options, values and options for the search can be checked and adjusted. Proceed by pressing OK.



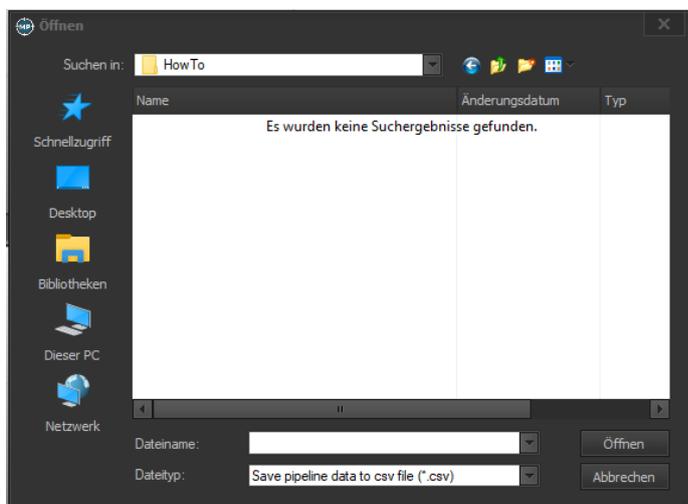
Proceed by clicking okay.

- Start the calculation by pressing:  The program will ask you to generate a project file (.ini), here called howto.ini, then click Save. The file will be saved automatically when the analysis is completed.

- Wait until the calculations are finished.

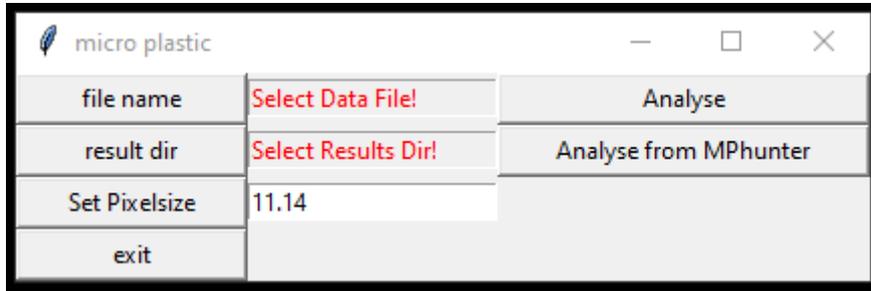
- Afterwards, click on the button: 

- Choose a file name (here howto.txt). Take care not to rename the database file (currently this is the default name generated by siMPle). Depending on the amount of identified MP this may take some time.

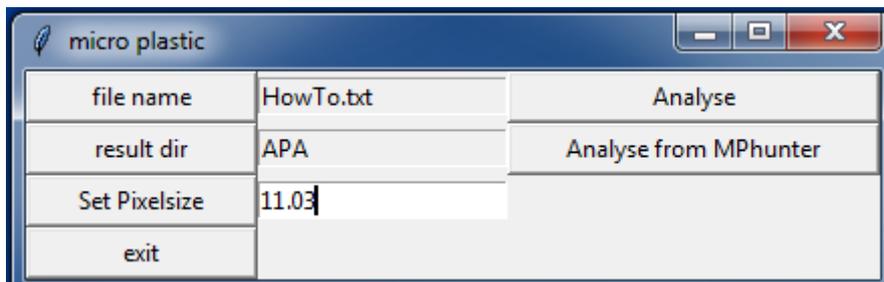


## APA.exe or MPAPP.exe

1. Depending on the pipeline you want to use, open the APA.exe or MPAPP.exe (both filenames contain the current version number, e.g. APAv1.01.exe, which may differ due to updates).
2. After a short while the user interface shows up (here AAP):

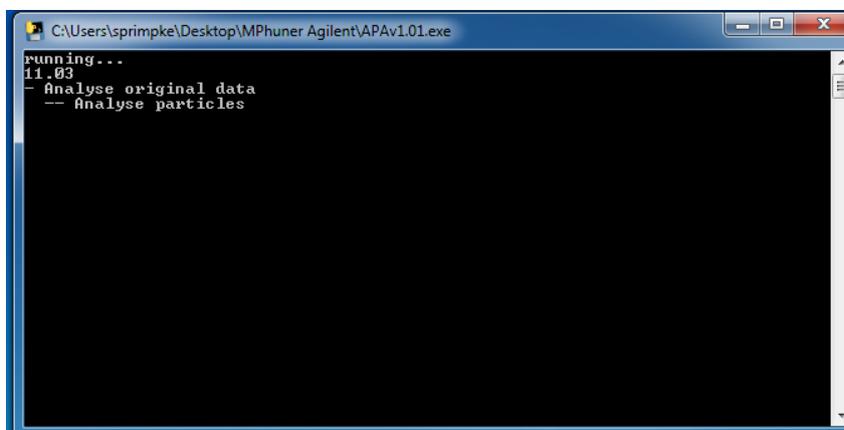


3. Set the pixel size to the value of your system, here 11.03, and press Set Pixelsize (otherwise 11.14 will be used as default).
4. By pressing "file name" you can choose the previously generated file (here HowTo.txt in the "MPhunter Agilent" folder).
5. Chose a result folder by pressing "result dir" (here called APA in the "MPhunter Agilent" folder). Important, this folder needs to be in the same folder as the executable, currently no sub folders are allowed (know bug, work in progress).

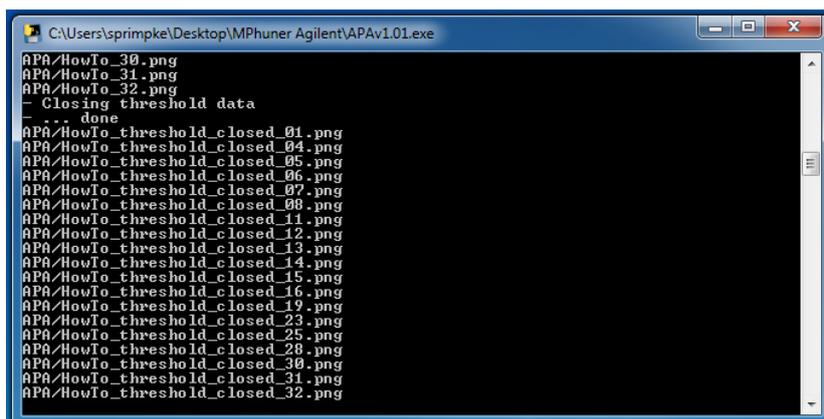


6. After folders and files are selected, and settings are set, press "Analyse from MPhunter" to start the analysis of generated data followed by an image analysis.

The start should look like this:



7. When the analysis is finished it will look like this:



8. Per samples various files are produced, all including data and images generated during image analysis. These are:

- a. A histogram for polymer type hits based on spectra.
- b. The combined (`_closed_image.csv`) and the individual (`_closed_imageID.csv`, ID = 01, etc.) image files.
- c. The number of particles found for each pixel area (`_particle_histogramID.csv`)
- d. All individual particle information (which number has which size) as `_individualID.csv`.
- e. A histogram for polymer type hits based on particle (`_particles_perpolymer.csv`)
- f. The size classes used in various publication:

`_sizeclassesID.csv` for individual particles types,

`_bio` for natural materials,

`_polymer` for plastic materials,

`_other` for all other materials,

`_bulk` (bio, polymer and other combined),

`_forcalculations` (all individual sizeclasses for all polymers in one file):

	A	B	C	D	E	F	G
1							
2	single	0	4	0	0	22	0
3	25	0	0	0	0	30	1
4	50	0	1	0	0	14	3
5	75	0	1	0	0	8	2
6	100	0	0	0	0	4	1
7	125	0	0	0	0	2	0
8	150	0	0	0	0	0	0
9	175	0	0	0	0	0	0
10	200	0	0	0	0	0	0
11	225	0	0	0	0	0	0
12	250	0	0	0	0	0	0
13	275	0	0	0	0	0	0
14	300	0	0	0	0	0	0
15	325	0	0	0	0	0	0
16	350	0	0	0	0	0	0
17	375	0	0	0	0	0	0
18	400	0	0	0	0	0	0
19	425	0	0	0	0	0	0
20	450	0	0	0	0	0	0
21	475	0	0	0	0	0	0
22	500	0	0	0	0	0	0
23	large	0	0	0	0	0	0
24							

Data reads as followed: The first column labeled with s does not contain data and should be deleted before starting further calculations. Subsequent columns contain numbers per polymer cluster (# 1-32):

	A	B	C	D	E	F	G
1	s		1	2	3	4	5
2	single	0	4	0	0	22	0
3	25	0	0	0	0	30	1
4	50	0	1	0	0	14	3
5	75	0	1	0	0	8	2
6	100	0	0	0	0	4	1
7	125	0	0	0	0	2	0
8	150	0	0	0	0	0	0
9	175	0	0	0	0	0	0
10	200	0	0	0	0	0	0
11	225	0	0	0	0	0	0
12	250	0	0	0	0	0	0
13	275	0	0	0	0	0	0
14	300	0	0	0	0	0	0
15	325	0	0	0	0	0	0
16	350	0	0	0	0	0	0
17	375	0	0	0	0	0	0
18	400	0	0	0	0	0	0
19	425	0	0	0	0	0	0
20	450	0	0	0	0	0	0
21	475	0	0	0	0	0	0
22	500	0	0	0	0	0	0
23	large	0	0	0	0	0	0
24							

\_forstatistics (all individual sizeclasses in one file):)

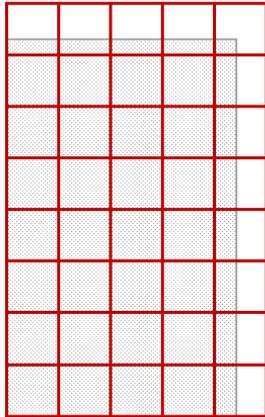
- g. Greyscale image of the raw and final images.

Analysis is finished! Welcome to the data Jungle!

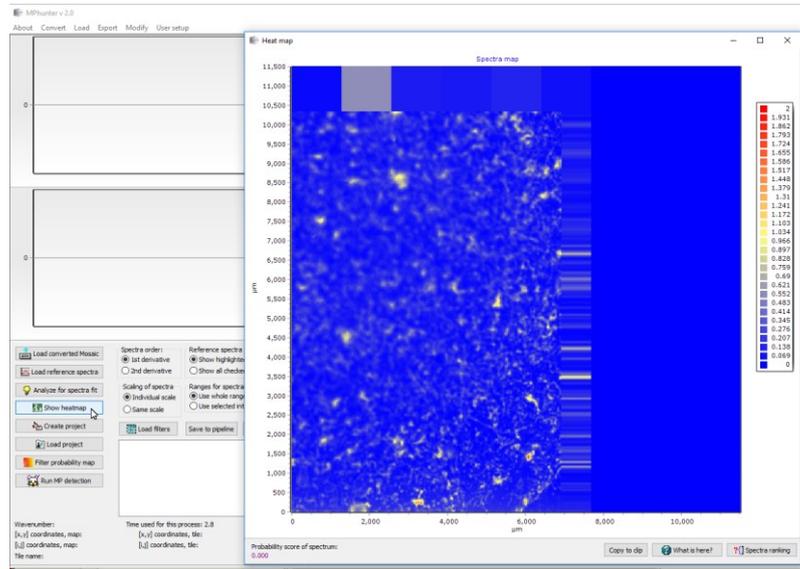
**Note, using an FTIR system without an FPA detector, e.g. The Thermo system Nicolet iN10**

simPLe

- Convert .jdx into .dx files, then convert .dx into .spe. While doing so MPH Hunter combines individual .jdx files into the shape of a 64x64 FPA format. At the right and upper edge, the last spectrum is repeated to fill up rows/ column. Do all steps with the MPH Hunter as indicated above. Before starting the APA/ MPAPP analysis remove these artefacts.



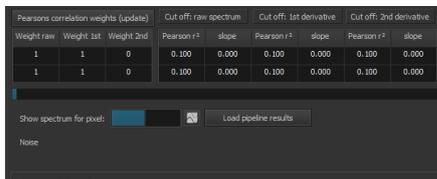
Grouping of .jdx files in dx files (64x64 spectra), that is the size of an FPA detector used by Bruker

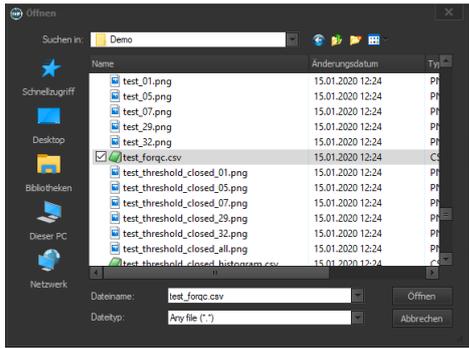


- This could be done using the software Origin, for this:
- Open – File – Import Wizard
- Choose file (howto.forqc.csv) – ‘Add file’ - ‘Ok’
- Mark column 3 and 4, right click ‘set as Z’

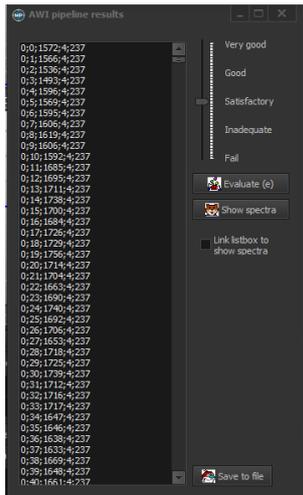
**QA/QC**

1. Load the file labeled “\_for\_qc” via the Load pipeline results function from Image Analysis:

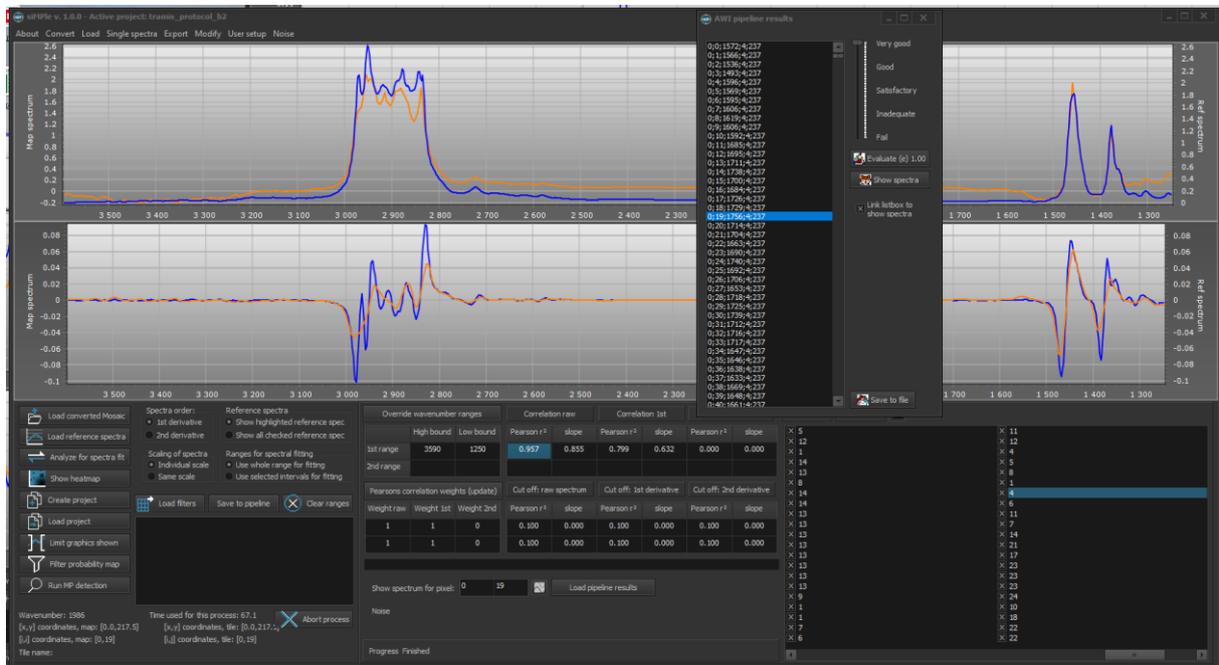




2. Activate “Link listbox to show spectra”:



3. Start QA/QC as described in the main document.

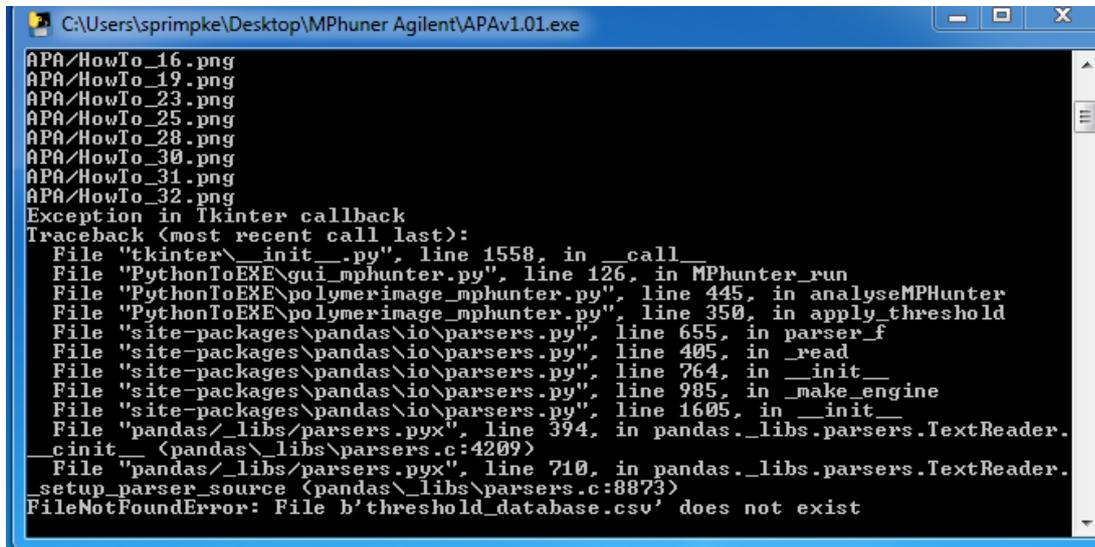


4. Click on “Save to file” so reassess the data later or for further evaluation.

## Troubleshooting:

### Error Messages (Traceback):

- a. Missing the threshold.csv file in the folder with the executable for APA/MPAPP:

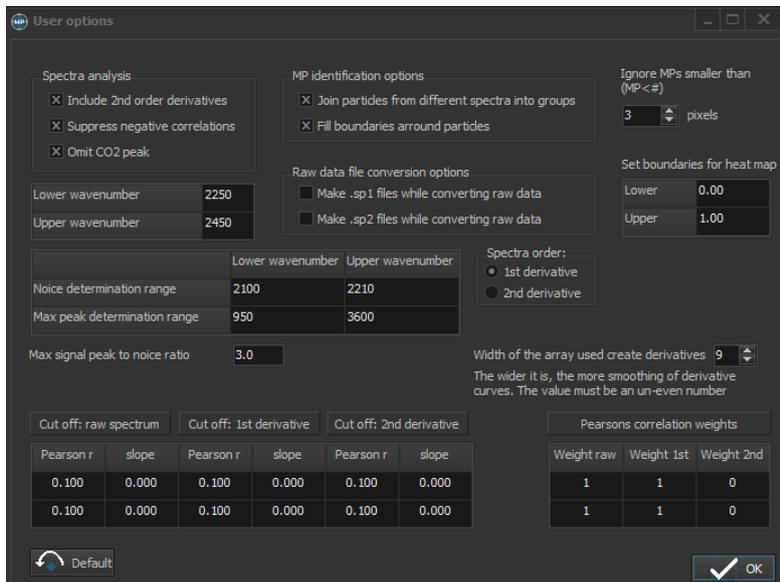


```
C:\Users\sprimpke\Desktop\MPHunter Agilent\APAv1.01.exe
APA/HowTo_16.png
APA/HowTo_19.png
APA/HowTo_23.png
APA/HowTo_25.png
APA/HowTo_28.png
APA/HowTo_30.png
APA/HowTo_31.png
APA/HowTo_32.png
Exception in Tkinter callback
Traceback (most recent call last):
  File "tkinter\__init__.py", line 1558, in __call__
  File "PythonToEXE\gui_mphunter.py", line 126, in MPHunter_run
  File "PythonToEXE\polymerimage_mphunter.py", line 445, in analyseMPHunter
  File "PythonToEXE\polymerimage_mphunter.py", line 350, in apply_threshold
  File "site-packages\pandas\io\parsers.py", line 655, in parser_f
  File "site-packages\pandas\io\parsers.py", line 405, in _read
  File "site-packages\pandas\io\parsers.py", line 764, in __init__
  File "site-packages\pandas\io\parsers.py", line 985, in _make_engine
  File "site-packages\pandas\io\parsers.py", line 1605, in __init__
  File "pandas\_libs\parsers.pyx", line 394, in pandas._libs.parsers.TextReader.
  cinit__ (pandas\_libs\parsers.c:4209)
  File "pandas\_libs\parsers.pyx", line 710, in pandas._libs.parsers.TextReader.
  _setup_parser_source (pandas\_libs\parsers.c:8873)
FileNotFoundError: File b'threshold_database.csv' does not exist
```

Please copy the provide threshold\_database.csv into the folder.

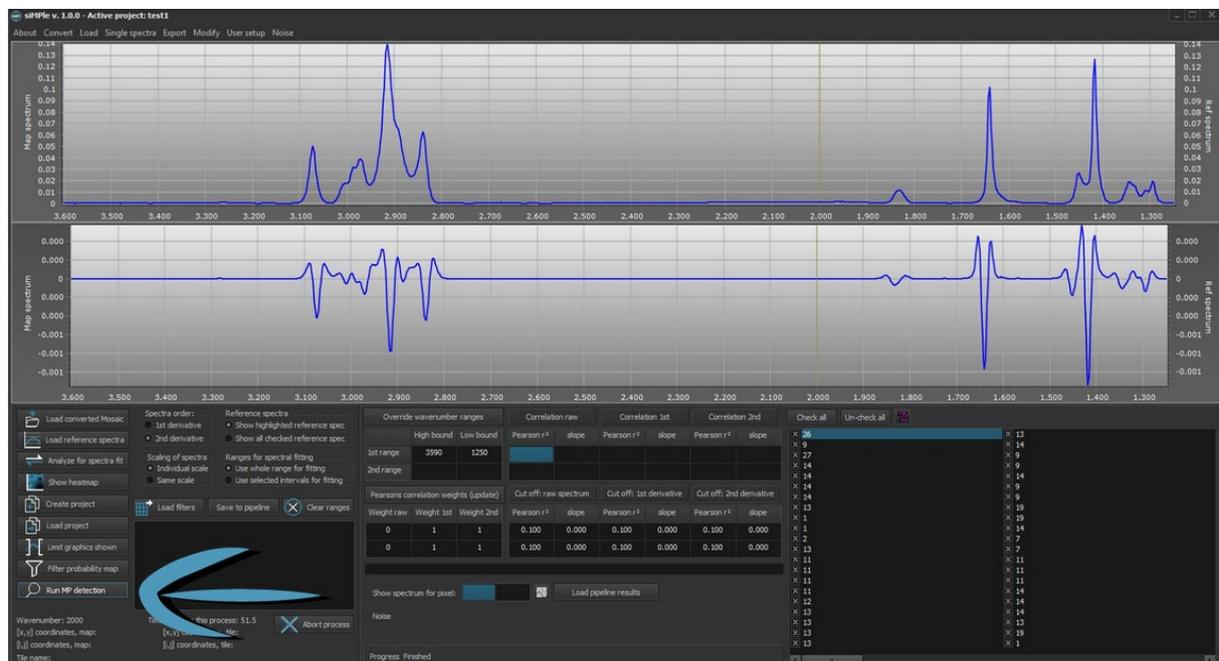
## The AAU pipeline:

The second pipeline uses the following user setting prior to spectral fit (see above for the procedure).

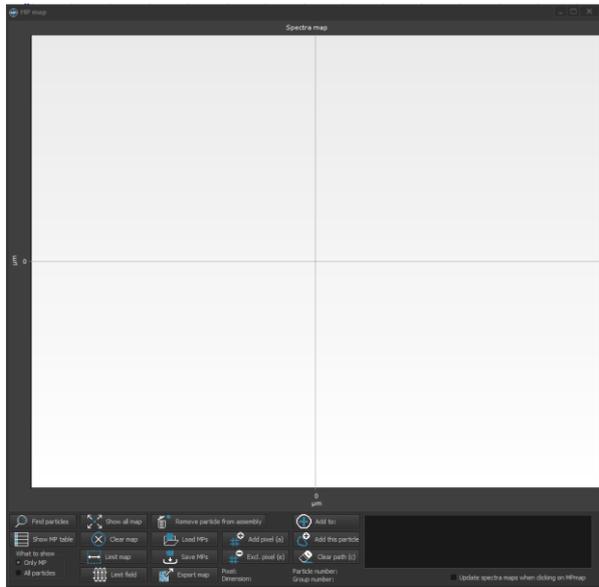


1. After spectral fit click on run MP detection

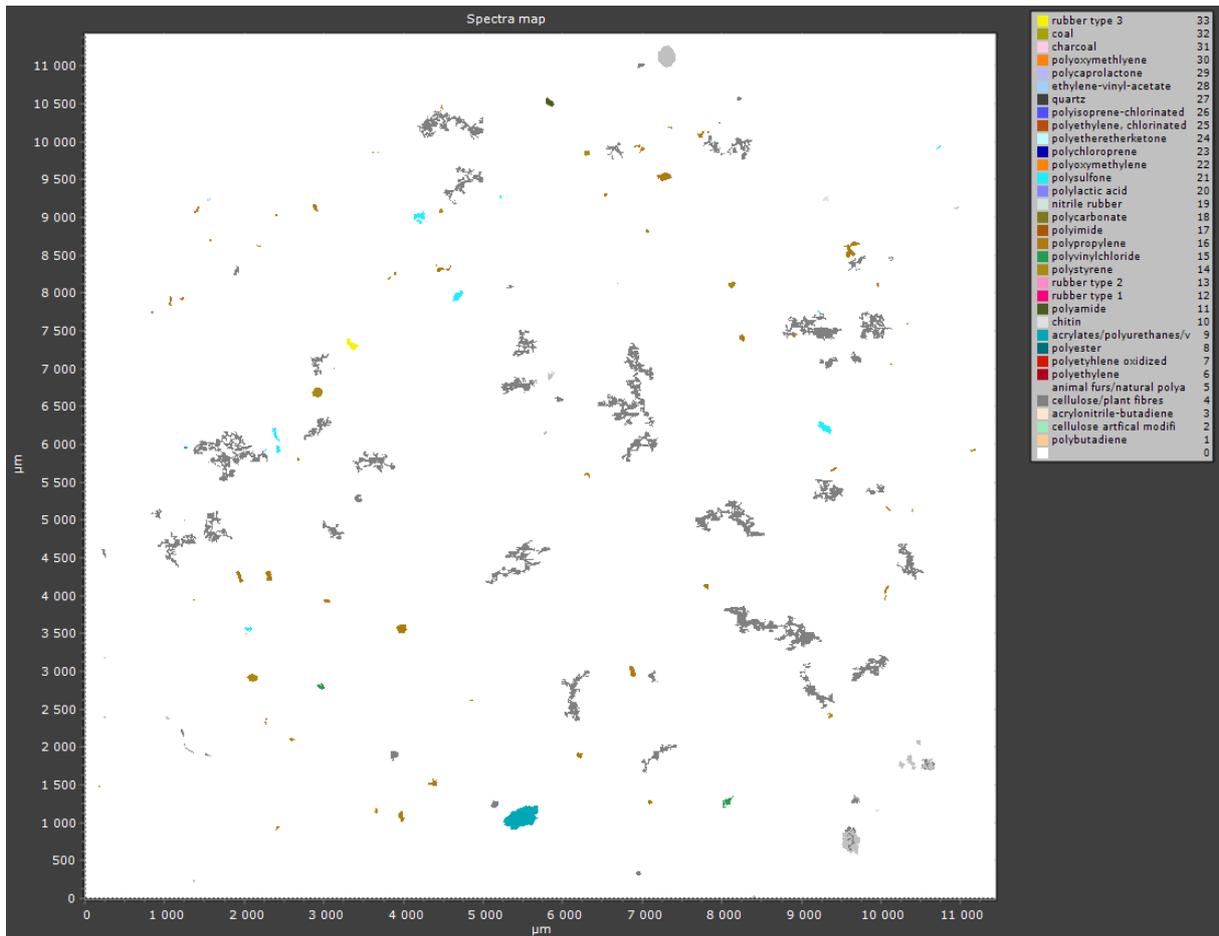
You can setup the probability thresholds before that: "Modify -> Modify loaded parameters for reference spectra"



2. Click on Run MP detection and afterwards Find particles in the following window:



3. This results in a false color image (here RefEnv1):



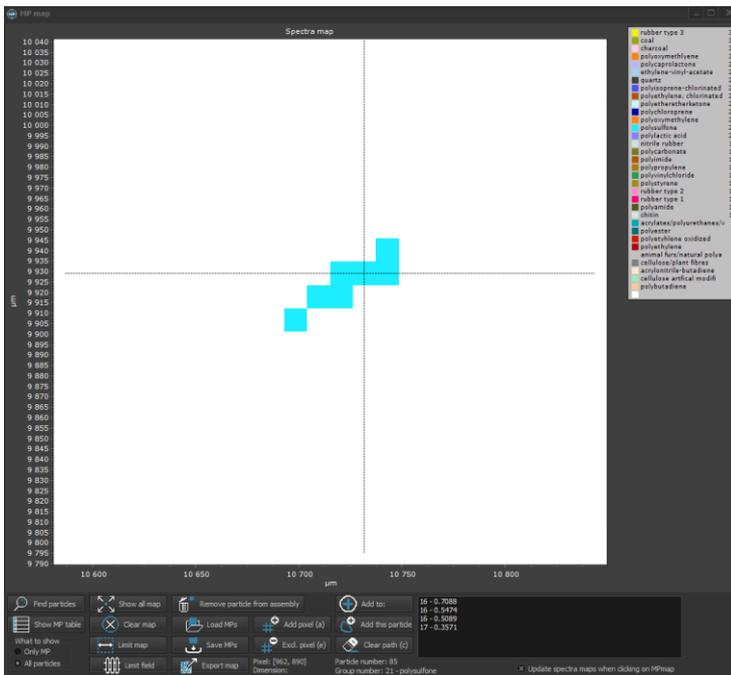
- All particles are listed via Show MP table containing all results which can be exported to .csv files for future analysis.

Show automatically detected MP particles

MP identifier	Coordinates [pixels]	Coordinates [µm]	Polymer group	Number of pixels [-]	Area on map [µm <sup>2</sup> ]	Major dimension [µm]	Minor dimension [µm]	Volume [µm <sup>3</sup> ]	Mass [ng]	Comments
Natural_1	[161,526]	[1795,5865]	cellulose/plant fibres	1223	152046.4	957.3	202.2	12299245	19063.830	
Natural_2	[775,320]	[8641,3568]	cellulose/plant fibres	1339	166467.8	1298.3	163.3	10870984	16850.024	
Natural_3	[819,675]	[9132,7526]	cellulose/plant fibres	747	92868.9	732.2	161.5	5999141	9298.667	
Natural_4	[862,75]	[9611,836]	cellulose/plant fibres	82	10194.4	294.4	44.1	179767	278.639	
Natural_5	[862,58]	[9611,647]	cellulose/plant fibres	5	621.6	36.1	21.9	5454	8.454	
Natural_6	[867,116]	[9667,1293]	cellulose/plant fibres	48	5967.5	126.8	59.9	143085	221.782	
Natural_7	[869,753]	[9689,8396]	cellulose/plant fibres	111	13799.8	282.0	62.3	343885	533.022	
Natural_8	[947,161]	[10559,1795]	cellulose/plant fibres	28	3481.0	129.4	34.3	47693	73.924	
Natural_9	[952,157]	[10615,1751]	cellulose/plant fibres	23	2859.4	176.9	20.6	23539	36.485	
Natural_10	[655,997]	[7303,11117]	al furs/natural polyan	382	47491.2	289.3	209.0	3970151	4565.673	
Natural_11	[863,66]	[9622,736]	al furs/natural polyan	374	46496.6	341.5	173.3	3223883	3707.465	
Natural_12	[92,213]	[1026,2375]	al furs/natural polyan	11	1367.5	51.4	33.9	18548	21.330	
Natural_13	[930,161]	[10369,1795]	al furs/natural polyan	86	10691.7	215.8	63.1	269778	310.245	
Natural_14	[950,158]	[10592,1762]	al furs/natural polyan	109	13551.2	183.4	94.1	509999	586.499	
Natural_15	[21,409]	[234,4560]	cellulose/plant fibres	22	2735.1	121.0	28.8	31496	48.819	
Natural_16	[145,438]	[1617,4884]	cellulose/plant fibres	394	48983.1	447.6	139.3	2730095	4231.647	
Natural_17	[834,829]	[9299,9243]	chitin	27	3356.7	90.8	47.1	63218	24.023	
Natural_18	[954,162]	[10637,1806]	chitin	4	497.3	26.9	23.5	4679	1.778	
Natural_19	[22,285]	[245,3178]	al furs/natural polyan	2	248.6	22.3	14.2	1412	1.624	
Natural_20	[220,393]	[2453,4382]	al furs/natural polyan	1	124.3	14.2	11.1	554	0.638	
Natural_21	[753,1]	[8396,11]	al furs/natural polyan	13	1616.2	58.5	35.2	22758	26.172	
Natural_22	[810,299]	[9031,3334]	al furs/natural polyan	7	870.3	46.4	23.9	8311	9.558	

What to report  
 Show only MP  
 Show all particles  
 Delete Particle  
 Max fit spectra output  
 Export particle-xy  
 Create a csv file  
 Save to MP file

- The particles can either be selected by clicking on the entry or on the MP map for QA/QC (Activate “Update spectra maps when clicking on MP map”)



Spectrum of the selected pixel of the identified particle.

