TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

MASTER THESIS

Evaluation of Microplastic Classifiers using Linear Hyperspectral Mixture Images

Submitted to the
Institute of Chemical Technologies and Analytics
Vienna University of Technology

under the supervision of

Ao.Univ.Prof. Mag. Dr. Johann Lohninger

by

Dieter Steiner

Identification No.: 01025618

Vienna, March, 2020

Dieter Steiner

Abstract

Hyperspectral images of microplastic samples can, under certain conditions, contain mixed spectra, for example when the resolution of the imaging instrument is lower than the particle dimensions. This work presents a tool for the evaluation of the performance of a microplastic classifier when faced with spectral mixtures and noisy data.

Linear hyperspectral mixture images comprising mixed spectra with varying amounts of added noise are created from two selected spectra and subsequently classified. The classified mixture images are translated into features that facilitate the analysis of the classifier under examination. Among these features are the parameters defining the logistic function, that fits the response values of a classified mixture image at a particular noise level.

Three random forest classifiers with different training sets and features were evaluated. Two of them use manually selected spectral descriptors (SPDCs) as features, while the last classifier works with automatically selected SPDCs. One of the first two classifiers, named experts classifier, was trained on a significantly larger training set compared to the other two.

The experiments conducted suggest that the features *"offset"* and *"midpoint"*, describing logistic function parameters, capture the behavior of a classifier applied to mixture images best. The results of the experiments covering the experts classifier revealed discrepancies in the performance of the individual binary classifiers. PAN showed much more rapid deterioration of the classifier responses with increasing noise level compared to ABS and PS. The classifiers using the manually chosen features showed comparable or better performance for mixtures of polymer and non-polymer spectra with the exception of EVOH and PE. Here the automatically selected features facilitate better separation despite similarities in the spectra of EVOH, PE and non-polymers.

Kurzfassung

Hyperspektrale Bilder von Mikroplastikproben können unter gewissen Umständen Mischspektren enthalten, beispielsweise wenn die laterale Auflösung eines Mikroskops nicht für die Erkennung der einzelnen Partikel ausreicht. Diese Arbeit präsentiert eine Methode zur Evaluierung der Performance von Mikroplastik-Klassifikatoren in Hinblick auf gemischte und verrauschte Spektren.

Lineare hyperspektrale Mischbilder enthalten unterschiedlich stark verrauschte Mischungen von zwei selektierten Spektren. Die Evaluierung der Klassifikatoren basiert auf verschiedenen Merkmalen, die aus den klassifizierten Mischbildern berechnet werden. Teil dieser Merkmale sind unter anderem die Parameter der logistischen Funktion die eine Zeile eines klassifizierten Mischbildes am besten approximiert.

Drei Random Forests mit verschiedenen Merkmalen und unterschiedlich großen Trainingssets wurden evaluiert. Zwei verwenden spektrale Deskriptoren (SPDCs) als Merkmale, die manuell ausgewählt wurden. Der dritte arbeitet mit einem Set aus automatisch selektierten SPDCs. Einer der Random Forests mit manuell gewählten Features wurde auf einem wesentlich größeren Trainingsdatensatz erzeugt als die anderen. Dieser wird folglich als ”Experten-Klassifikator” bezeichnet.

Die Merkmale *”offset”* und *”midpoint”* beschreiben Parameter logistischer Funktionen und konnten das Verhalten von Klassifikatoren bei der Anwendung auf Mischbilder am besten wiedergeben. Die Untersuchung des ”Experten-Klassifikators” zeigte Unterschiede in den einzelnen binären Klassifikatoren auf. So fiel die Antwort des PAN Klassifikators wesentlich schneller mit steigender Amplitude des Rauschens im Vergleich zu ABS und PS. Die auf den manuell selektierten SPDCs basierenden Klassifikatoren erzielten mit Ausnahme von EVOH und PE die besseren Ergebnisse. Bei letzteren konnten die automatisch ausgewählten Merkmale die Klassen EVOH, PE und Non-Polymer trotz Ähnlichkeiten in den Spektren besser voneinander unterscheiden.

I want to thank everyone who contributed to this thesis, first and foremost my supervisor Hans Lohninger for his guidance and patience.

Contents

Chapter 1

Introduction

Spectroscopy, the theory of the interaction of matter with electromagnetic radiation, is the foundation for many optical analytical instruments, including spectroscopes. A sample is illuminated with light of a specific energy region, for example ultraviolet (UV), visible or infrared (IR) light. After different interactions of the electromagnetic radiation with the matter of the sample, the transmitted or reflected light is detected in form of a the spectrum and used to gain qualitative and quantitative information. These spectra describe the absorbance or reflectance of the sample in dependence of frequency or wavenumber. The spectra treated in this work are absorption spectra measured in the infrared region and show peaks or absorption bands characteristic for the investigated polymers. This facilitates the classification and detection of polymers on the basis of their infrared spectra.

Images generated by spatially scanning a sample in two dimensions and acquiring a spectrum for every pixel are called hyperspectral images (HSI). They feature two spatial and one frequency dimension. If particles of different polymers are spaced closely together such that they occupy the same pixel in the associated HSI or the particles overlap, mixtures of the spectra belonging to the members are observed for the related pixel. These mixed spectra and noisy data can deteriorate the classification performance and particle detection of microplastics.

In order to investigate the effects of mixed spectra and noise in the data on the classification performance, a tool is developed during the work of this thesis, that creates and analyzes mixture images. The concept of such linear hyperspectral mixture images is provided in figure 1.1. They comprise along the x-axis linear mixtures of the two clean spectra they are generated from and located themselves in the bottom corners of the mixture. Along the y-axis increasing amounts of noise are added to the linear mixtures of the two reference spectra. Application of a classifier on a mixture image yields a matrix with the same dimensions. Its entries reflect the classifiers response to the mixed spectrum located at the related pixel. The analysis of these classified mixture images provide insights into the

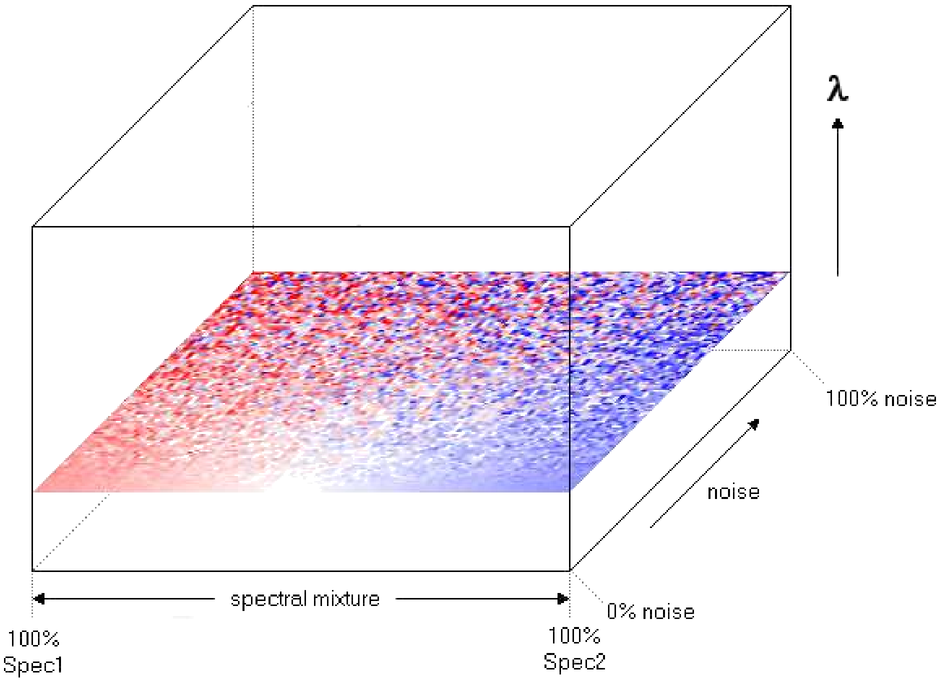


Figure 1.1: Concept of a spectral mixture image. Along the x-axis are mixtures of two selected spectra, here Spec1 and Spec2. Noise is added in increasing amounts along the y-axis [Epi20].

classifiers behaviour in regards to mixed and noisy spectra. Multiple mixture images are created for the analysis of a single combination of two classes, say polyethylene (PE) versus polypropylene (PP), to get more general perceptions about the classifier and avert possible negative implications of mixture images created from outliers.

The results of the tool are finally evaluated for different random forest classifiers. One was created outside the scope of this work through the use of chemical knowledge about the underlying problem of polymer IR spectroscopy, while another random forest is created automatically. For this purpose spectroscopic features are generated in huge numbers, which are subsequently narrowed down to a compact set of a few descriptive features. Lastly their performance is discussed on the basis of the experiments utilizing artificially created aqueous microplastic samples.

Chapter 2

Theory

2.1 Spectroscopic Fundamentals

2.1.1 Fourier Transformed Infrared Spectroscopy

Infrared (IR) spectroscopy is an analytical technique, where the absorption, reflectance or transmission of infrared light of an analyte is measured and used for quantitative and qualitative analysis. Electromagnetic waves, including infrared light, can be characterized by frequency ν or wavelength λ. In infrared spectroscopy light is usually described by its wavenumber $\bar{\nu} = \frac{1}{\lambda} = \frac{\nu}{c}$, where c is the velocity of light. Therefore the wavenumber is defined as the number of wavelengths per unit distance, which is typically provided in centimeters. The energy spectrum occupied by infrared radiation ranges from near IR starting at 0.78 µm or $12\,800\,\text{cm}^{-1}$ to the end of far IR at 1 mm or $10\,\text{cm}^{-1}$ [SHC16]. The lower end of far IR is also called terahertz (THz) radiation. In this work part of the middle IR spectrum with $3500\,\text{cm}^{-1} > \bar{\nu} > 1250\,\text{cm}^{-1}$ is used.

A molecule can absorb radiation by three basic processes. Depending on the energy of the absorbed photons, rotational, vibrational or electronic transitions are induced in the energy states of the molecule. Electronic transitions require photons with energy higher than IR, e.g. visible light, ultraviolet and even higher frequencies. Infrared light therefore causes rotational and vibrational transitions. For a molecule to absorb IR radiation, it must undergo a change in its dipole moment, otherwise the molecule can not interact with the electrical field of the IR radiation. Absorption occurs if the frequency of the radiation matches the frequency of a molecular vibration or rotation [CDS13]. Since diatomic molecules that have no permanent dipole moment, e.g. N≡N can neither express one by rotation or vibration, such molecules can not interact with and therefore do not absorb IR radiation at all. Triatomic molecules with no permanent dipole moment like O=C=O on the other hand, may exhibit a dipole moment through asymmetric vibration and hence are IR active. Rotational and vibrational states are both quantized and would result in sharp absorption maxima in the measured spectrum, but interactions with surrounding groups spread the sharp absorption peaks in solids and liquids, causing broadened vibrational bands. The energies for purely rotational transitions correspond to the far-infrared region with $\lambda > 100\,\mu\text{m}$ and vibrational transitions match the mid-IR with $2.5\,\mu\text{m} < \lambda < 50\,\mu\text{m}$ or $4000\,\text{cm}^{-1} > \bar{\nu} > 200\,\text{cm}^{-1}$ in wavenumbers [SHC16]. Figure 2.1 shows the most important modes of molecular vibrations present in polymers by means of a CH_2 chain. Stretching vibrations periodically vary the bond length, while deformation or bending vibrations (scissoring, wagging, rocking, twisting) have oscillating bond angles [RD17].

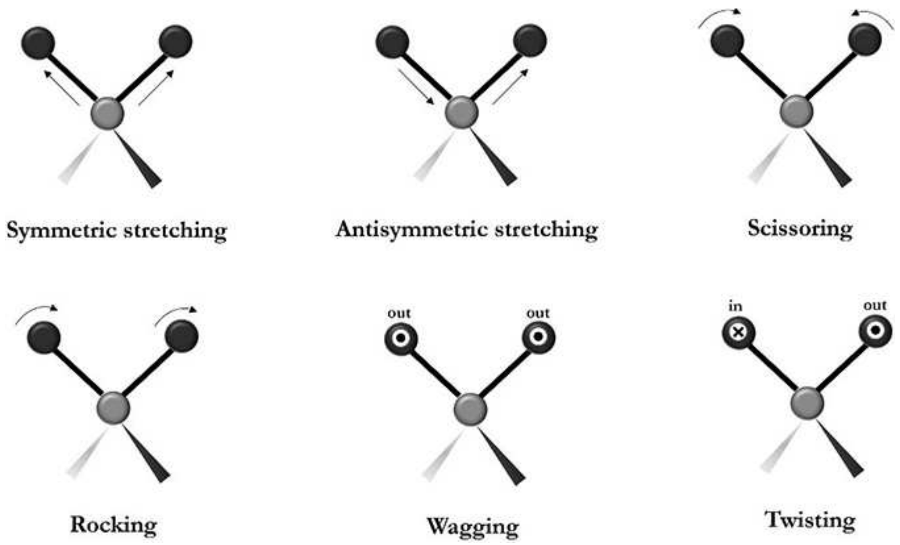


Figure 2.1: Different vibrational modes for a $-CH_2-$ group. © 2016 João Cajaiba Da Silva, Alex Queiroz, Alline Oliveira and Vinícius Kartnaller. Originally published in [Sil+17] under CC BY 3.0 license. Available from: `https://doi.org/10.5772/65552`.

IR spectrometers

A traditional dispersive spectrophotometer with a grating monochromator consists of a source, providing continuous radiation over the spectrum of interest, a monochromator, dispersing the light and guiding only a narrow band of wavelengths through the sample and a detector, capturing the radiation passed through or reflected by the sample. These conventional IR spectrometers have been largely replaced by Fourier transform infrared (FTIR) spectrometers [CDS13]. Figure 2.2 schematically shows the principal components of an FTIR instrument. It uses an interferometer to split the beam along two paths, one of them subjected to a phase

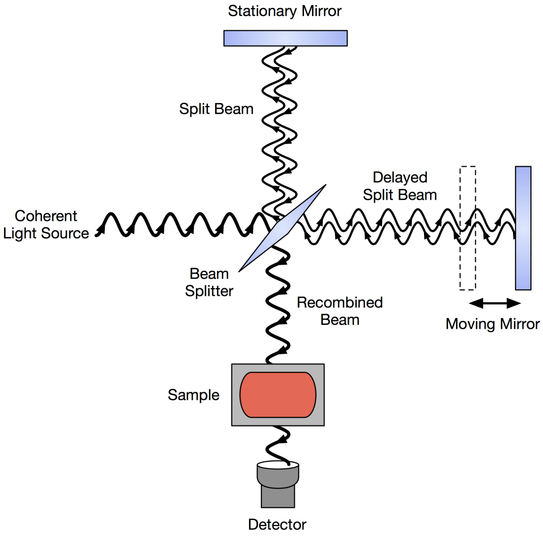


Figure 2.2: Fourier transform infrared spectrometer setup consisting of an IR light source, an interferometer, the sample and a detector.

shift with a movable mirror. The sample is illuminated with the interference pattern created by the superimposed beams joined again after the semi-transparent mirror paths. In dependence of the position of the movable mirror certain frequencies are eliminated from the radiation due to destructive interference and are therefore not part of the illumination of the sample. The position of the moveable mirror is varied and the light captured at the detector is scanned and recorded together with the mirror position. From the whole dataset, containing values for a wide range of mirror positions, the absorption of the sample in dependence of the wavelength, can be calculated. The raw spectrum acquired at the detector through scanning of the mirror position is called interferogram and can be converted into frequency domain by Fourier transformation. With this setup a single spectrum can be obtained much faster compared to conventional IR spectroscopy. A second advantage is the increased signal-to-noise ratio achieved by a higher throughput in terms of radiation, meaning less attenuation through fewer optical elements between source and detector [SHC16].

FTIR Microscopy

In order to create images of spatially coherent FTIR spectra analogous to optical microscopy, the sample is segmented into a grid and the absorption of each pixel has to be measured. An early method involved the restriction of the illumination in the sample plane to each point of interest and subsequent detection of the related attenuation [LB05]. In a more recent approach, used for the acquisition of the data examined in this work, the segmentation takes place in the detection plane, where not a single, but a 2D array of detectors, a so called focal plane array (FPA), records thousands of spectra simultaneously [Löd+15; LB05]. If, for example, a FPA with 64 × 64 pixels is used to create an FTIR image with a resolution of 1024 × 1024, $256 = 16^2$ separate adjacent views of the sample are subsequently irradiated and the attenuated radiation imaged onto the FPA detector. A significant improvement compared to the individual measurement of all 1024^2 spectra.

2.1.2 Polymer Spectroscopy

Polymers are large molecules consisting of many smaller substructures called monomers. There are polymers of natural origin, for example DNA, RNA, Cellulose, etc. and of synthetic nature, like plastics. The focus in this work lies on plastics. They can be categorized into plastics or thermoplastics, thermosets, elastomers, fibers and paints and coatings. Plastics (e.g. Polyethylene (PE), polypropylene (PP), polyvinyl chloride (PVC) and polystyrene (PS)) can be shaped under heating and reshaped a number of times if heated again. Thermosets are heavily cross-linked after being hardened into their final form. Polyester and polyurethanes (PU) are common examples of thermosets. Elastomers are flexible materials used for tires, damping and insulating elements, etc. Among the polymers used as fibers are nylon made of polyamide (PA) or polyacrylonitrile (PAN). Paints and coatings consist of polymers that can form a film [SV13].

All of these polymers contain IR active groups. Figure 2.3 shows reference spectra taken from the ImageLab Database [Epi13] and spectra measured from artificially generated samples (see section 4) of PE and PP. The plots of the measured spectra show the absorption of infrared light of the sample. As can be seen in the figure, different polymers with unlike functional groups also express differing characteristics in their spectra. The database spectrum of PE shows two major absorption bands around $2900\,\text{cm}^{-1}$ and one peak roughly around $1450\,\text{cm}^{-1}$. In the PP spectrum visible in figure 2.3(b) on the other hand 4 close peaks around $2900\,\text{cm}^{-1}$ can be identified as well as two more at approximately $1450\,\text{cm}^{-1}$ and $1375\,\text{cm}^{-1}$. The distinction of these polymers is also easily possible when looking at the measured spectra at the bottom of the visualization. Table 2.1 contains frequencies and assignments of absorption bands of 5 selected polymers according to Noda et al. [Nod+07] and their structural formulae. The comparison of the graphs and the table reveals the exact wavenumbers of the aforementioned peaks and the group and vibration causing it. The two peaks at high wavenumbers in the PE spectrum for example, are due to the symmetric stretching of the CH_2 groups, while their bending vibrations are responsible for the third peak.

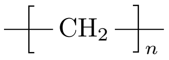
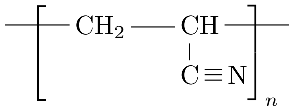
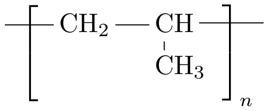
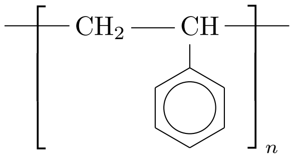
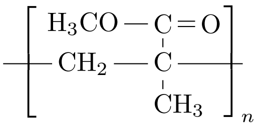
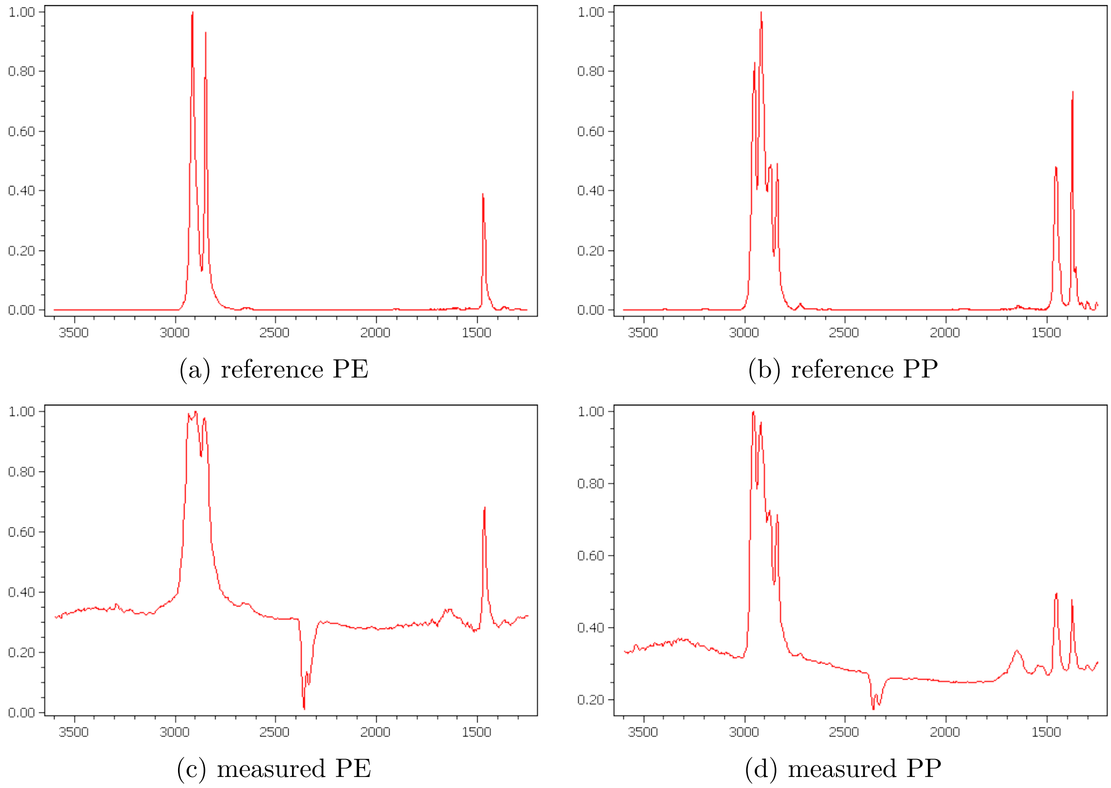
Polymer	Frequency cm^{-1}	Assignment
Polyethylene (PE) 	1303	CH_2 wag (gtg conformation)
	1353	CH_2 wag (gtg conformation)
	1368	CH_2 wag (gtg conformation)
	1463	CH_2 bend
	1473	CH_2 bend
	2850	CH_2 symmetric stretch
	2918	CH_2 symmetric stretch
Polyacrylonitrile (PAN) 	1247	CH bend
	1358	CH_2 asymmetric bend
	1452	CH_2 bend
	2241	$C{\equiv}N$ stretch
	2243	$C{\equiv}N$ stretch
	2870	CH_2 symmetric stretch
	2940	CH_2 asymmetric stretch
Polymethylmethacrylate (PMMA) 	1241, 1277	$C-O$ stretch
	1387	CH_3 symmetric (umbrella) deformation
	1448	CH_2 symmetric (scissors) deformation, $O-CH_3$ deformation
	1485	CH_3 (α-methyl) asymmetric deformation
	1732	$C=O$
	2850	Overtone of ester CH_3 deformation
	2950	CH_3 (ester methyl) stretch
	2998	CH_3 (α-methyl) stretch
Polystyrene (PS) 	1220	Ring in-phase CH bend
	1454	Ring semicircle stretch + CH_2 symmetric (scissors) deformation
	1494	Ring semicircle stretch
	1601	Ring quadrant stretch
	2850	CH_2 symmetric stretch
	2924	CH_2 asymmetric stretch
	3000 ∼ 3100	Aromatic CH stretch
Polypropylene (PP) 	1256	CH bend + CH_2 twist + CH_3 rock
	1377	CH_3 symmetric bend + CH_2 wag
	1458	CH_2 scissors
	2837	CH_2 symmetric stretch
	2868	CH_3 symmetric stretch
	2919	CH_2 asymmetric stretch
	2951	CH_3 asymmetric stretch

Table 2.1: Frequencies in wavenumbers and assignments for molecular vibrations of 5 common polymers according to Noda et al. [Nod+07]. Only bands in the frequency range detected by the instrument, the samples used in this thesis were measured with, $3600\,\mathrm{cm}^{-1}$ to $1250\,\mathrm{cm}^{-1}$, are listed.



(a) reference PE

(b) reference PP

(c) measured PE

(d) measured PP

Figure 2.3: The top sub figures show reference spectra of Polyethylene (PE) (a) and polypropylene (PP) (b). Below are spectra from artificially created samples with microplastic particles mixed with a freshwater plankton sample. (c) shows PE and (d) shows PP. The data of all plots is normalized and shows absorption vs. wavenumber (cm^{-1}).

In contrast to the spectra taken from the Imagelab database the two bottom plots exhibit two negative peaks between $2400\,\mathrm{cm}^{-1}$ and $2300\,\mathrm{cm}^{-1}$. They arise from the different amount of IR radiation absorbed in the two split beams of the instrument in case of contamination with IR active carbon dioxide from the surrounding air. The sign of the amplitude of the CO_2 peak is plainly determined by whichever optical path is exposed to more CO_2. Another less noticeable difference can be seen at the peak with the highest wavenumber in the measured PE spectrum. It shows signs of total absorption or overloading. Generally the thicker the sample is, the more light it absorbs as described by Lambert-Beer's law. After a few tens of micrometers almost all light is absorbed in the sample. Therefore even thicker particles do not produce higher absorbances which leads to sharp peaks turning into a plateau of random noise. The most prominent contrast between the top and bottom spectra however is the non-zero baseline in the measured spectra. This effect arises due to scattering and partial reflectance. Scattering leads to linear or parabolic baselines and partial reflectance causes cosine-shaped baselines [RD17]. Additionally, Mie scattering occurring when electromagnetic radiation interacts with particles in the size of the radiation wavelength further impacts the baseline and induces even slight shifts in the positions of the absorption bands [Bas+09]. The resonant Mie scattering is caused by the change in the scattering efficiency, depending on the particle diameter, the wavelength of the illumination and the refractive index of the particle.

This adverse effect can be overcome by highly computationally expensive operations modeling and subtracting the baseline or through the usage of so called spectral descriptors introduced in the following section.

2.2 Hyperspectral Images

A grayscale image with n_r rows and n_c columns has $n_r \cdot n_c$ pixels. Every single pixel has a gray value, where low values present dark and high values bright pixels. This image can be interpreted as 2D matrix, where the entries are the gray values. In colored photographic images every pixel holds a value for red, green and blue. This can be represented by three grayscale images and three 2D matrices, which is just another representation of a 3D matrix with the third dimension having size $n_l = 3$, meaning three layers, one for each color red, green and blue. If we move from the three very broad wavelength bands representing the colors to a higher number of narrower regions, e.g. 50 nm wide intervals between 400 nm and 800 nm, we talk about multispectral images, where each pixel holds the values of multiple, in this case 8, wavelength bands [GGB07]. Hyperspectral images (HSIs) contain even more variables. According to Geladi, Grahn and Burger [GGB07] hyperspectral images are characterized by having many wavelength bands, often more than 100 and by the possibility to represent a pixel as a spectrum with spectral interpretation, spectral transformations, spectral data analysis, etc. Wang and Zhao [WZ15] emphasize, that hyperspectral images have high spectral resolution, enabling it to solve many problems unsolvable by multispectral imaging and that adjacent bands are correlated. HSIs are stored as three-dimensional data matrices often called hypercubes. Hyperspectral images can even feature a fourth dimension representing the time dependent changes of the image.

Figure 2.4 shows a hyperspectral image of a piece of meat. The top shows the different layers or images of adjacent wavelength bands. The graph in the bottom shows the spectra belonging to two pixels showing fat and lean meat. A single layer in the top stack of gray value images relates to a specific wavelength in the bottom plot, depicted by the vertical dashed line.

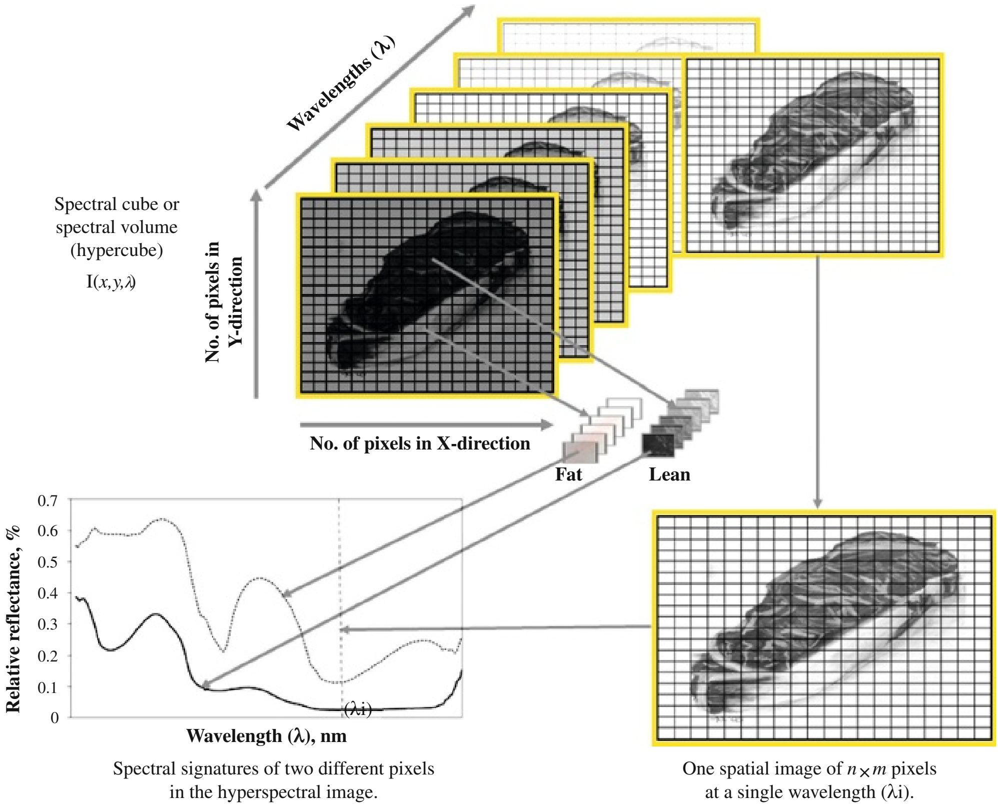


Figure 2.4: Hyperspectral image of a piece of meat taken from ElMasry and Sun [ES10]. The figure shows the relation between spectral and spatial information.

2.2.1 Spectral Descriptors

The hyperspectral images analyzed in this thesis contain a total of 609 layers, describing the spectrum from approximately 3595 cm^{-1} to 1250 cm^{-1}. The 609 input variables span a feature space with 609 dimensions. Calculations in this huge space are very time intensive and prone to error. Reduction of the dimensionality majorly reduces computation time and also decreases complexity and increases the generalization of supervised learning models (see section 2.3) [DL97]. There are many possibilities to reduce the size of the feature space. Here spectral descriptors (SPDCs), transformations of the 609 raw input variables are used. They reduce the effective dimensionality and include chemical knowledge at the same time.

Figure 2.5 motivates the usage of spectral descriptors. The top and bottom row show the Raman spectra of two pixels of the hyperspectral image displayed in-between. The upper left is labeled L1 and the pixel located at the center of the right edge L2. The spectrum of L1 features a peak around 2917 cm^{-1}, while the same region of the spectrum of L2 shows similar, but apparently random intensities. The images in the middle line visualize from left to right the intensity at $2917\,\mathrm{cm}^{-1}$, the integral of the spectrum over the area indicated in the graphs above and below, and the value of the correlation between the spectrum and the triangle peak illustrated as red line in the related plots. The image displaying the raw intensity of the selected wavelength or input variable shows high values around L1 and intermediate values around L2, even though the spectrum of L2 does not hold any identifiable spectral information at this wavenumber. The intensity at L2 is caused by a non-negative baseline and noise. Using the area under the spectrum as feature most of the background is successfully rejected, though L2 and other spectra still show non-zero values. The triangle template correlation (TC) spectral descriptor, however, almost completely removes the false positives including L2, since the random noise shown in the bottom graphs has no correlation with the triangle template peak, regardless of the height of the baseline and the amount of noise. The rightmost HSI demonstrates the capacity of the TC descriptor, that enables the identification of spectrum L1 and rejection of all other spectra from this single, transformed variable.

The next figure 2.6 lists four, in *ImageLab* commonly used, spectral descriptors. The SPDC in (a), referred to as ABL in *ImageLab*, calculates the area under the curve and subtracts the area below a straight line. The resulting area is intended to approximate the baseline-corrected integral of the spectrum in the specified interval. The ARW descriptor calculates the raw area below the spectrum in the selected region without baseline correction. The descriptor DV1 (b) approximates the first derivative and is calculated by a smoothed spectrum around the wavenumber of interest. The first derivative is unaffected by constant baselines. (c) and (d) describe the triangle template correlation descriptor TC and the inverse triangle template correlation descriptor TCI. They describe the correlation between the spectrum and a triangle template in the interval defined by the edges of the triangle, visualized by dashed vertical lines in the respective plots. The response is weighted by the approximated baseline-corrected integral of the spectrum, calculated like (a), favoring larger peaks with higher values.

Table 2.1 lists reasonable candidate positions for spectral descriptors specific for the polymers PE, PAN, PMMA, PS and PP. The type of the selected SPDC depends on the width and form of the absorption band. For sharp, narrow peaks the best choice is usually a TC descriptor.

2.2.2 Spectral Mixtures

Under certain conditions the spectrum belonging to a single pixel can consist of a mixture of possible multiple other spectra. This effect may arise from spectroscopy of polymer particles to remote sensing if either

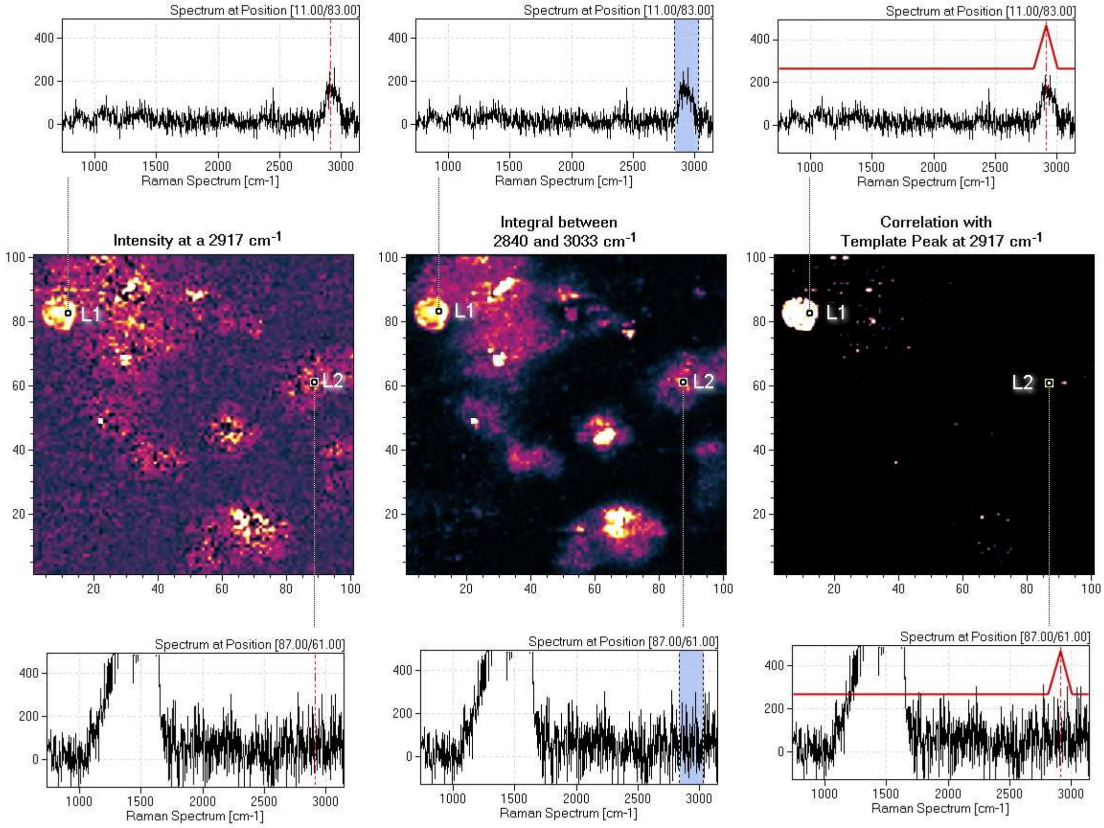


Figure 2.5: Comparison of raw intensity, integrated intensity and correlation of the spectrum with a triangle peak for two pixels of a HSI taken from Lohninger and Ofner [LO14]. The morgenstemning color map, used in the images, illustrates values from low to high with transitions of the colors from black, purple and yellow to white.

- the resolution of the imaging instrument is low enough, so that particles of different polymers contribute to the signal of a single pixel,
- two or more different particles physically overlap and the measured, transmitted light passes through and interacts with both particles subsequently
- or the particle in the measured pixel is itself a blend of different polymers.

In remote sensing point one refers to different members, e.g. water bodies, fields, urban areas, forests, etc. occupying a single pixel and an example for the last point would be sparse trees on a plain. A common task in remote sensing involves the decomposition of these mixed spectra into a collection of pure spectra, called endmembers and their abundance fractions [ND05; KM02]. These endmembers

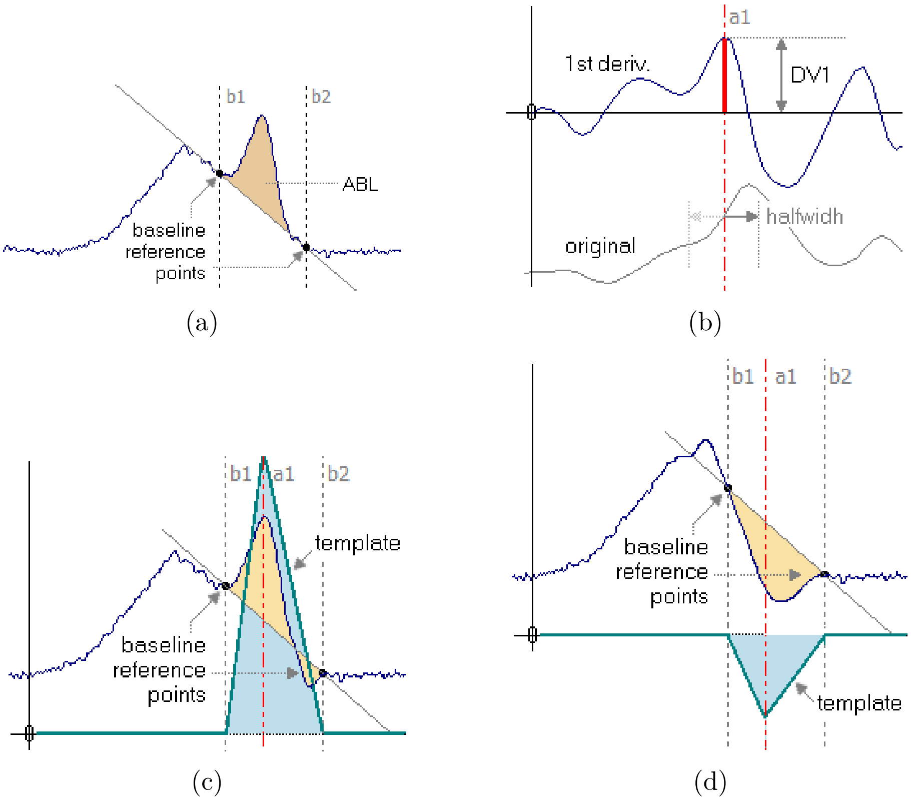


Figure 2.6: 4 types of spectral descriptors available in ImageLab [Epi20]. (a) ABL – baseline-corrected area between b_1 and b_2 (b) DV1 – first derivative (c) TC – weighted correlation with a triangle spanned by $(b_1, 0)$, $(a_1, 1)$ and $(b_2, 0)$ (d) TCI – same as TC with triangle tip pointing downwards $(a_1, -1)$

are calculated under the assumption of either a linear mixture model or non-linear models. Linear models apply if the mixing scale is macroscopic and the individual incident light rays interact with a single material. The mixing takes place in the sensor, measuring the light beams originating from different members because the detectors resolution is too low. Non-linear mixture occurs due to light being scattered by multiple materials and eventually reaching the sensor [Bio+12]. Even though linear mixture models only strictly hold if the endmembers are segregated and light interacts only with a single material, it is considered as acceptable approximation [Bio+12; KM02].

2.2.3 Hyperspectral Linear Mixture Images

In order to investigate a classifiers response to mixed spectra and noisy data, new evaluation data is created from two selected polymer spectra. The newly created data is arranged in a hyperspectral image with 101 × 101 pixels covering $10201 = 101^2$ spectra. Figure 2.7 visualizes a single layer of such a hyperspectral

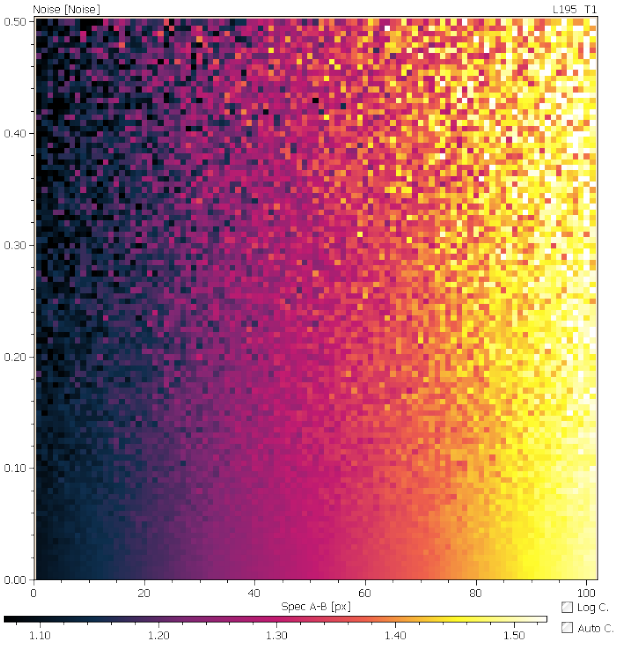
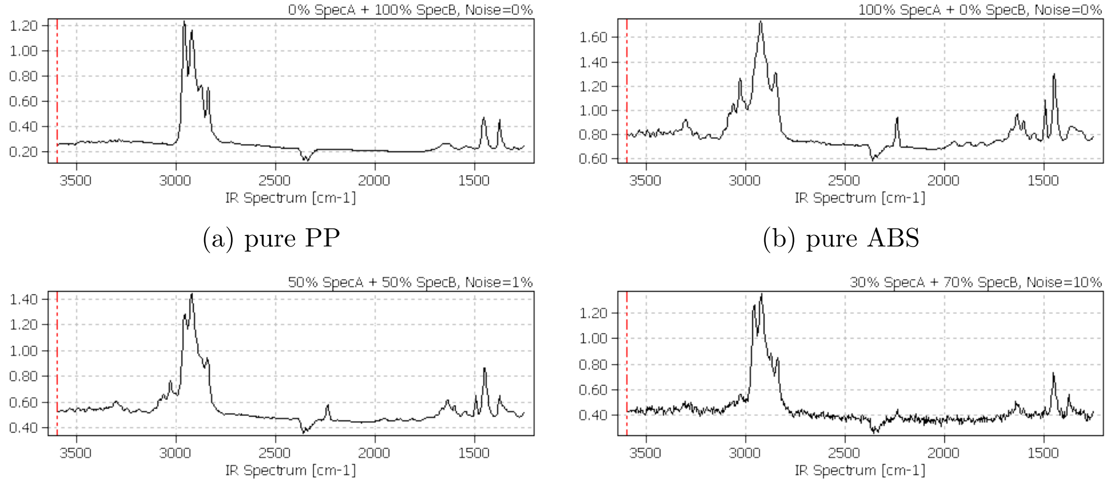


Figure 2.7: Mixture image showing the linear mixture of two reference spectra with noise increasing along the y-axis. The bottom corners hold the reference spectra and pixel in-between their mixture with varying amounts.

linear mixture image. The pixels in the bottom left and right corners are the two selected, pure polymer spectra used to create the whole hyperspectral image. The pixels in the bottom line in-between are calculated as linear mixtures of the corner spectra according to $s_{mix}(x, 0) := xs_{p_1} + (100 - x)s_{p_2}$, where s_{p_1}, s_{p_2} are the selected reference spectra and $x \in \{0, 1, \ldots, 100\}$ the zero-based x-coordinate. The lines above hold the same mixtures, but have increasing amounts of normally distributed noise added to them. The total amount of noise added in the top line can be controlled by means of a parameter. The spectrum of each pixel can therefore be described as $s_{mix}(x, y) := (xs_{p_1} + (100 - x)s_{p_2}) + \frac{y}{100}\epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and $y \in \{0, 1, \ldots, 100\}$ being normally distributed noise and the y-coordinate, respectively. σ^2 depends on the selected amount of total added noise and the amplitude of the spectra (heteroscedastic noise). The mixture in figure 2.7 was created from a polypropylene (PP) and a acrylonitrile butadiene styrene (ABS) spectrum. The difference in the color gradient along the x-axis for the different rows indicate the varying amounts of noise added, hence smooth transitions in the bottom row and increasingly discontinuous steps above.

Figure 2.8 contains the plots of two measured polymer spectra in (a) and (b)



(a) pure PP

(b) pure ABS

(c) mixture containing $\frac{1}{2}$ PP and $\frac{1}{2}$ ABS with 1% of added noise

(d) 30% ABS and 70% PP with 10% added noise

Figure 2.8: Four spectra of the mixture image 2.7, where the x and y coordinates in the image are described by the percent of *SpecA* contained in the mixed spectrum and percent of *Noise* added respectively, written atop the sub figures. (a) and (b) show the unmixed reference spectra and (c) and (d) show the spectra of two of the $10199 = 101^2 - 2$ mixed spectra in a mixture image.

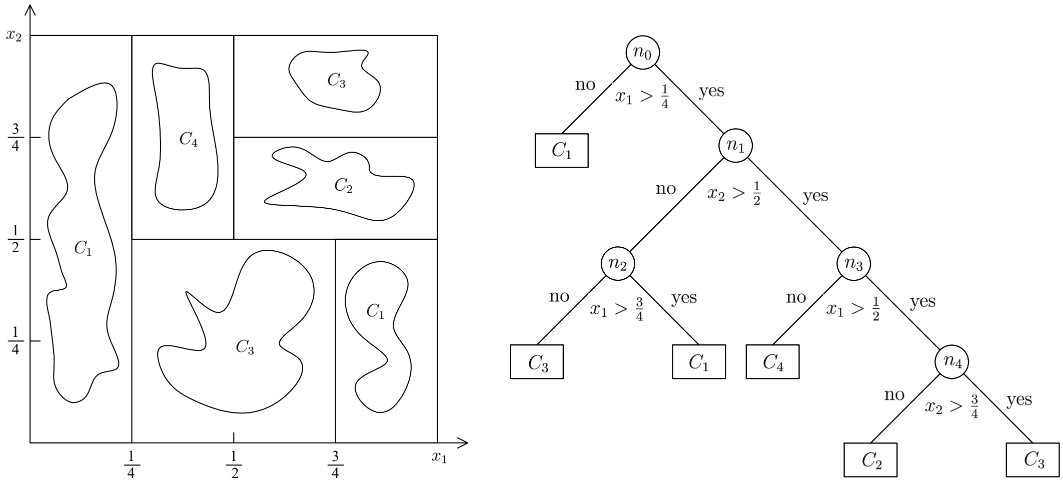
used to create the hyperspectral mixture image above and two of their mixtures in (c) and (d). The pure, unmixed spectra are located at the lower corners of the mixture image at position $(0, 0)$ and $(100, 0)$. The mixture made up of equal amounts of PP and ABS (c) can be found at $x = 50$ and $y = 1$, since it consists of 50% PP and contains 1% added normally distributed noise. The spectrum shows features describing both polymers, for example the highly descriptive peak around $2242\,\mathrm{cm}^{-1}$ caused by the C≡N group still exists, with lower amplitude though. Analogously the pixel coordinates of the spectrum seen in sub figure (d) are $(x, y) = (30, 10)$ in a ABS-PP mixture image.

2.3 Image Classification with Random Forests

Supervised learning is the category of problems, where on the basis of training data $(\mathbf{x}, y(\mathbf{x}))$ with input $\mathbf{x}$ and known target values or labels $y(\mathbf{x})$, a prediction of the unknown target function $y : A \to B; \mathbf{x} \mapsto y(\mathbf{x})$ is calculated. With this predicted target function $\hat{y} : \mathbf{x} \mapsto \hat{y}(\mathbf{x})$ a label can be assigned to new, hitherto unseen and unlabeled data. If the codomain B of y consists of discrete values, for example class labels like different polymers $\{\text{PAN, PE, PP}, \dots\}$ the task of learning the target function y is called classification. If the codomain is continuous the task is termed regression. Thus classification is the task of finding or training a model or function based on training data, to predict the labels of new, unseen data. Examples for classification are the recognition of handwritten digits or letters using photographic images or the identification of microplastic particles using spectroscopic data. A regression example is the prediction of the yield in a chemical process using the concentrations, temperature, etc. as input.

2.3.1 Decision Trees and Random Forests

The definition of a random forest given by Breiman, who first used the name ”random forest” states: ‘A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(\mathbf{x}, \Theta_k),\ k = 1, \dots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\mathbf{x}$.’ ([Bre01]) Generally random forests refer to the training method and the collection of classification trees obtained through this training. Each tree is independent identically distributed and has high variance and, ideally, low bias. Since the collection of tree is independent and have low bias, averaging over those trees greatly reduces the variance of the entire forests.



(a) A partition of the input space, resulting in four not connected subspaces with different prediction values.

(b) Binary tree partitioning the input space according to (a).

Figure 2.9: Partition (a) and tree (b) of a classification problem.

Desicion Trees

A simple and widely used type of a decision tree is the Classification and regression tree (CART). Figure 2.9 shows a regression problem with inputs x_1 and x_2 which illustrates the concept of CARTs. The input space is recursively split into subregions using a simple decision like $x_i < t$ until a stopping rule applies. The predicted value for new data is, in case of regression, the average of all training data points in the subregion of the new data point and the class of the majority in case of classification. Every split is calculated by a splitting criterion, e.g. for classification, the variable x_i and threshold t are selected in a way, that best separates the data points by their class. The partition in figure 2.9(a) was created by the decisions seen in the tree in figure 2.9(b). First the space in (a) is split with $x_1 > \frac{1}{4}$, separating one cloud of class C_1 completely. The next best split parts the right subspace further by separating the three clouds in the top C_2, C_3 and C_4 from the the bottom clouds of C_1 and C_3 through the decision rule $x_2 > \frac{1}{2}$. This continues until either all classes are separated or a stopping criterion is met. Since the individual divided subregions are axially parallel cubes, the complexity of the entire data set can usually not be captured by a single partition of a single tree.

Random Forests

Random forests use bagging to reduce the variance of the individual trees, i.e. each tree is grown only on a randomly drawn bootstrap sample of the whole training data. For the calculation of the best split in every step of each tree, again only a random subset of all variables are considered. This method, called random subspace method [Ho95; Ho98] results in different, much less correlated trees, since the splitting decisions are highly dependent on the available data and variables. The averaging over a forest of uncorrelated, independent and identically distributed trees, each with high variance, largely reduces the variance of the forest and leads to better generalization on unseen data. The output of a random forest is, in case of regression, the average of all trees. For classification, the majority of the trees decide the class. Considering a binary classification problem, characterized by having only two exclusive classes, say class C_0 and C_1, the average of the predictions of all trees also provides valuable insights. An average of 0.55 for example results in the classification of class C_1, even though almost halve of the trees predicted class C_0. Therefore outputs around 0.5 indicate uncertain decisions, while outputs close to 0 or 1 imply unity among the decision trees in the forest.

Random forests can be tweaked through a number of parameters. Most implementations feature the following parameters: The number of trees n_{trees} used, more trees mean longer computation time but better models with less variance; the size of the bootstrap sample $B \subseteq X$ as a fraction of the whole data set X, used to grow each tree, $0 < r \leq 1$, larger bootstrap samples mean more robust, but also more similar trees, a drawback, since the potency of the forest stems from the trees being uncorrelated; the number of features considered in each split, again, higher numbers increase the correlation between the trees, but increases the chance, that features particularly important for a given split are taken into account.

2.3.2 Out-of-bag estimates

Out-of-bag error

An important benefit of random forests are the out-of-bag estimates for error, correlation and variable importance. The generalization error is the error rate of the classifier on unseen data. The out-of-bag estimate for the generalization error is calculated using only the training data and is defined by Breiman [Bre01] in the following way. Let T be a training set and T_k the bootstrap samples on which the decision trees $h(\mathbf{x}, T_k)$ are grown. For every $(y, \mathbf{x}) \in T$ the out-of-bag classifier is simply the collection of trees $\{h(\mathbf{x}, T_i) : (y, \mathbf{x}) \notin T_i\}$ for which $(y, \mathbf{x})$ is not part of the bootstrap sample used to grow the tree. The out-of-bag classifier therefore operates on unseen data. The error rate or miss-classification rate of the out-of-bag classifier applied on the training set is finally the out-of-bag error. The random forest hence provides an error estimation without the need to split the data into a separate training and test sets.

Variable importance

Having trained a random forest, the interpretation of how the response came to be and which input or feature was significant is often important. It is key for the understanding of the underlying problem to know which input variables were significant for the classification result, and which contributed almost nothing. This information is also highly relevant for dimension reduction. Such insights can be gained by the calculation the importance of each variable. Breiman [Bre01] described the variable importance as follows. To obtain the variable importance of a single feature f_i, first the values of this variable are randomly permuted across the entire data set. The out-of-bag error or miss-classification rate with permuted ith variable is now compared to the out-of-bag error with intact variables. The percent increase in miss-classifications for permuted variable f_i is its variable importance. Features that increase the miss-classification rate if permuted, are therefore important for the classification result and attributed with higher variable importance.

According to Strobl et al. [Str+07; Str+08] variables that are correlated to descriptive variables, but have no causal relation to the problem have an importance measure > 0. This drawback means that descriptive and redundant variables may be selected alike if only the permutation importance is used.

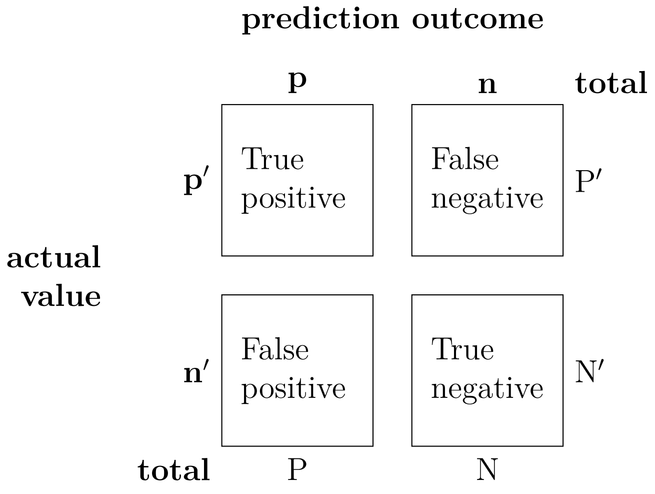
2.3.3 Model validation and Performance Characteristics

Once the training has finished and a classifier is obtained, its predictive performance is of interest. Therefore the resulting model is validated using test data, of which, like the training data, the real classes or labels are known. The model output of the test data is calculated and compared with their true label. Validation ensures the predictive quality of the model and its applicability to the problem at hand.

In order to validate a model on unseen data the available labeled data has to be divided into two disjunct data sets before training, the training set used to train the model and the test set for the calculation of performance measures needed to validate the model. One popular technique is k-fold cross-validation, where the whole sample is randomly split into k sub samples. Each of the sub samples is used as validation data on a classifier trained on the union of all other sub samples. The results can be visualized, for example, in k confusion matrices or one confusion matrix containing the averaged results of all k models.

Confusion Matrix

The confusion matrix briefly describes the amount of correctly and miss-classified data. Figure 2.10 shows a confusion matrix for a binary problem with classes 'positive' and 'negative' abbreviated with p and n. The rows differentiate the actual value of the data points, while the columns reflect the outcome of the model. True positive (TP) is the number of correctly classified positives and true negative (TN) is the number of correctly classified negatives. False positive (FP) and false negative (FN) are samples wrongly predicted as positive and negative, respectively. The models accuracy can be calculated by $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ and denotes the fraction of correct predictions.



		prediction outcome		
		p	**n**	**total**
actual value	**p′**	True positive	False negative	P′
	n′	False positive	True negative	N′
	total	P	N	

Figure 2.10: Confusion matrix for a two class (binary) problem.

2.4 Microplastics

Microplastic particles or microplastics have been detected in all environmental compartments, including marine and coastal environment, freshwater environment, wastewater, soils, air and in biota [SAP19]. They are defined as small polymer particles with a diameter below 5 mm [ABB09; GES15] and usually above 1 µm. Smaller particles are referred to as nanoparticles.

2.4.1 Origin

Microplastics entering the environment are either already manufactured in the microplastic size range or result from weathering, breakdown and fragmentation of larger plastic debris. The former group is often called primary while the latter are secondary microplastics. Primary microplastics are for example used as micro beads in cosmetics or to blast clean surfaces, while secondary microplastics are formed during the lifetime of products, for example through wearing of clothes or friction of tires, or after their life cycle, e.g. through UV induced breakdown of packaging on beaches [And11; Bro15]

2.4.2 Effects of microplastics

In this section a few documented adverse interactions of microplastic with the marine environment are presented. Microplastic can interact with the environment in many different ways. Ingestion of plastic and microplastic particles can have direct effects, like blockages of intestines while others are more subliminal, like the transport of toxins through plastic particles or the change of permeability of the sediment, which influences, for example, sex ratio of sea turtle eggs [Lus15]. Ingestion of particles by species of various trophic levels, including sea birds, marine mammals, fish, turtles and plankton was observed by numerous authors [GES15; GES16]. Plastic particles can cause a blockage in the intestines of animals and thus result in rapid death [KBV15]. Particles also take up space in the stomach of animals, which limits the food intake and leads to deterioration of the body condition of those animals [KBV15]. Smaller particles can have effects on a cellular level, for example high density polyethylene (HDPE) particles with sizes below 80 µm ingested by blue mussels (*Mytilus edulis L.*) in a study performed by Von Moos, Burkhardt-Holm and Köhler [VBK12], led to accumulation in the lysosomal system and strong inflammatory responses.

Sea turtle egg's sex ratio is dependent on the sand temperature. Below 28 °C all embryos are male and at 30 °C there is an equal amount of males and females. Microplastic in sediment increases the sediment permeability and thus increases the time needed for sand to heat up, which in turn may introduce sexual bias in sea turtles [Lus15].

Besides the physical effects microplastics exert on the environment and its life, microplastics also exhibit deleterious chemical effects. During production of plastics and microplastics several substances are added in order to give the plastic certain characteristics, e.g. color, flexibility and strength. Those additives include pigments, flame retardants, stabilizer and fillers. According to Lithner, Larsson and Dave [LLD11] more than half of the plastics produced are hazardous, based on their monomers, additives or by-products. Next to the chemical contaminants already present from production, microplastics also accumulate pollutants present in the surrounding seawater [Roc15].

It is therefore key to improve the detectability and classification of microplastics.

Chapter 3

Methods

3.1 Spectral Descriptor Selection

One of the following experiments of this thesis is about the evaluation of classifiers, that are trained to discriminate different polymers from one another and from background and matrix material by their FTIR spectra (see section 3.2 and 3.2). This section describes how a random forest classifier can be obtained without any prior chemical knowledge like the position of absorption bands (see table 2.1). This classifier will later on be compared with a random forest obtained, when experts handpick spectral descriptors (SPDCs) specific for the 20 polymers contained in the samples. The features used for the first approach are selected solely through their variable importance calculated on a labeled training set. The resulting feature set, consisting of highly descriptive SPDCs can be used to gain further insight into the problem, highlighting spectral regions of interest. On the other hand, the classifier could be overfitted to the training data and might perform worse on unseen data, compared to the random forest constructed from experts.

In order to train a classifier, a random forest in this case, training data and a set of features, here the spectral descriptors are needed. The training data are the FTIR spectra of the samples described in section 4 including their labeled class affiliations. The feature set is obtained by the generation of a large, unspecific set of features, which is subsequently narrowed down to a few descriptive SPDCs. Since the spectra used in this work comprise 609 layers, there are around 609 possible options for the location of SPDCs. Most SPDCs have in addition to the center or location of the descriptor additional parameters. In the case of the often used TC descriptor, which represents the weighted correlation of the spectrum with a triangle template, two additional parameters are used. The left extent relating to the left rising flank (b1) and the right falling flank (b2) complement the location of the center (a1) (see figure 2.6(c)). ImageLab offers about 20 different SPDCs, each with 1 to 4 parameters. Creating a feature set of all sensible types of SPDCs and reasonable variations of their parameters therefore results in a huge number of features.

This section covers an algorithm for the feature generation and subsequent selection of the most important features using the out-of-bag variable importance property of random forests (see section 2.3.2). First a list of candidate spectral descriptors is generated and through iterative training of random forests the best features in terms of their variable importance, are selected. Algorithm 3.1 depicts the simplified pseudo code used to find the SPDCs with highest variable importance.

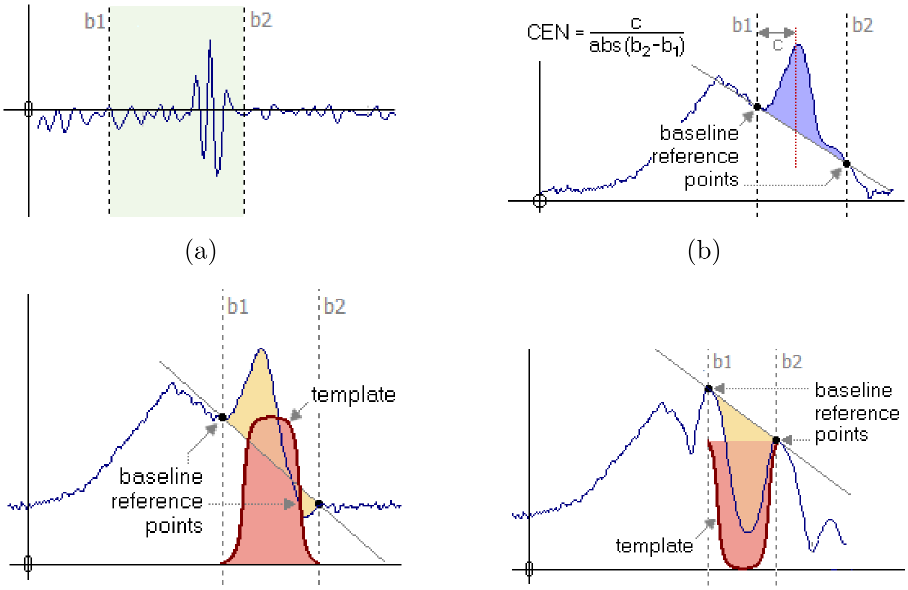
```
Data: Trainingsdataset T
Result: List of features F with highest variable importance
features $F_G$ = generateFeatures()
for $i \in \{1 \dots m\}$ with $t_1, \dots t_m$ types of SPDCs do
    $F_t = \{f \in F_G : type(f) = t_i\}$
    $F_i$ = featureSelection($F_t$)        // select features for each SPDC type
end
$F = \bigcup_{i \leq n} F_i$                // remove duplicates
$F_S$ = featureSelection(F)
Function featureSelection(F)
    for $i \in \{1 \dots n\}$ do
        draw random sample $S_i \subset F$ with sample size $|S_i| = d$
        var_imp = RF($S_i$, T)
        $F_i = S_i(\text{var\_imp} \geq c)$        // keep features of $S_i$ with higher var_imp
    end
    return $F_S = \bigcup_{i \leq n} F_i$        // remove duplicates
```

Algorithm 3.1: Feature Selection

3.1.1 Algorithm

Feature Generation

First, an exhaustive list of features is generated across the 609 layers. For every SPDC type and layer multiple descriptors are created with different parameters, sampling the parameter space in a reasonable area. Parameters are for example the center of a SPDC and its lower and upper boundary. In addition to the SPDC types described in section 2.2.1 and figure 2.6 the VAR, CEN, GC and GCI descriptor types were used as well (see figure 3.1). VAR describes the variance in the spectrum between two points b_1 and b_2. CEN is the position of the centroid in a selected region $[b_1, b2]$. GC and GCI are similar to the TC and TCI spectral descriptors. Here the correlation is calculated between the spectrum and a Gaussian.



(a) (b) (c) (d)

Figure 3.1: 4 further types of spectral descriptors [Epi20]. (a) VAR – variance of the spectrum between b_1 and b_2 (b) CEN – location of the centroid relative to b_1 and b_2 (c) GC – weighted correlation with a Gaussian curve with amplitude 1 (d) GCI – same as GC with negative amplitude

The TC and TCI descriptors have three parameters and the DV1 features two parameters. VAR, CEN, ABL, GC and GCI all have the same parameters and are therefore treated alike during the feature generation. They possess, like the DV1 two parameters, but with different meaning. For each layer $l \in \{1 \dots 609\}$ n descriptors of the last mentioned group of types are created with $b_{1,i} = l$, $b_{2,i} = l + i \cdot s$ with $0 < i \leq n$ and s the sample rate. The DV1 is calculated using a certain window of the spectrum of size b around the descriptor location a_1. The DV1 descriptors are created simply by sampling all possible positions $a_1 \in \{1 \dots 609\}$ with fixed window size. Both TC and TCI take three parameters. For each layer of the spectrum l descriptors of these types with parameters $a_{1,i} = l$ and all possible combinations of $b_{1,i} = l - i \cdot \frac{s}{2}$ and $b_{2,i} = l + i \cdot \frac{s}{2}$, where s is the step size, are created and appended to the list of generated features. If three different values for b_1 and b_2 each are used, resulting in 9 combinations, the descriptors of type TC alone account for approximately 5300 SPDCs, a few less than $5481 = 9 \cdot 609$

since the options for b_1 and b_2 are restricted for points close to the boundary of the spectrum.

Feature Selection

As described in algorithm 3.1, the list of features is first split and narrowed down separately for each SPDC type. The union of these intermediate results is afterwards again trimmed through the featureSelection function, now yielding the final, selected set of SPDCs. The features of each type are first filtered to reduce the size of the intermediate feature set and therefore greatly reduce calculation time.

Inside the function featureSelection, first a simple random sample without replacement of size d is drawn. A random forest is trained using the features of the drawn random sample and the training set discussed earlier. The out-of-bag calculated property variable importance of the random forest is then used to eliminate the least important features from the sample. Since correlated features can skew the variable importance, the features with highest variable importance are not necessarily the most descriptive. Features with very low variable importance on the other hand were rarely selected in a split during training, indicating low relevance. Therefore the features are eliminated from the sample over multiple iterations, removing in each iteration only the least important SPDCs. After all N random samples $\mathsf{S}_1, \ldots \mathsf{S}_N$ are processed, duplicate features are removed.

According to algorithm 3.1 the best features of each SPDC type are selected through the featureSelection function. The best features of all types are then combined and again put through featureSelection. The finally resulting set of spectral descriptors $\mathsf{F} = \bigcup_{n \leq N} \mathsf{S}_n$ is utilized for the training of the final random forest classifier.

3.2 Mixture Evaluation

The response of a classifier to a single mixture image is highly dependent on the selected spectra, the mixture was created from and can potentially vary greatly even if different spectra within the same class are used. Figure 3.2 shows four classified mixture images of polyethylene (PE) and polypropylene (PP). The figure illustrates the discrepancies in the classifiers response caused by different spectra. Darker areas represent a lower response from the classifier trained to detect PE, while brighter, more yellow areas indicate high certainty for the underlying spectra to be PE. Starting in the bottom right corner, the response gradually decreases along the x-axis in reverse direction as the fractions of PE and PP decrease and increase respectively. The two images on the right show mixtures with dominating and weak PE spectra. The top right plot displays high classifier responses even in the upper rows, where higher amounts of noise were added, while the bottom right features high responses only closely to $(101, 0)$. The classified mixture image in the top left represents an average result with high responses down to 60% PE and 40% PP.

In order to evaluate the classifier characteristics in dependence of the polymer types only and remove such variations caused by inner class variance of the spectra, multiple mixture images are created and features calculated from the classified images used for the interpretation.

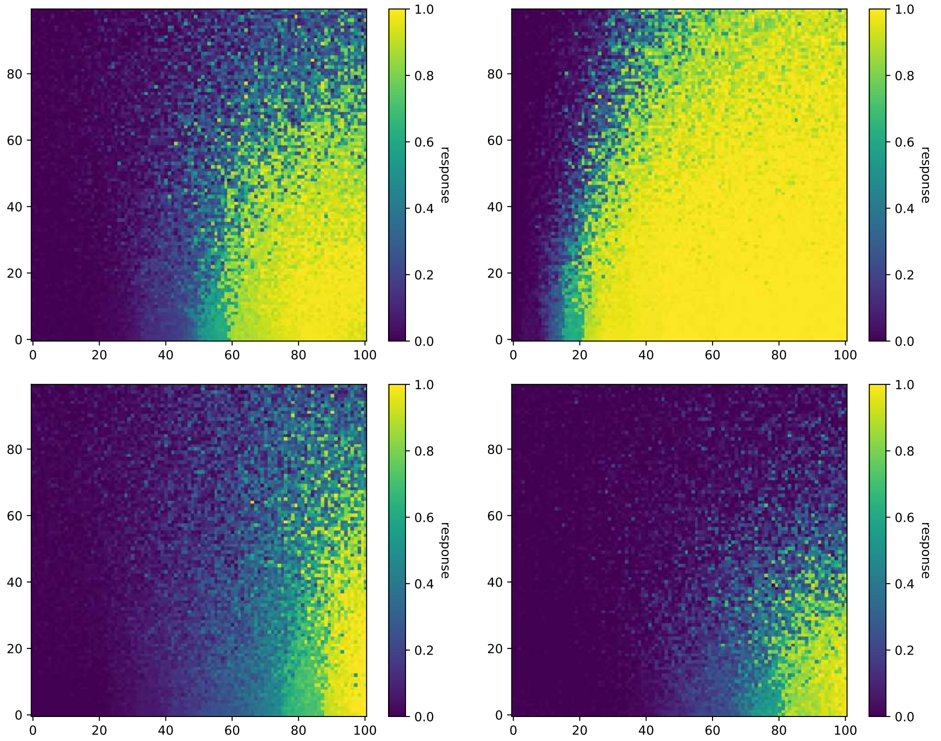


Figure 3.2: Shows four classified mixture images created from different PE and PP spectra. The original PE and PP spectra are located in the lower right and left corner respectively. The figures show the response of a random forest trained to detect PE.

3.2.1 Mixture Image Features

To investigate the behaviour of a specific mixture, a couple of mixture images are created and classified. If, for example, 10 spectra of both classes are used, 100 mixture images are obtained if every spectrum of class C_1 is mixed with every spectrum of class C_2. Because it is unreasonable to interpret all these images manually, features describing the classified images are created.

3.2.2 Areal Features

Three simple features used to describe the classified images as a whole are the number of pixels with classification results $\hat{y} \geq \frac{2}{3}$, the number of pixel with $\hat{y} < \frac{1}{3}$ and those in-between with values $\frac{1}{3} \leq \hat{y} < \frac{2}{3}$. Equation 3.1 shows these features, where $s_{mix}(x_1, x_2)$ is the spectrum with coordinates (x_1, x_2) in the mixture image, x_1 being the direction along which the fraction of the mixture changes and x_2 the direction of increasing amounts of noise. The indicator function is defined as $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise.

$$\begin{aligned} A_p &:= \sum_{x_1,x_2} I_{[\frac{2}{3},1]}(\hat{y}(s_{mix}(x_1,x_2)) \\ A_b &:= \sum_{x_1,x_2} I_{[\frac{1}{3},\frac{2}{3})}(\hat{y}(s_{mix}(x_1,x_2)) \\ A_n &:= \sum_{x_1,x_2} I_{[0,\frac{1}{3})}(\hat{y}(s_{mix}(x_1,x_2)) \end{aligned} \tag{3.1}$$

High values of A_b denoting a large number of classified spectra with response $\hat{y} \approx 0.5$ and are attributed to weak performance with regards to noise, since most of the mixture images pixels contain added noise. A rather sharp border with A_b close to 0 would be ideal. The following function of A_p, A_n and A_b describes the ratio of pixels the classifier was confident about (high and low values) and pixels with values around $\frac{1}{2}$, meaning unsure decisions.

$$f(A_p, A_n, A_b) = 4\frac{A_p A_n}{1 + A_b^2} \tag{3.2}$$

This function is showing high values if the mixture is well separated (low A_b) and favors equal values of A_p and A_n. The factor of 4 scales the function to 1 if the classified mixture image contains only pixels with response values $\hat{y}$ above $\frac{2}{3}$ and below $\frac{1}{3}$ in equal amounts, hence $A_p = A_n = 0.5 \implies A_b = 0$.

3.2.3 Logistic Function Features

If we consider only a single row of the classification results of a mixture image and plot the classifier response values of this row against the mixture fraction (x_1), we end up with a function similar to a logistic function. Figure 3.3 shows exactly this discrete function for the bottom row (no added noise) of the classified mixture images seen in 3.2.

The smooth function approximating the response of the classifier is a logistic function

$$f(x) = \frac{a}{1 + e^{-k(x-x_0)}} + y_0 \tag{3.3}$$

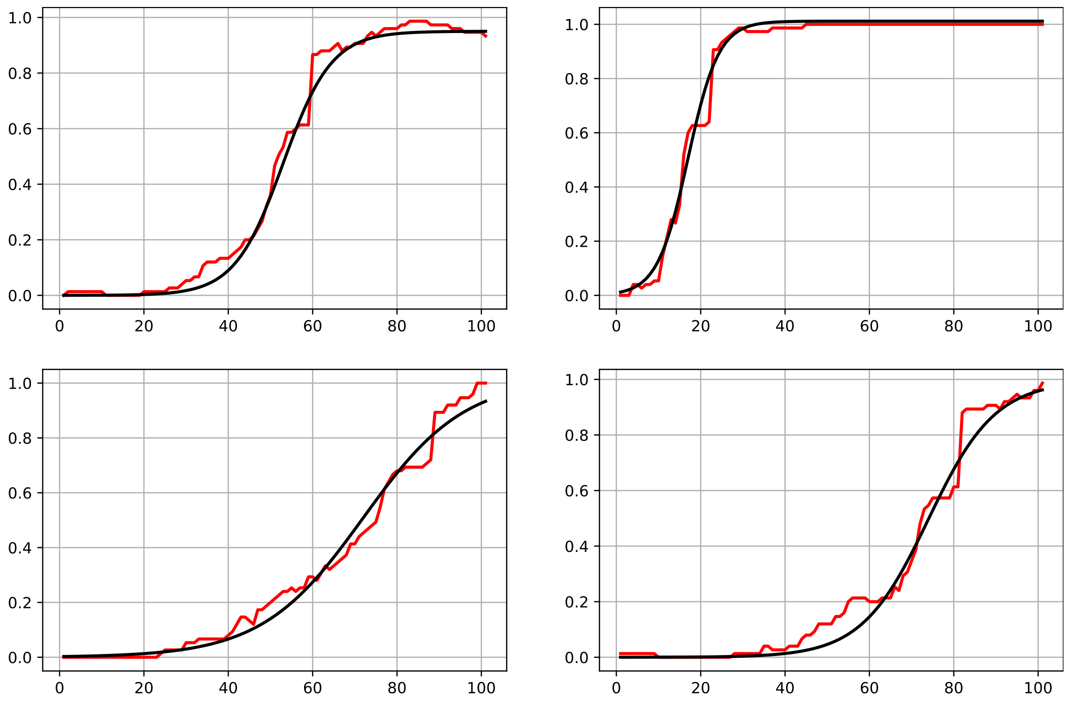


Figure 3.3: Plot of the bottom line values of the classified mixtures seen in figure 3.2 connected through a red line and logistic function with best fit drawn in black.

where a is its amplitude, y_0 is the offset from the x-axis, k the steepness of the curve and x_0 the midpoint or inflection point of the sigmoid function.

The parameters of the logistic function, that best approximates the classification results of a given row, provide descriptive information about that row. How this best approximation is determined is discussed in the next section. Specifically the midpoint x_0, offset y_0, steepness k, amplitude a and the range of the logistic function $r(f) := f(100) - f(0)$ are used as features.

High steepness k and $r(f) \approx 1$ indicate, that the classifier is able to separate both components well. $r(f) = 1$ is the ideal case while $r(f) << 1$ suggest classification results for the positive spectrum below 1 and or response for the negative class above 0. The midpoint can be interpreted as the classifiers sensitivity to the positive class. In figure 3.3 the position of x_0 shows the fraction of the positive class needed, in order for the classifier to yield a positive response. Low x_0 as seen in the bottom right plot could be caused by dominating descriptive features in the spectrum of the positive class and or the presence of similarities between the two classes in the spectrum of the other class.

Logistic Function Fitting

The logistic function with best fit is determined with linear regression, since maximum likelihood estimation is not available in *ImageLab* as of version 3.15. Linear regression calculates the fit, amplitude β_1 and intercept β_0 for a linear model

$$Y = \beta_1 X + \beta_0 \tag{3.4}$$

with given observations $(x_1, y_1), \ldots (x_n, y_n)$ of X and Y. We can use a transformation to calculate the amplitude a and the offset y_0 of the logistic function that best describes the classification results Y for fixed midpoint x_0 and steepness k.

$$X_{x_0,k}(x) := \frac{1}{1 + e^{-k(x-x_0)}}, \tag{3.5}$$

If we combine the transformation 3.5 with the linear model used for regression 3.4 we get

$$Y(x) = \beta_1 \frac{1}{1 + e^{-k(x-x_0)}} + \beta_0. \tag{3.6}$$

We are regressing the row values of a classified mixture image $Y(x)$ onto a real logistic function $X_{x_0,k}(x)$ and get $\beta_1 = a$ for the amplitude and $\beta_0 = y_0$ for the offset as a result. β_0 and β_1 a are chosen in a way, that minimizes the squared errors between $Y(x)$ and $X_{x_0,k}(x)$. Table 3.1 shows the values of $Y(x)$ and $X_{x_0,k}(x)$ for columns 31 to 40 of the data in the bottom left plot of 3.3. To obtain all the parameters of the logistic function, that approximates the response function best, the regression is calculated for every combination of x_0 and k. x_0 is sampled in $\{0, \ldots 100\}$ and for k 100 values logarithmically spaced between 0.1 and 1 are tried. The combination of x_0 and k that results in the regression with the highest fit between Y and X determine the values of those parameters.

x	31	32	33	34	35	36	37	38	39	40
$X_{x_0,k}(x)$	0.16	0.19	0.22	0.26	0.30	0.35	0.40	0.45	0.50	0.55
$Y(x) = \hat{y}(x)$	0.09	0.09	0.12	0.32	0.32	0.32	0.32	0.32	0.73	0.73
$f(x)$	0.13	0.16	0.20	0.23	0.28	0.32	0.37	0.42	0.47	0.52

Table 3.1: Observations 31 to 40 of the input for the linear regression of classifier response $\hat{y}(x, 0)$ onto $X_{x_0,k}(x)$ with midpoint $x_0 = 62$ and steepness $k = 0.2089$. The data correspond to the bottom left plot in figure 3.3.

Further, only logistic functions f with $-\epsilon < f(x) < 1 + \epsilon$ for $x \in [0, 100]$ and $\epsilon = 0.025$ will be considered for the best fit. Functions with values far below

0 or above 1 do not represent meaningful classifier responses and are therefore discarded. Algorithm 3.2 shows the pseudo code used to find the logistic function with the best fit.

```
Data: $Y = \hat{y}$, row of classified mixture image
$x = (1, 2, \dots 100)$
$\text{fit}_{\text{max}} = 0$
foreach $x_{0,i}$ and $k_i$ do
    $X_{x_{0,i},k_i} = (1 + e^{-k_i(x - x_{0,i})})^{-1}$          // calc. logistic function values
    $(\text{fit}, \beta_0, \beta_1) = \texttt{LinReg}(X, Y)$
    if $\text{fit} > \text{fit}_{\text{max}}$ and $-\epsilon < f(x) < 1 + \epsilon$ then
        $\text{fit}_{\text{max}} = \text{fit}$;  $a = \beta_1$;  $y_0 = \beta_0$;  $x_0 = x_{0,i}$;  $k = k_i$
    end
end
Result: logistic function with parameters $a, y_0, x_0, k$ approximating $\hat{y}$
```

Algorithm 3.2: Pseudocode to find the logistic function with best fit

3.2.4 Mixture Evaluation Tool

The main focus of this chapter was the development of a tool in *ImageLab's* scripting language, that accepts a classifier and a test dataset as input and supplies information about performance of the classifier with regards to mixtures and noisy data in the form of features as output. Additionally a multiplier to control the amount of noise added to the spectra is queried from the user. Depending on the data the classifier was trained with, standardization or other scaling mechanisms can be applied to the test data as preparation for the mixture evaluation. The features extracted from the generated and classified mixture images are exported and stored in a CSV file and all mixture images classified in the process are saved as matrices with the .asc extension. The visualization of the results is realized in a later step with python.

Chapter 4

Experimental

Software. The experiments, data analysis and visualization was carried out with *ImageLab* 3.15, a software application for analyzing hyperspectral images with an built-in Pascal-like scripting language. Furthermore *Python* 3.8.1, using the modules *Pandas* 0.25.3, *Matplotlib* 3.1.2 and *Numpy* 1.18.1 were used.

Data. The samples used in the following experiments were artificially created and contain microplastic particles on an aluminium oxide filter. The particles are arranged in a circular region measuring about 10 mm in diameter. The particle sizes vary between 10 µm and 100 µm. Each of the 8 samples addressed in table 4.1 holds particles/fibers of 5 different polymers, which were mixed with clear water (samples are named R1 to R4) and with a freshwater plankton sample (samples R1.1 to R4.1) before eventually being sucked through aluminium oxide filters. In addition to the artificial data, the environmental samples used by Primpke et al. [Pri+18] were also used. These two samples will be referred to as RefEnv1 and RefEnv2.

R1/R1.1	PU	PET	PA	PC	PMMA
R2/R2.1	PVC	PE	PP	POM	ABS
R3/R3.1	PAN	PS	PBT	PPSU	CA
R4/R4.1	EVAc (14% Ac)	PEEK	EVOH	PSU	Silicone

Table 4.1: Sample composition – The table shows the different combinations of polymer particles found in the used samples along with the label they will be referenced with.

Out of every freshwater plankton sample R1.1 to R4.1 40 spectra of each polymer as well as 40 substrate spectra and 40 spectra containing material belonging

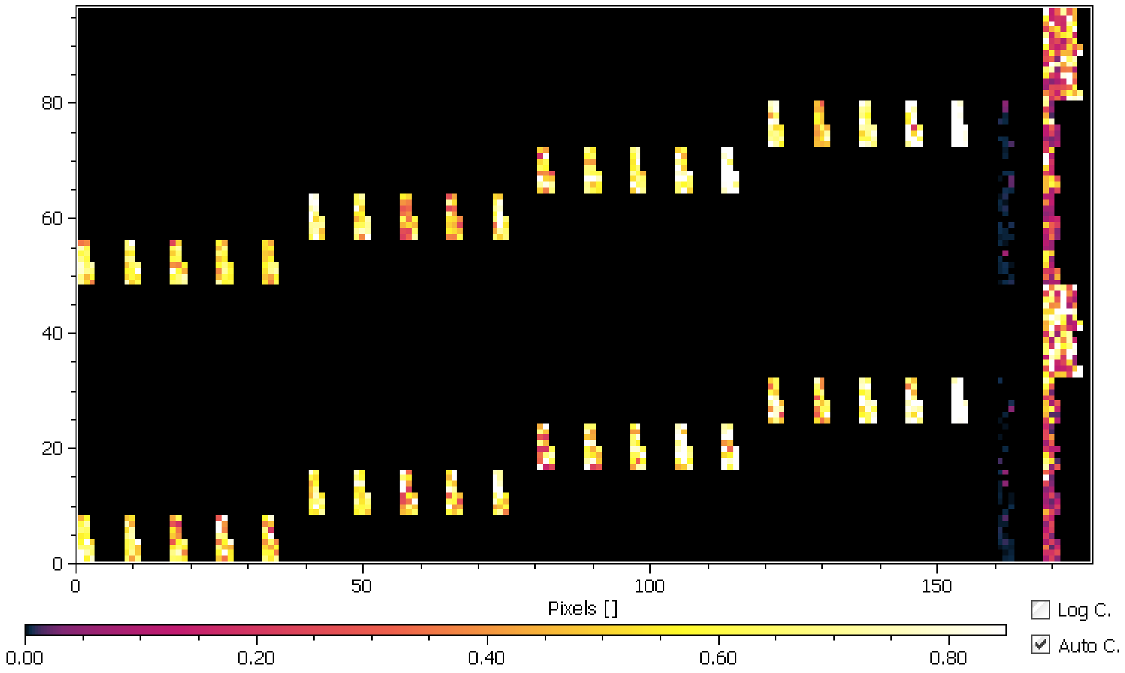


Figure 4.1: Class map of used data

to neither polymer have been labeled, resulting in 1120 labeled spectra. The latter group of spectra or class has been named non-polymer. In order to better capture the variety of this non-polymer class, an additional 100 spectra were labeled from each of the RefEnv samples.

The samples in table 4.1 were measured with a Bruker Hyperion 3000 FTIR microscope resulting in spectra containing 609 data points from approximately $3600\,\mathrm{cm}^{-}1$ to $1250\,\mathrm{cm}^{-}1$. The data of the environment samples RefEnv1 and RefEnv2 was recalibrated to match the spectral resolution of the former samples.

The class map shown in figure 4.1 contains the labeled training and test data. The sampling was carried out in two stages explaining the partition of the data in an upper and lower half. Each row corresponds to a different sample. In the barely visible second to last column are the low intensity substrate spectra and the last column holds the spectra attributed to the non-polymer class. Since only non-polymer spectra were sampled in the RefEnv samples, only the last column contains data in the first two rows. The first 20 columns contain the labeled polymer spectra. The image is colored according to the intensity at the arbitrarily selected $\bar{\nu} = 2241\,\mathrm{cm}^{-1}$.

Experts Feature Set. As input for the evaluation tool developed as part of this thesis (see section 3.2), a random forest classifier with manually selected spectral descriptors is used. This random forest or rather its feature set is also briefly compared to the classifier constructed in the following section 4.1 using the methods explained previously in section 3.1. These manually selected features are placed at characteristic regions of the spectrum, therefore introducing chemical knowledge into the selection of suitable features. This set contains template correlation (TC), Gaussian correlation (GC), first derivative (DV1), one raw area (ARW) and several IGF descriptors. The IGF descriptor calculates, much like the TC(I) and GC(I) descriptor, the correlation between a spectrum and a template. The template here is in contrast to the aforementioned SPDC types a section of another spectrum, usually selected to be a characteristic region for a class or polymer type. These not necessarily connected sections of a spectrum are commonly taken from reference spectra.

The feature set comprises 150 SPDCs in total with the majority being IGF and DV1 descriptors. The single ARW descriptor separates the usually low intensity substrate spectra.

4.1 Spectral Descriptor Selection

As described in 3.1 an exhaustive list of potential spectral descriptors is generated and subsequently narrowed down to contain only a few specific features.

The SPDCs having two parameters delimiting the spectral region they are defined on (ABL, GC, GCI, VAR, CEN), are generated with five different widths, ranging from 12 to a maximum of 60 layers or from approximately 42 to 212 wavenumbers. The DV1 descriptors are specified to calculate the first derivative in windows sizes of 10 layers around the selected location. Three parameter SPDCs (TC and TCI) are created with widths of 20, 40 and 60 layers. For this purpose descriptors with all combinations of $b_1 = a_1 - \frac{s}{2}$ and $b_2 = a_1 + \frac{s}{2}$ with $s \in \{20, 40, 60\}$ are created for all a_1. For example 9 TC and TCI descriptors with parameters $\{(a_1, b_1, b_2) : a_1 = 100, b_1 \in \{70, 80, 90\}, b_2 \in \{110, 120, 130\}\}$ for the location $a_1 = 100$ are generated. In total the list of features generated this way sums up to 25732 spectral descriptors.

For the selection of features 30 independent randomly drawn samples with a sample size of a fifth of the input feature set size are drawn. Each such sample is iteratively halved and hence trimmed down to only contain the ten best features. The resulting distinct $n_t \leq 300 = 30 \cdot 10$ best features of each SPDC type obtained like this are combined and finally narrowed down again using the same strategy to acquire $n \leq 300$ unique descriptors. The parameters of the random forest classifiers trained in each iteration to obtain the variable importance were chosen to be $n_{trees} = 75$ trees and the size of the bootstrap sample of the training data used in each split as a fraction of the size of the whole data set $r = 0.3$.

These parameters were selected as a trade-off between computation time and performance. More trees mean a more stable response of the classifier but take more computation time to calculate. With increasing bootstrap sample size fraction, again, the computation time rises as more sample spectra have to be separated. Higher r yield more robust but also more similar trees. Therefore a low value with $r = 0.3$ was selected to ensure more variance among the trees.

In order to compare the acquired feature set with the one created from experts, two random forests are created with the same parameters and training set but different features. These classifiers use $n_{trees} = 300$ and $r = 0.5$ and are tested on the left-out test set containing another 20 spectra per polymer. Since these random forests are created only once, the higher calculation time needed to compute 300 trees is negligible. The results of the classified test set is turned into confusion matrices to gauge and discuss their implications.

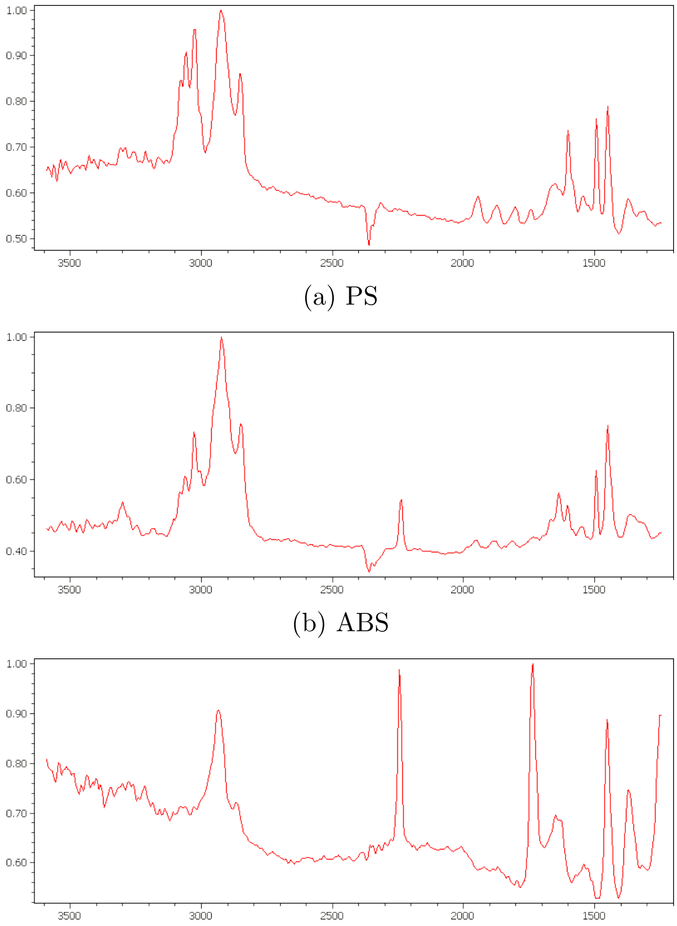
4.2 Evaluation of Mixture Images

The examination of the classifiers mentioned in the previous section 4.1 gives insight in their performance characteristics like miss-classification rate or accuracy. However from these experiments alone, one cannot deduce the classifiers reaction on noisy data and spectra containing mixtures of two classes.

Three classifiers are used for the various experiments explained hereafter. Mostly employed is the experts classifier, this time not retrained with a limited training set of 20 spectra per polymer class, but in its original configuration with a significantly larger training set. The other two random forests evaluated during the following experiments are the already described classifiers obtained with the feature selected descriptors and the experts feature set trained on the training set. Since the size of the training sets of the first and the latter two classifiers differs greatly, divergent results can be expected. The test data set, all consecutive experiments are based on, is as before the complement of the training set (see beginning of chapter 4). In the course of the subsequent experiments a great number of pairwise linear mixture images of selected groups of polymers are created and investigated, therefore only summarizing statistics and visualizations will be shown.

4.2.1 ABS, PS, PAN

The combination of acrylonitrile butadiene styrene (ABS), polystyrene (PS) and polyacrylonitrile (PAN) is interesting, since they have similar absorption patterns. Both PAN and PS share, due to the C≡N bond and the aromaticity respectively, distinct absorption bands with ABS. Comparing the group assignments of PS and



(a) PS

(b) ABS

(c) PAN

Figure 4.2: Three spectra of PS, ABS and PAN taken from the training/test sets. Absorbance is plotted vs. wavenumbers (cm^{-1}).

PAN listed in table 2.1 with the spectra shown in figure 4.2 reveals these absorption similarities.

The 4 consecutive peaks slightly below $2000\,\mathrm{cm}^{-1}$ are present in the PS and in the ABS spectrum. In the latter the amplitudes are lower since the fraction of styrene is also lower in ABS. The pronounced peaks at $1494\,\mathrm{cm}^{-1}$ and $1454\,\mathrm{cm}^{-1}$ are also present in both spectra. The key similarity between the ABS and PAN spectrum is the prominent (double-)peak at $\approx 2241\,\mathrm{cm}^{-1}$ caused by the stretch of the C≡N triple bond in the acrylonitrile group. The presence of these similar absorption bands in varying intensities promise interesting linear mixtures.

20 spectra of each ABS, PAN and PS will be translated into $400 = 20 \cdot 20$ pairwise mixtures per combination, resulting to a total of $1200 = 400 \cdot 3$ mixtures for all three combinations. Each mixture image is classified with both binary

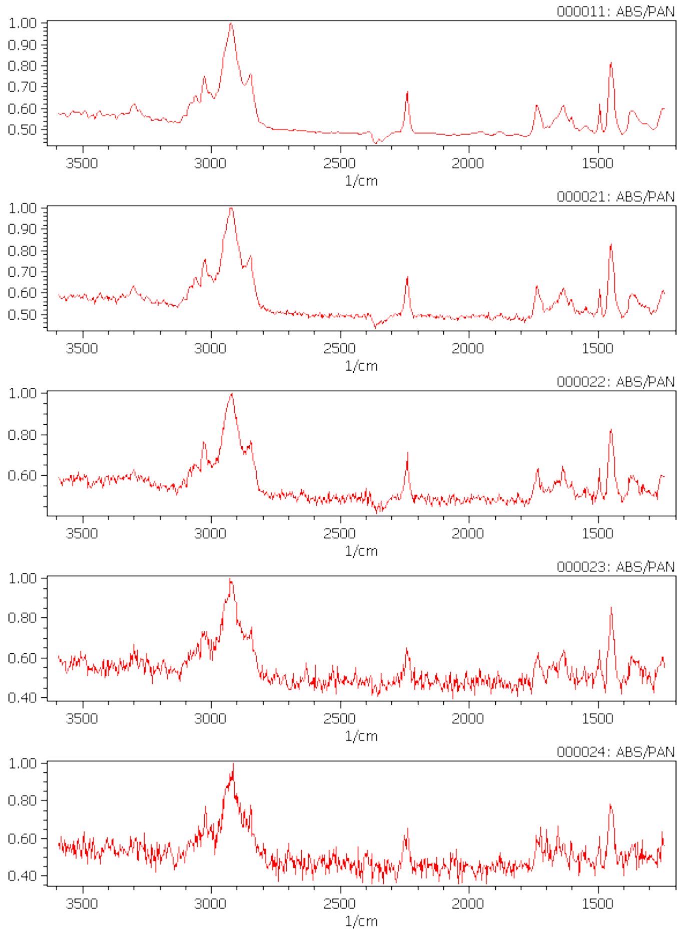


Figure 4.3: Mixture of ABS and PAN with fraction 0.5 each. Added noise is, from top to bottom 0, 0.1, 0.25, 0.5 and 0.75 of total noise = 1.

classifiers of the classes constituting the image. Or, in other words, the response matrices for those two classes only are calculated and processed further. Logistic functions are fit into the bottom row and rows 10, 25, 50, 75 containing zero, 10%, 25%, 50% and 75% of the maximum of the normally distributed noise. Mixtures of ABS and PAN containing the amounts of noise mentioned are displayed in figure 4.3, where the top spectrum contains no noise and the bottom one 75%. The multiplier controlling the total amount of noise is set to 1 for this experiment. While the features of the spectrum are clearly visible with 10% of added noise, only the major peaks can still be seen in the bottom spectrum.

In the following $2400 = 1200 \cdot 2$ classified mixture images and $12.000 = 2400 \cdot 5$ logistic functions are calculated. The classifications of this experiment were carried out with the experts classifier.

4.2.2 Most Produced Polymers

The aim of this experiment is to evaluate the behaviour of a classifier if faced with mixed spectra of some of the most abundant or most produced polymers. This includes polyethylene (PE), polypropylene (PP), polyvinyl chloride (PVC), polyethylene terephthalate (PET) and polyurethane (PU), the top 5 nonfiber plastics produced from 1950 to 2015 [GJL17]. Here 10 samples per class are used for each combination of polymers, yielding for all ten combinations 10^3 mixture images and 2k classified mixture images. The noise settings are, with total noise being 1 and investigated rows being 0, 10, 25, 50 and 75, the same as before.

4.2.3 Polymer vs. Non-Polymer

During this experiment the ability of the classifiers to discern the polymers in a polymer and non-polymer mixture is evaluated under different noise conditions. For this approach mixture images for all polymer classes as class C_1 with non-polymer as C_2 are generated. Only five observations of each class are used to create the mixture image. Using all 20 spectra for all 20 classes in the test set, would result in $8000 = 20^3$ mixture images being created, which would be very computationally expensive. The total amount of noise is again set to 1 and rows 0, 30 and 75 are investigated, hence logistic functions fit into the responses of these rows. Another aim of this experiment is the comparison of the classifiers treated in this work. This includes next to the experts classifier mentioned above, both classifiers obtained from the experts feature set as well as the feature selected set trained on the training set.

Chapter 5

Results

5.1 Spectral Descriptor Selection

As described in the previous chapters, an exhaustive list of spectral descriptors (SPDCs) was created and narrowed down to contain only the $n \leq 300$ best features. The quality of the features was measured using the out-of-bag property of the random forest variable importance, which describes for a feature or input variable, how much the miss-classification rate increases, if the values of this variable are randomly permuted prior to classification.

5.1.1 The Selected Features

The last step of the feature selection algorithm described in section 3.1 resulted in 30 random samples of the union of all selected features of all SPDC types narrowed down to ten features each. From these 300 features holding many duplicates, 185 unique SPDCs remain as the final result of the algorithm. Table 5.1 holds the distribution of the 185 descriptors across the 8 SPDC types used. Descriptors calculated as correlation to a template (TC, TCI, GC, GCI) were primarily selected, whereas centroid (CEN), variance (VAR) and the first derivative (DV1) are only represented in low numbers. Area descriptors (ABL) are also under-represented with 15 out of 185. The distribution of spectral descriptors over the 609 layers is displayed in figure 5.1. The data is split in three histograms. The top one holds the positions of the five DV1 SPDCs. The other histograms display for each layer

TC	GC	TCI	ABL	GCI	CEN	VAR	DV1
56	51	32	15	14	7	5	5

Table 5.1: Number of features of each SPDC type occurring in the result set.

or wavenumber the number of SPDC operating on that wavenumber. The distributions are unsurprisingly similar and depict the areas of the spectrum, that were useful in the separation of the classes. All peaks previously discussed are located in high density areas of the histograms.

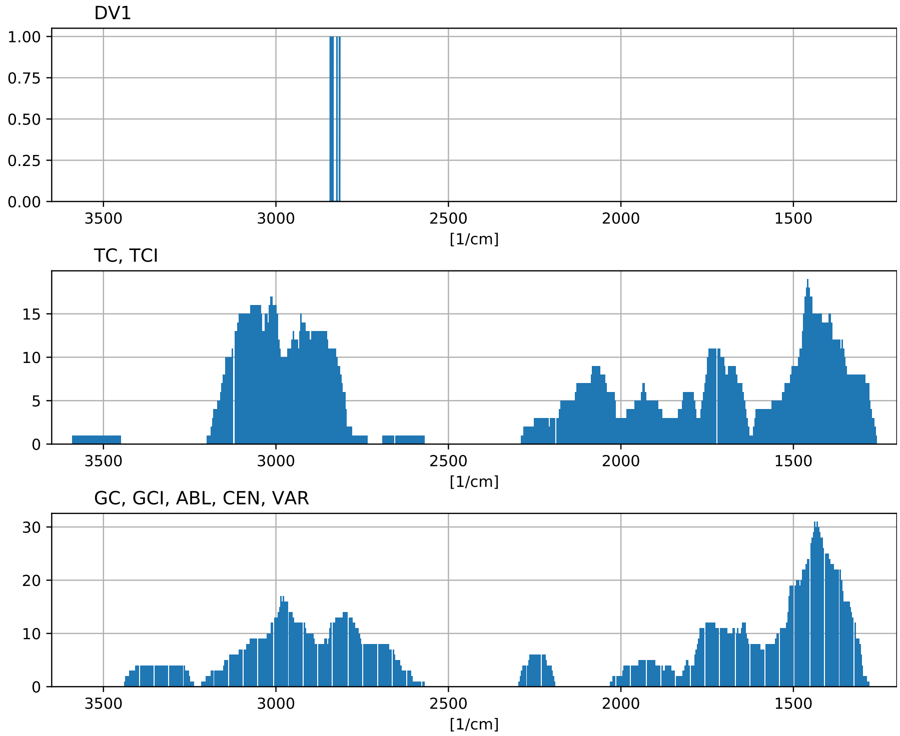


Figure 5.1: Distribution of the location of the best 185 features selected.

5.1.2 Validation

Both random forest classifiers were trained using the training set, comprising half the available labeled data. The other half of the labeled spectra is used here to validate and briefly compare the two classifiers. Figures 5.2 and 5.3 show the confusion matrices resulting from the test set being classified by the random forests. Both figures show 22 binary confusion matrices. Each describing the performance of the individual 22 binary classifiers. Everyone of those 22 binary classifiers calculates its response for every spectrum in the test set. If the response is above the threshold, depicted as solid red line at 0.5 in the matrices, the spectrum is

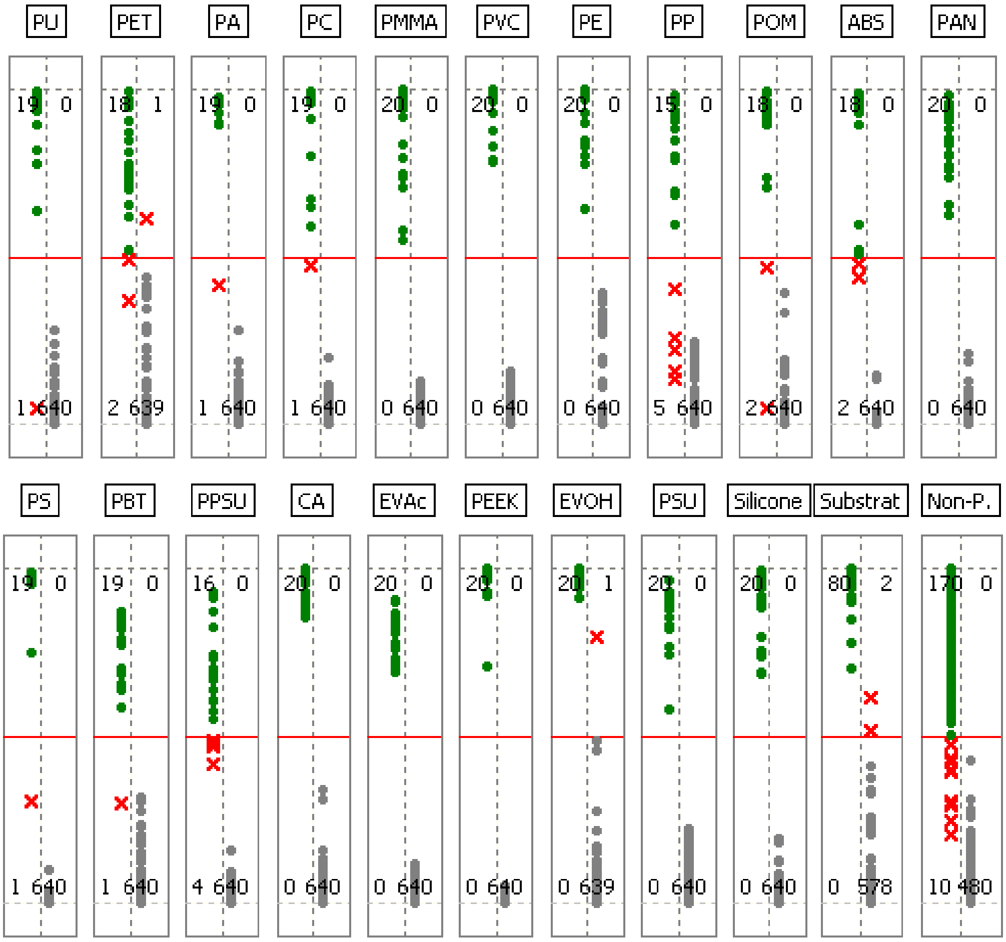


Figure 5.2: Validation of the random forest with the feature selected descriptor set. The confusion matrices of the 22 binary problems consist of the responses of the classifier plotted along the y-axis. The two gray dashed horizontal lines in each matrix are positioned at values of 1 and 0. The solid red line represents the classifier threshold at 0.5.

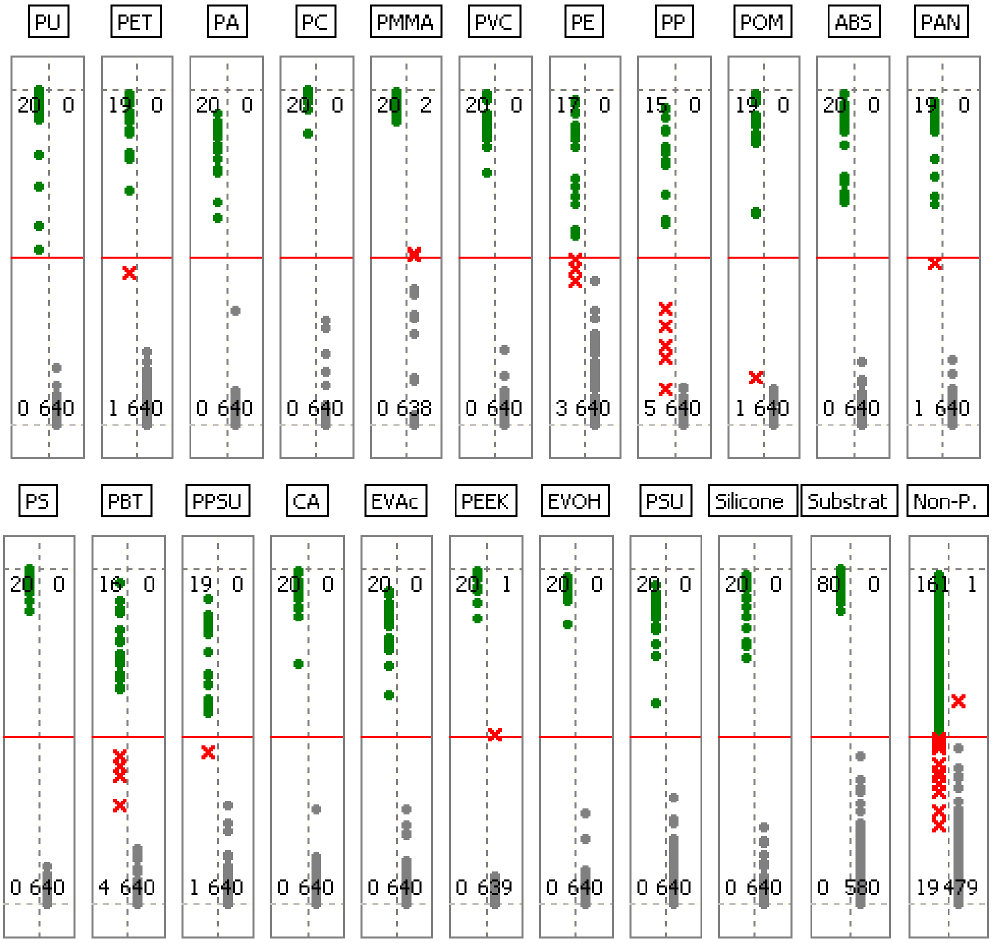


Figure 5.3: Binary confusion matrices of the binary problems calculated with the retrained experts classifier on the test set.

classified as positive or belonging to the class associated with the classifier and negative otherwise. The numbers from left to right and top to bottom are the true positives, false positives, false negatives and true negatives. For example the binary classifier for PET constructed from the feature selected set of descriptors in 5.2 exhibits TP = 18, FP = 1, FN = 2, TN = 639.

Comparing the individual binary classifiers in 5.2 and 5.3, nine of the experts binary polymer classifiers show higher accuracy $\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{P}+\mathrm{N}}$, another six have equal accuracy while the remaining five classifiers have lower ACC. The classifier for substrate compares favorably for the experts random forest with no miss-classifications, while the feature selected non-polymer classifier shows higher accuracy.

The confusion matrices depicted in figure 5.4 and 5.5 are multi-class general-

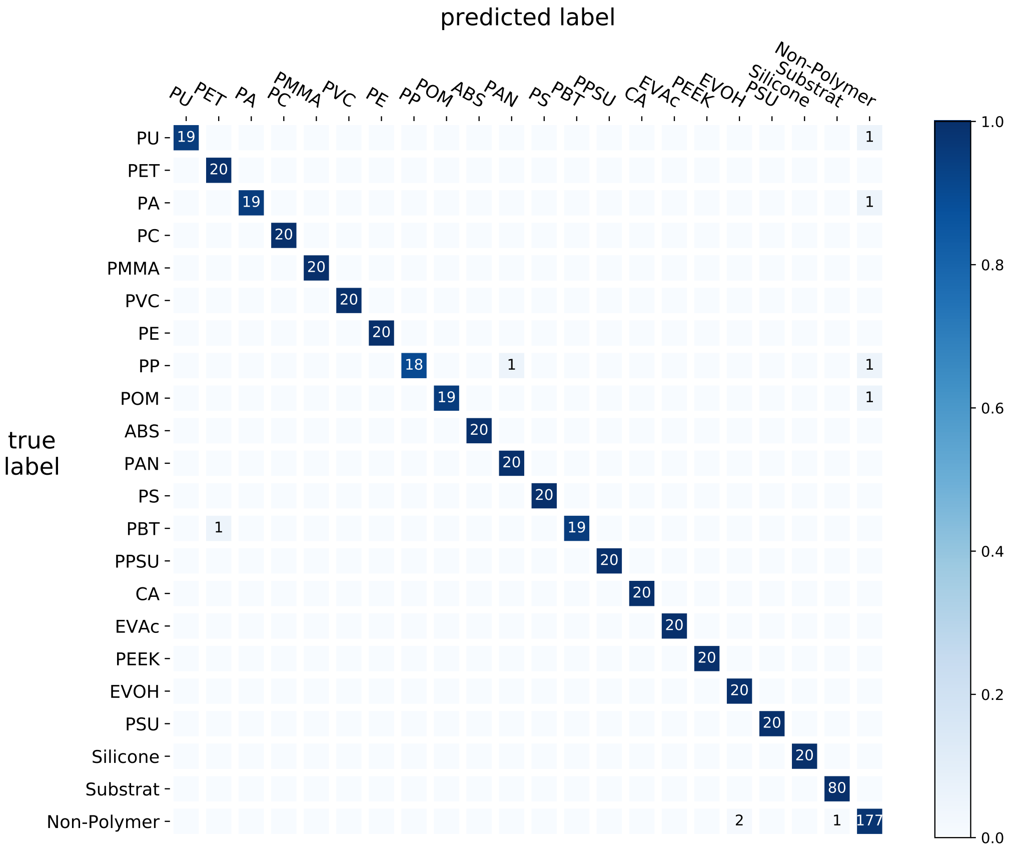


Figure 5.4: Confusion matrix of feature selected classifier validated on the test set. The labels are in absolute numbers and were omitted for empty combinations to increase readability. The color denotes the relative amount of assignments in each class.

izations of the binary ones. The rows describe the true label, while the columns present the labels predicted by the classifier. Unlike the confusion matrices of the binary classifiers above, the response of the individual classifiers are not compared to a threshold for this mapping. Simply the classifier with the highest response dictates the predicted class affiliation, which explains the discrepancy between the binary and multi-class matrices. The binary confusion matrices for non-polymer contain 10 and 19 false negatives for the feature selected and experts classifier respectively. Conversely the multi-class matrix reveals that only three and one of the non-polymer spectra produced higher responses for one of the other classes for the two classifiers. It is further evident, that some polymer samples were falsely classified as non-polymer. This effect could stem from the unequal class sizes of each polymer and the non-polymer class in the training set.

Both types of confusion matrices compare favorably for the experts feature set,

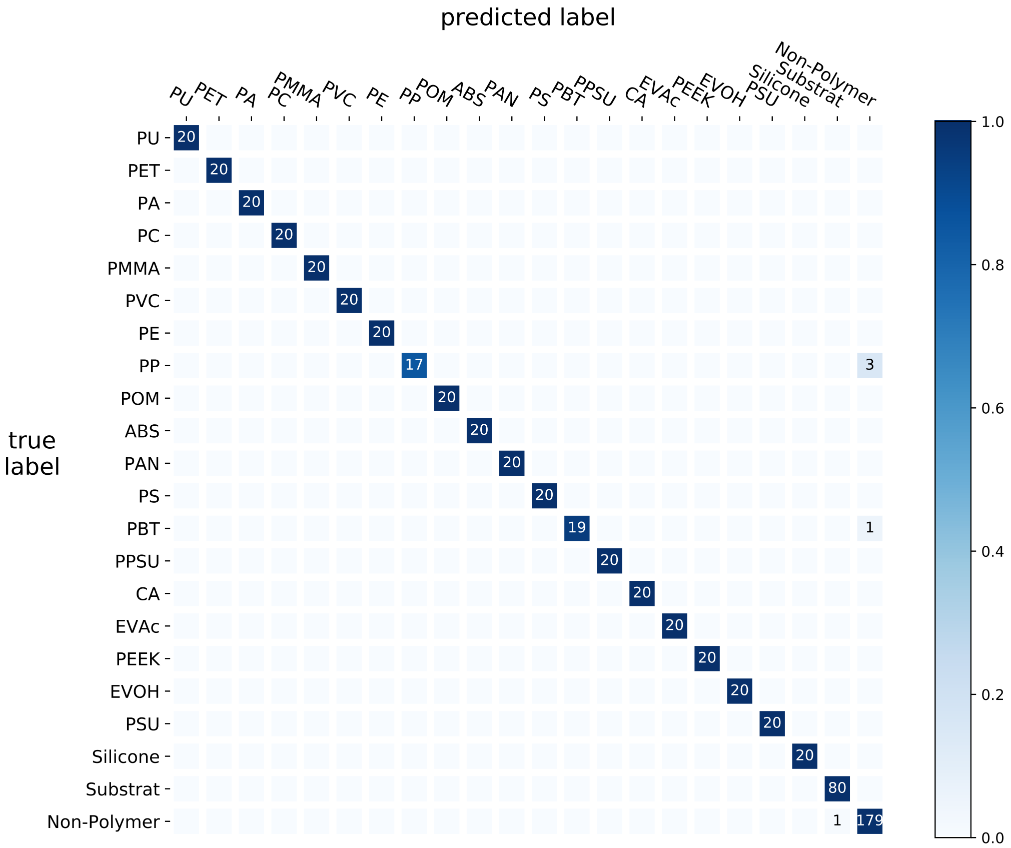


Figure 5.5: Confusion matrix of the retrained experts classifier. Three PP spectra were falsely classified as non-polymer.

but give only quantitative information about the classification performance.

5.2 Mixture Image Evaluation of ABS, PS and PAN

Running the mixture evaluation with the experimental setup described before, a table containing the values of the features described in section 3.2 for every mixture of the samples of the three polymers in the test set, hence 12k rows are created. A transposed version of the first 8 rows with values displayed in three significant digits is visible in table 5.2. The first rows until and including *total noise* describe the experiment setup or the classified mixture image, while all further rows are extracted features. Class C_1 denotes the class affiliation of the positive class corresponding with the classifier that produced the results, while class C_2 states the negative or other class. The coordinates x_i and y_i for $i \in \{1, 2\}$ refer to the

column	1	2	3	4	5	6	7	8
class C_1	ABS	ABS	ABS	ABS	ABS	PAN	PAN	PAN
x_1	73	73	73	73	73	81	81	81
y_1	57	57	57	57	57	65	65	65
class C_2	PAN	PAN	PAN	PAN	PAN	ABS	ABS	ABS
x_2	81	81	81	81	81	73	73	73
y_2	65	65	65	65	65	57	57	57
noise	0	10	25	50	75	0	10	25
total noise	1	1	1	1	1	1	1	1
midpoint x_0	48	49	46	50	38	61	58	50
offset y_0	0.981	0.983	0.916	0.573	0.39	0.958	0.96	0.912
steepness k	0.0661	0.0661	0.0631	0.0724	0.1	0.11	0.105	0.0832
amplitude a	1.03	1.03	0.984	0.608	0.375	0.982	0.985	0.936
y-range	0.953	0.961	0.9	0.577	0.365	0.967	0.971	0.907
A_p	0.151	0.151	0.151	0.151	0.151	0.189	0.189	0.189
A_b	0.25	0.25	0.25	0.25	0.25	0.147	0.147	0.147
A_n	0.599	0.599	0.599	0.599	0.599	0.664	0.664	0.664
f	0.341	0.341	0.341	0.341	0.341	0.491	0.491	0.491

Table 5.2: First 8 result rows from ABS, PS, PAN experiment.

classmap image (fig. 4.1) in section 4. The related features are calculated with the classifier belonging to the positive class C_1 (cf. column 1 and 6 relate to the exact same mixture image, classified once with the ABS and once with the PAN classifier in the columns 1 and 6 respectively.). The first five features midpoint x_0, offset y_0, steepness k, amplitude a and y-range define the logistic function fitted into the classification results of the row specified with the row label *noise*, and the 4 subsequent features describe the whole mixture image and hence are equal for all different values of *noise*.

Figure 5.6 shows the analysis of the features midpoint x_0 and offset y_0. The data in this figure is grouped by class C_1. Each group is displayed in a different color, consistent over all subplots, as can be seen in the legend in the bottom left corner. In the centered scatter plot only 200 randomly sampled data points of all $4000 = \frac{12k}{3}$ are displayed per group as not to overcrowd the plot and thus increase perceptibility. Histograms depicting the estimated distribution of the features are displayed to the left of and below the scatter plot. The bar diagrams present the mean μ in colored bars and the standard deviation σ as black error lines. Their numeric values are displayed below in table 5.3 with three significant digits. The ellipses visualize the covariance between the features in the x- and y-axis, x_0 and

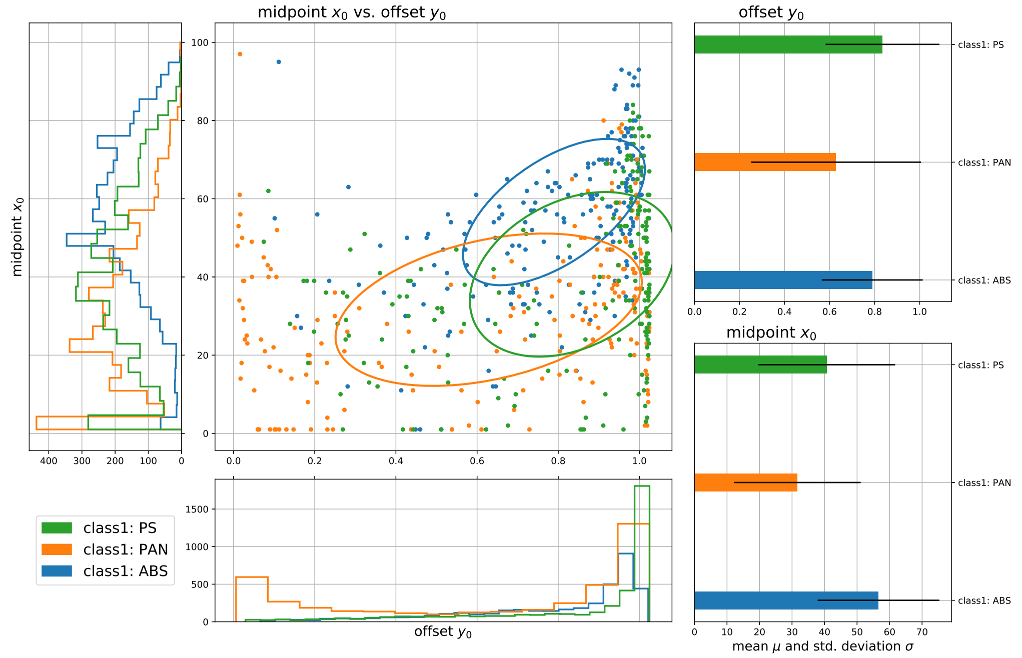


Figure 5.6: Figure showing values of the features x_0 against y_0 in a scatter plot and statistics in two histograms and bar graphs separately for both features. The data is grouped and colored by their label of class C_1.

y_0 here and are centered around the mean μ of each class. The size of the ellipsoid, or rather the dimensions of the smallest rectangle with edges parallel to the axes outlining the ellipse, denote the standard deviation of the respective dimension, hence y_0 along the x-axis and x_0 along the y-axis and coincide with the size of the black error lines in the bar graphs.

x_0 describes the inflection point of the sigmoid, with low values indicating weak detection of the positive class C_1. y_0 is the limit of the asymptote of the logistic function for $x \to \infty$ and correlates with the highest, stable, classifier response in a mixture image row. For this particular scatter plot of x_0 against y_0 we would assume the points to be centered around $(1, 50)$, since the decision boundary is to be expected around the center of x_0, hence $50 = \frac{101-1}{2}$ and the classifier response for mixtures with high ratios of class C_1 should ideally be 1.

Most noticeable in figure 5.6 is the discrepancy of the ellipse relating to PAN, being positioned further away from the ideal point of $(1, 50)$ compared to PS and ABS. This also shows in the low mean and high standard deviation in the bar graph for y_0 as well as in the long tail in the associated histogram. Looking closer at the PAN mixtures only in figure 5.7(a) reveals, that the noise is in fact responsible for the divergence of y_0 from 1. The data in this arrangement comprises only the classified mixtures with $C_1 =$ PAN, hence mixtures containing PAN and

	offset y_0		midpoint x_0	
class C_1	μ	σ	μ	σ
ABS	0.789	0.224	56.6	18.7
PAN	0.629	0.377	31.6	19.5
PS	0.835	0.253	40.7	21

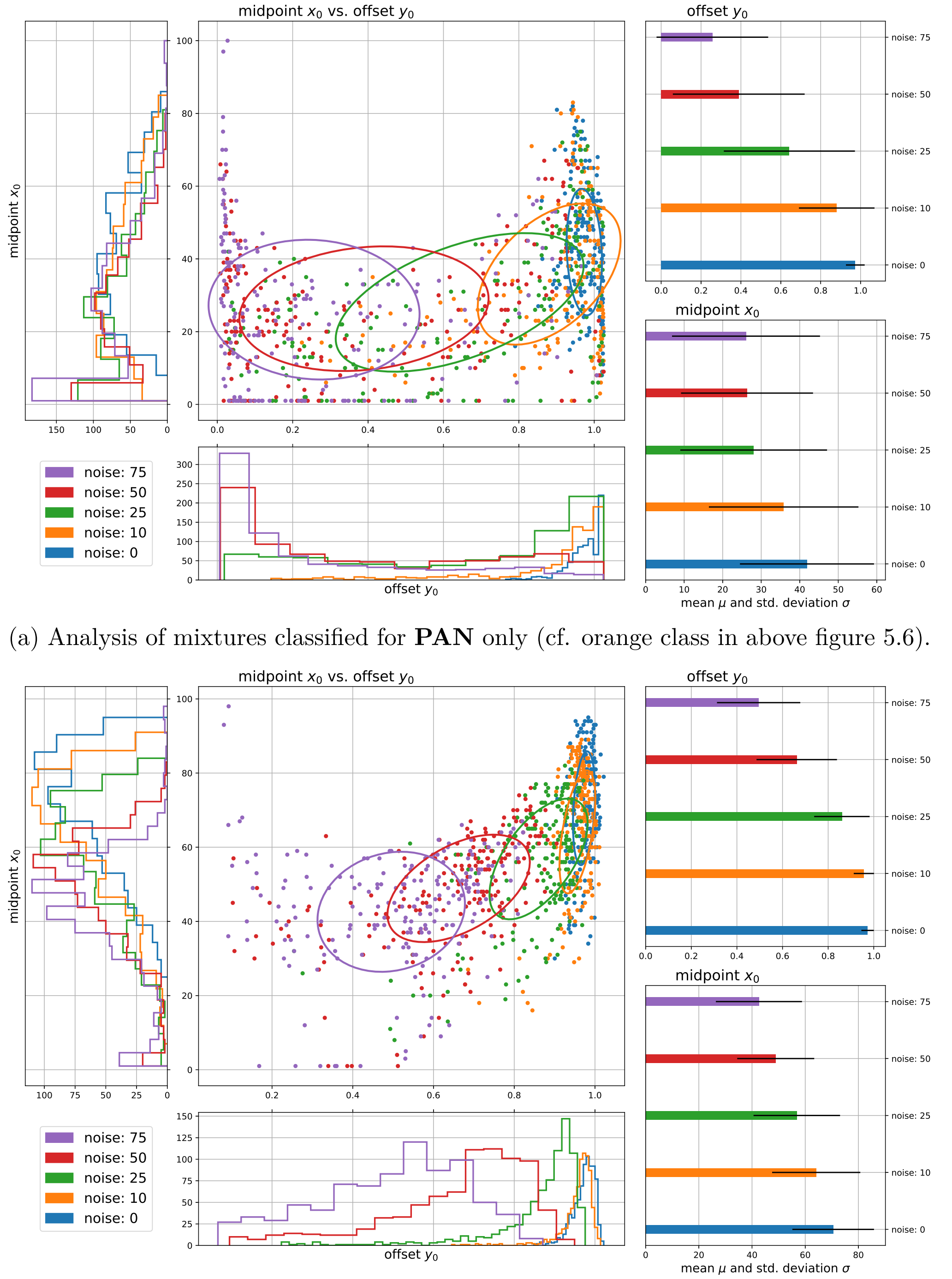
Table 5.3: Statistics displayed in the bar graphs of figure 5.6.

	PAN				ABS			
	offset y_0		midpoint x_0		offset y_0		midpoint x_0	
noise	μ	σ	μ	σ	μ	σ	μ	σ
0	0.974	0.047	41.9	17.4	0.973	0.0273	70.5	15.3
10	0.88	0.19	35.8	19.4	0.957	0.0444	64.1	16.6
25	0.643	0.329	28	19	0.86	0.122	56.8	16.3
50	0.389	0.331	26.3	17.1	0.662	0.176	48.9	14.5
75	0.257	0.28	26.1	19.2	0.495	0.183	42.6	16.2

Table 5.4: Statistics displayed in the bar graphs of figure 5.7.

classified with the PAN classifier. The same holds true for sub figure (b) and ABS. Comparing the two graphics, the different behavior of the classifiers becomes apparent. Note the different scaling and limits of the x-axis of the scatter plots and the associated histograms below. It is clearly visible, that mixtures without added noise are focused at a point close to $(1, 42)$ and rapidly diverging towards lower offset y_0 for increased noise. On the first glance the mean of the midpoint in sub figure (a) does not seem to be deteriorating with increasing noise, but with offset $y_0 \to 0$ the midpoint looses its significance, explaining the scattered points with high midpoint and offset $y_0 \approx 0$ close to zero. The ABS figure on the other hand shows much less standard deviation for both features and all noise levels. The mean of the undisturbed group is with $(0.973, 70.5)$ close to 1 for the offset and beyond 50 for the midpoint indicating, that the ABS classifier detects mixtures of ABS and both PAN and PS as ABS up to approximately 70% C_2 in the spectrum.

These results indicate bad performance of the PAN classifier in regards to noisy data in mixtures with ABS and PS. Another visible characteristic in figure 5.6 is the high midpoint mean $\mu(x_0)$ for class $C_1 =$ ABS, which is also visible in the ABS only sub figure 5.7(b). This effect stems from the fact, that the monomer of ABS itself is a mixture of styrene, acrylonitrile and butadiene. Since the monomers of PAN and PS also contain acrylonitrile and styrene respectively (see table 2.1), linear mixtures of ABS and PS or PAN are expected to be similar to ABS.



(a) Analysis of mixtures classified for **PAN** only (cf. orange class in above figure 5.6).

(b) Mixture images where class C_1 = **ABS** (cf. blue in figure 5.6).

Figure 5.7: PAN and ABS data grouped by different amounts of noise (y-coordinate of the analyzed row in the mixture image.

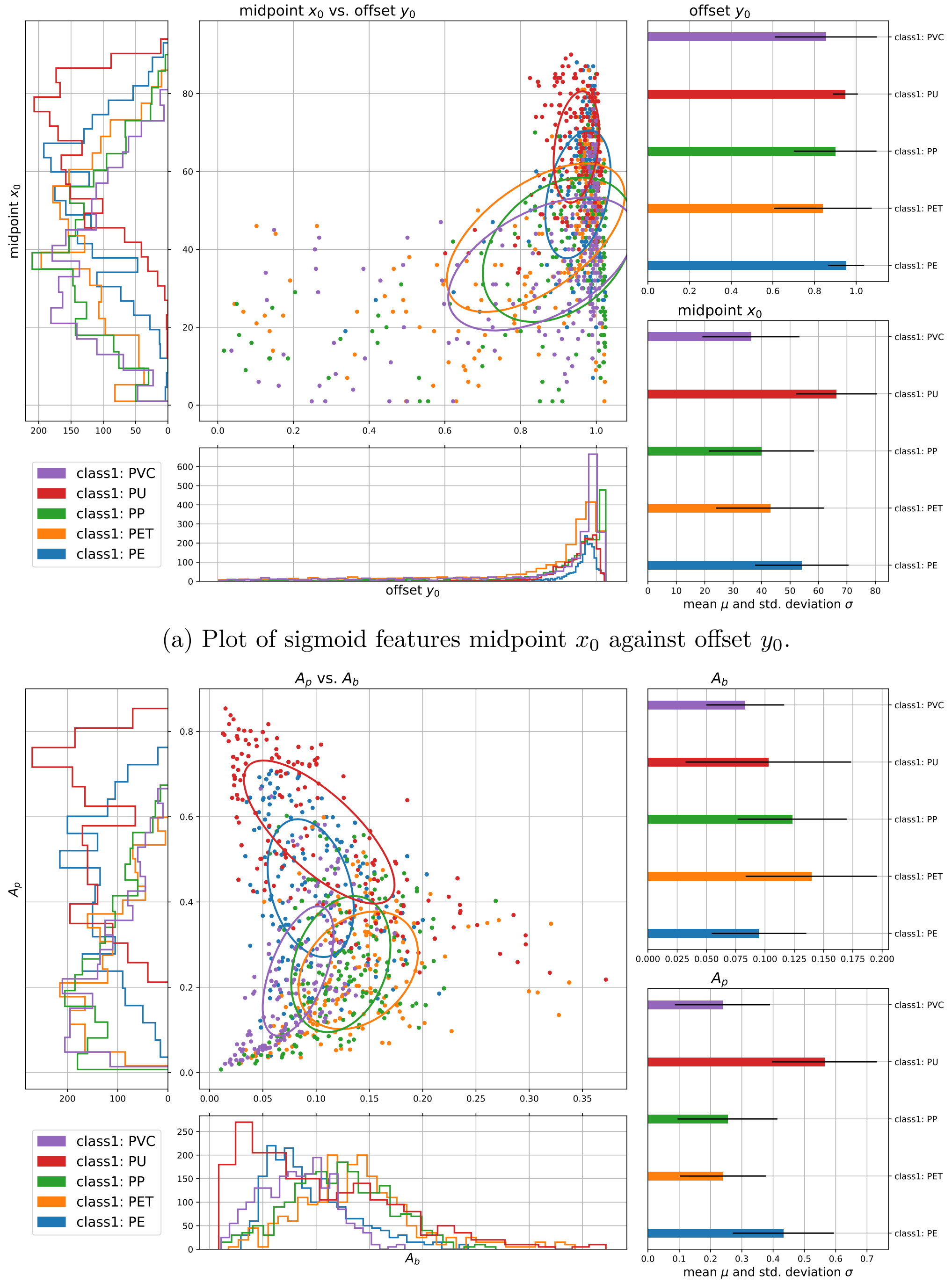
5.3 Mixture Image Evaluation of the Most Produced Polymers

The following section covers the analysis of mixtures of the five most produced polymers, including PE, PP, PVC, PET and PU. Figure 5.8 shows the results using the same sigmoid features as above, midpoint x_0 and offset y_0 in the upper sub plot, while the bottom plot depicts the areal features A_p and A_b. The numeric values of the means and standard deviations are listed in table 5.5. PU and PE show high mean and low standard deviation for offset y_0 signaling consistent high responses from the corresponding classifier for all noise settings. The other classes show much higher standard deviation and slightly lower mean suggesting more impact of added noise on the classifiers performance. The feature midpoint x_0 shows similar behavior with less severe differences in its standard deviation and more pronounced contrast regarding its mean.

	offset y_0		midpoint x_0		A_b		A_p	
class C_1	μ	σ	μ	σ	μ	σ	μ	σ
PE	0.952	0.0862	54.1	16.4	0.0949	0.0404	0.433	0.162
PET	0.84	0.235	43	19.1	0.14	0.0561	0.24	0.138
PP	0.899	0.199	39.9	18.5	0.123	0.0466	0.254	0.159
PU	0.948	0.0602	66.3	14.3	0.103	0.0707	0.564	0.168
PVC	0.854	0.245	36.2	17.1	0.0831	0.0332	0.238	0.152

Table 5.5: Statistics displayed in the bar graphs of figure 5.8.

The second sub plot 5.8(b) displays the areal feature A_p, the number of spectra in the mixture images with classifier responses between $\frac{2}{3} \leq \hat{y} \leq 1$ and A_b, the number of spectra with $\frac{1}{3} \leq \hat{y} < \frac{2}{3}$. A_b represents with the mid third of the responses those spectra, that the classifier is unsure about, with roughly half of the individual trees voting for class C_1 and the others against. A_p encompasses the top third, where most trees vote for the classification with C_1. Thus mixture images with high A_b comprise lots of spectra the classifier is unsure about and high A_p indicates good detection of class C_1. Since $A_p + A_b + A_n = 1$, the value of A_n can be calculated for any point in the scatter plot. PU shows the highest values for A_p. A tight cluster of red points can be seen around $A_p = 0.7$ meaning that 70% of the spectra in those mixture images were classified as PU with response above $\frac{2}{3}$ correlating with the high values of the midpoint x_0 feature. The purple group PVC shows lots of points close to $(0, 0)$, thus $A_n \approx 1$ suggesting bad detectability of PVC in mixtures. This could mean, that noise and mixed spectra, starting with low amounts of other polymers already rapidly deteriorate the classifier's ability



(a) Plot of sigmoid features midpoint x_0 against offset y_0.

(b) Plot of areal features A_p against A_b.

Figure 5.8: Analysis of the five most produced polymers (PE, PP, PVC, PET, PU).

to detect PVC. On the other hand however, the mean of A_p for PVC with 23.8% states that almost a fourth of spectra contained in the mixture images got high responses.

	PU				PE			
	offset y_0		midpoint x_0		offset y_0		midpoint x_0	
noise	μ	σ	μ	σ	μ	σ	μ	σ
0	0.95	0.0465	71.5	12.5	0.981	0.0213	58.6	16
10	0.953	0.0454	70.8	12.6	0.984	0.0174	58.3	16
25	0.959	0.0391	68.6	12.8	0.982	0.0196	56.7	15.7
50	0.953	0.0522	63.3	13.5	0.946	0.0684	51.8	15.3
75	0.922	0.0941	57.2	14.7	0.866	0.146	45.3	15.2

(a)

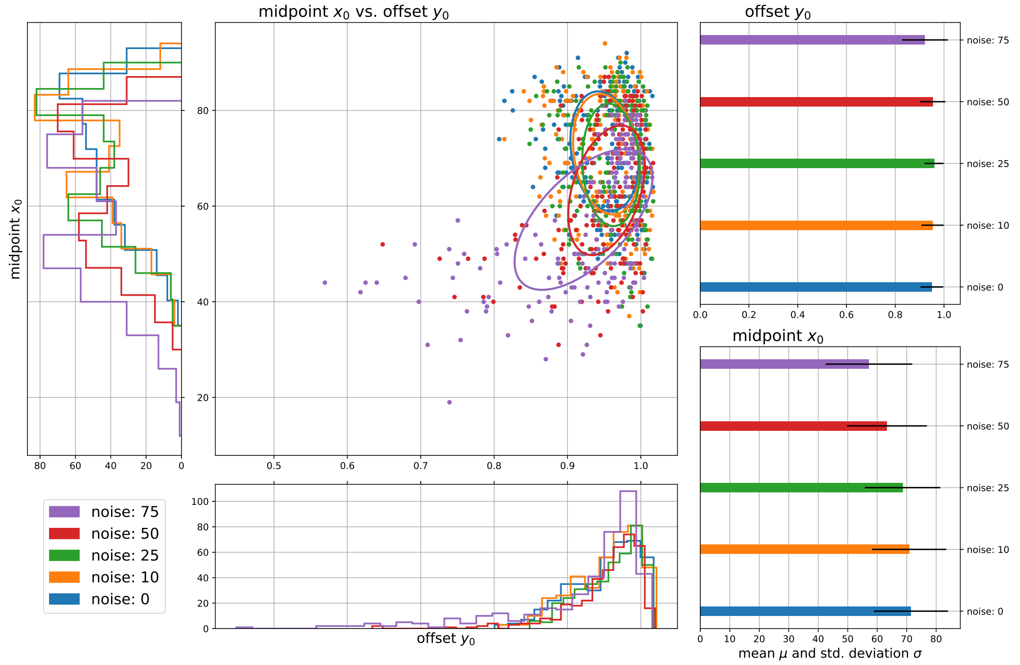
	PET				PVC			
	offset y_0		midpoint x_0		offset y_0		midpoint x_0	
noise	μ	σ	μ	σ	μ	σ	μ	σ
0	0.989	0.0201	41.7	16.6	0.955	0.0527	56.6	13.6
10	0.987	0.022	41.1	16.5	0.962	0.0433	53.4	14.5
25	0.954	0.0898	38.5	16.2	0.935	0.106	44.8	16.4
50	0.774	0.245	32.4	16	0.769	0.252	33.4	16.2
75	0.567	0.313	27.5	15.6	0.579	0.295	26.8	15.3

(b)

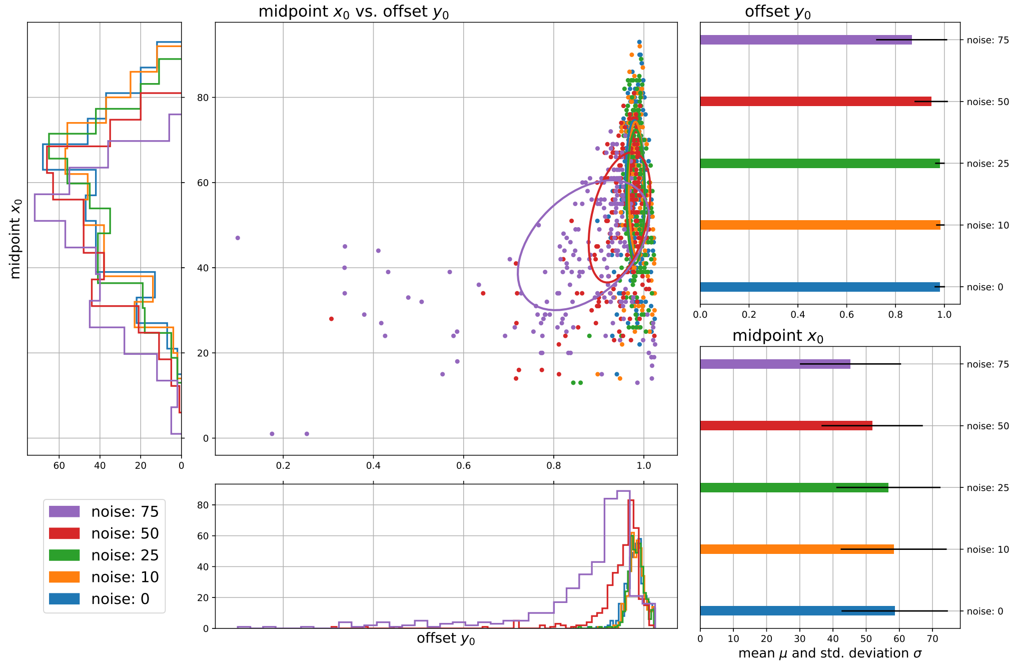
Table 5.6: Statistics of the four polymers PU, PE, PET and PVC shown in figure 5.9.

Figure 5.9 shows the classes PU, PE, PET and PVC separately and grouped by the different noise levels. Table 5.6 holds the associated statistics including mean and standard deviation. The separate views and related tables of PU and PE reveal, that their classifiers are not affected by noise up to 25%. PET and PVC on the other hand show increasing standard deviations for offset y_0 starting at 25% noise and rapidly decreasing midpoint x_0 in the case of PET.

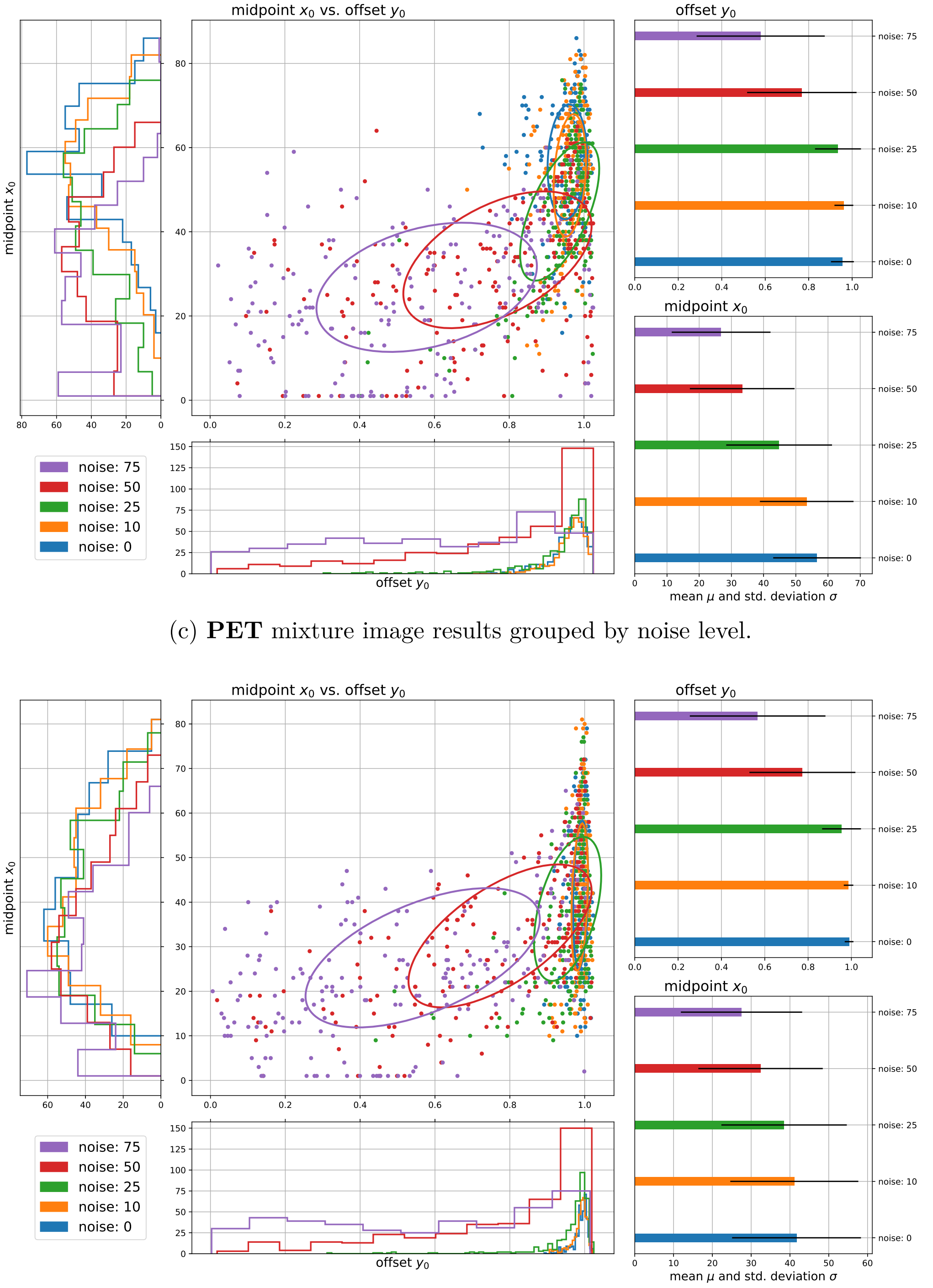
The difference, especially, in the histogram depicting the estimated distribution of offset y_0 points out the performance gaps for noisy data between PU and PE, and PET and PVC. If we look at the statistics for the midpoint x_0 feature it is evident, that out of the five polymer classifiers the PU classifier can deal best with impure spectra. It is further of interest, that the distributions of this specific feature are generally shifted towards lower values for higher amounts of added noise. From the



(a) **PU** mixture image results grouped by noise level.



(b) **PE** mixture image results grouped by noise level.



(c) **PET** mixture image results grouped by noise level.

(d) **PVC** mixture image results grouped by noise level.

Figure 5.9: Closer analysis of PU, PE, PET and PVC first analyzed in figure 5.8(a).

histograms can be seen that the extent of this shift is most pronounced for the PVC classifier, supporting the above claims of the PVC classifier loosing performance for noisy impure spectra.

5.4 Mixture Image Evaluation of Polymers vs. Non-Polymer

The last experiment treats mixtures of polymers and the non-polymer class. For this purpose mixtures of 20 different labeled polymers and non-polymers were created and classified with the corresponding polymer classifiers of all three random forests. Table 5.7 and 5.8 show the mean and standard deviation of the features

	offset y_0					
	experts trn.		experts		feature selected	
	μ	σ	μ	σ	μ	σ
ABS	0.572	0.252	0.723	0.252	0.551	0.306
CA	0.721	0.186	0.786	0.24	0.798	0.135
EVAc	0.683	0.152	0.91	0.129	0.732	0.0823
EVOH	0.645	0.322	0.664	0.386	0.824	0.195
PA	0.542	0.311	0.731	0.343	0.689	0.309
PAN	0.684	0.265	0.75	0.284	0.679	0.277
PBT	0.534	0.237	0.786	0.258	0.625	0.207
PC	0.853	0.108	0.935	0.127	0.644	0.214
PE	0.512	0.197	0.884	0.177	0.772	0.233
PEEK	0.932	0.0544	0.966	0.0459	0.734	0.192
PET	0.676	0.265	0.742	0.298	0.484	0.27
PMMA	0.826	0.212	0.773	0.273	0.691	0.212
POM	0.737	0.229	0.823	0.26	0.559	0.281
PP	0.532	0.268	0.817	0.248	0.556	0.293
PPSU	0.674	0.139	0.816	0.169	0.486	0.173
PS	0.596	0.312	0.647	0.297	0.823	0.156
PSU	0.719	0.17	0.678	0.237	0.724	0.16
PU	0.777	0.191	0.95	0.0926	0.813	0.185
PVC	0.734	0.196	0.825	0.242	0.769	0.208
Silicone	0.511	0.301	0.63	0.338	0.567	0.222

Table 5.7: Statistics showing mean and standard deviation of feature offset y_0 of all 20 evaluated polymer classifiers.

offset y_0 and midpoint x_0 for all 20 classifiers in alphabetical order. From the 20 different evaluated mixtures, five are displayed in 5.11. These five polymers are printed in bold in the table for easier reference. Midpoint x_0 and offset y_0 were selected for the visualization, because they proved to be most informative in the previous experiments.

	midpoint x_0					
	experts trn.		experts		feature selected	
	μ	σ	μ	σ	μ	σ
ABS	62.6	15.9	58	24.7	43.4	24
CA	51.7	17.8	51.4	25.1	54.1	16.2
EVAc	64.3	24.5	64.6	23.3	70.5	14.2
EVOH	42.9	28	44.6	29.5	52.6	22.6
PA	51.5	30.6	53.6	31.8	53.3	28.5
PAN	54.6	26.4	58.2	24.7	47.1	24.8
PBT	43.8	20.8	54.8	25.3	41	22.7
PC	66.9	13.9	62.5	18	49.9	19.4
PE	62.5	20.9	63.5	23.8	52	20.8
PEEK	67.5	9.9	73.2	11.6	61.4	11.6
PET	38.2	20.8	46.7	24.6	35.7	19.5
PMMA	59.5	21.6	57.5	25.6	56.5	19.9
POM	53.5	23.5	57.9	28.7	48.9	25.4
PP	48.4	19	56.5	23.8	47.6	20.4
PPSU	63.1	14.7	61.7	16.2	48.3	20.1
PS	42.2	25.9	52.9	26.4	53.3	17.1
PSU	57	14.7	57.2	16.1	60	16.4
PU	47.2	19.7	59.3	18.9	43.2	21.2
PVC	50.5	21.1	52.8	25.5	47.5	19.2
Silicone	40	25.8	47.7	30.5	47	18.7

Table 5.8: Statistics showing mean and standard deviation of feature midpoint x_0 of all 20 evaluated polymer classifiers.

The sub figure presenting the analysis of polysulfone (PSU) 5.11(a) reveals very similar results for all three classifiers. Even though the experts classifier shows lower mean for y_0, it also produced a group of points in the top right corner of the scatter plot with the highest values for this feature, indicating better performance at low noise levels. The results for polyether ether ketone (PEEK) 5.11(c) differ drastically for feature selected and the experts classifiers. The former classifier

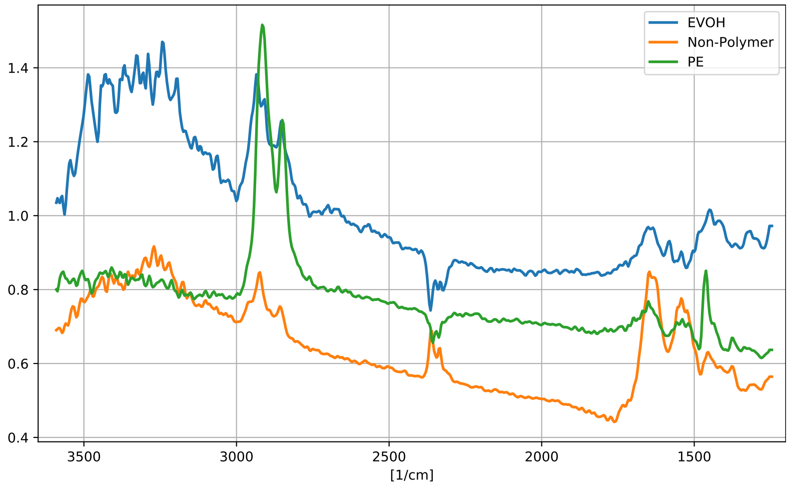
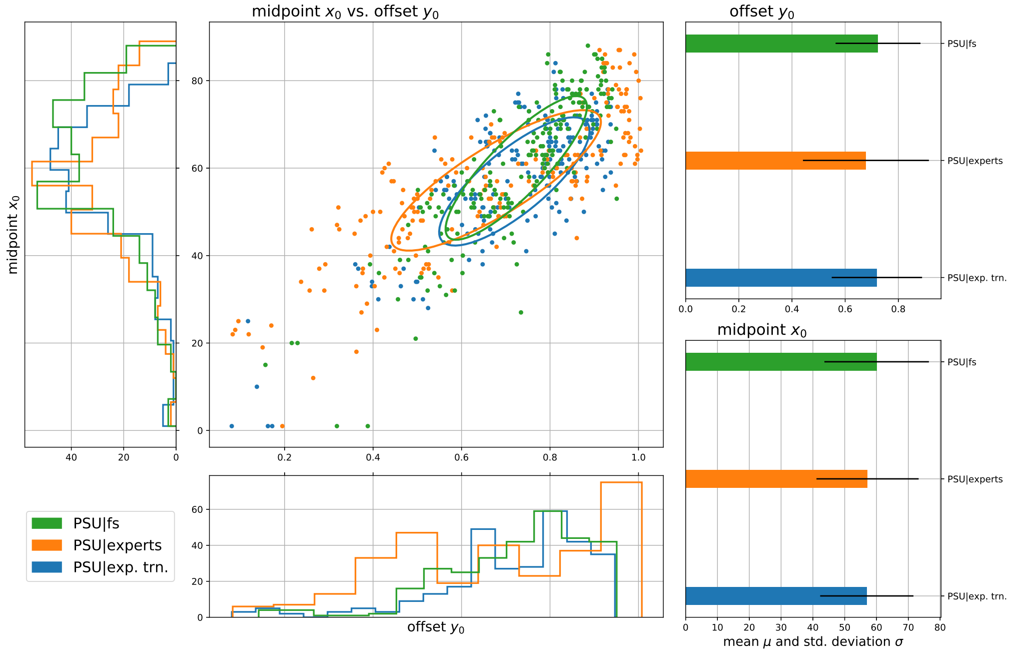
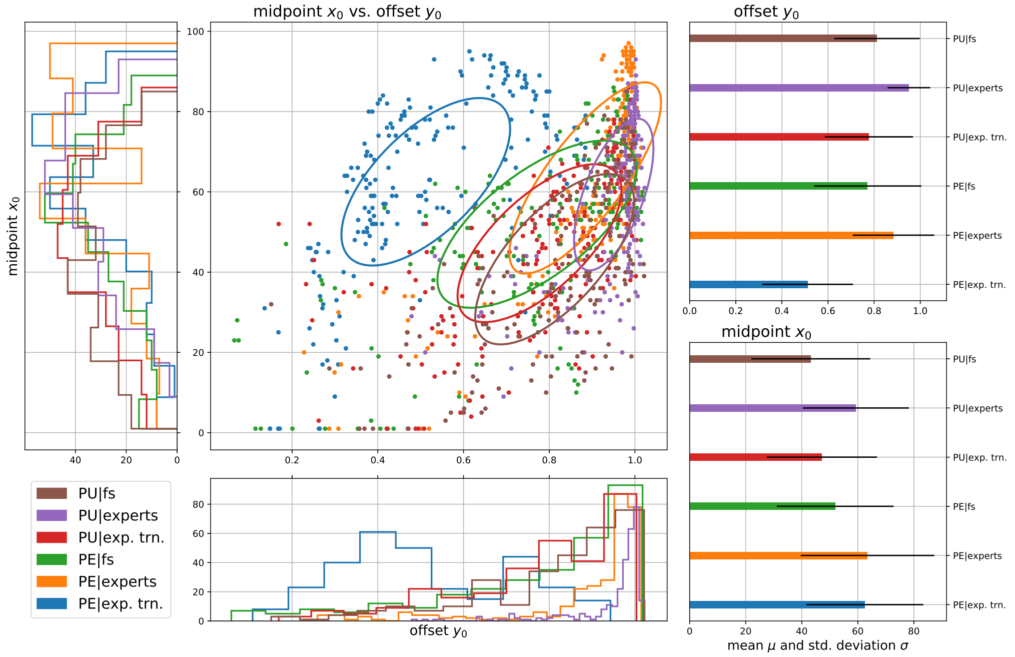


Figure 5.10

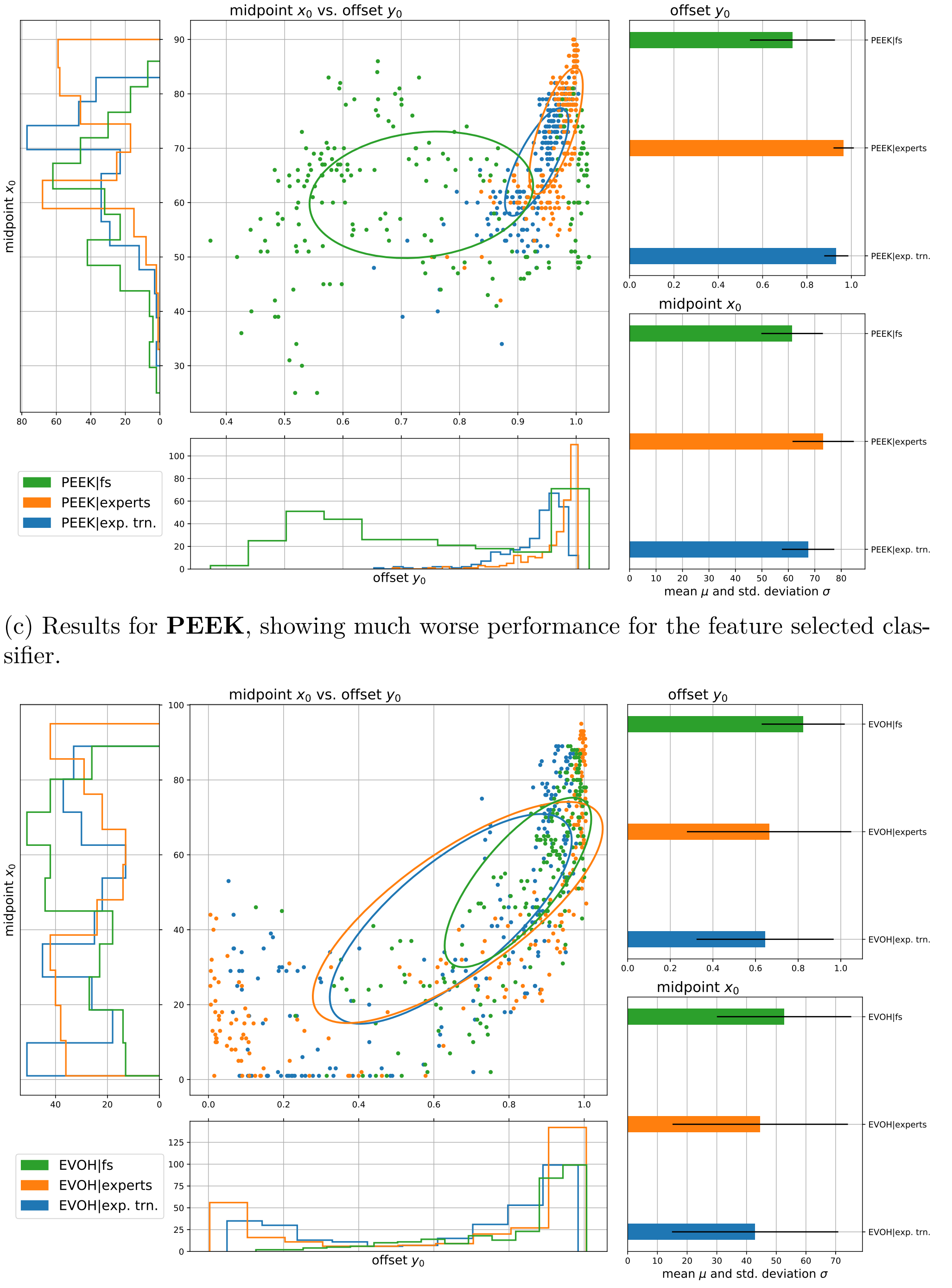
yields much worse offsets y_0, with lower μ and higher σ. The group of green data points focused around $y_0 = 1$ indicate, that the effect could be noise induced. The experts classifier and the retrained experts classifier using the same features, show comparable results. As expected however, the experts classifier outperforms the others due to its significantly larger training set. In contrast the graphic depicting ethylene vinyl alcohol (EVOH) 5.11(d) attributes best, or most stable results to the feature selected classifier, suggesting the existence of features better suited for the detection of EVOH against non-polymers in the feature selected random forest. Both experts versions produce similar statistics. That the performance of the experts EVOH classifiers is among the worst of all 20 polymer and non-polymer combinations could be explained by the broad band between 3500 cm^{-1} and 3000 cm^{-1} shared by spectra of the EVOH and non-polymer category seen in figure 5.10. The last remaining sub figure 5.11(b) depicting both PE and PU again expose the experts classifier as best in terms of its offset mean being closest to 1 and having an acceptable midpoint mean. Prominently different from the other ellipses, the retrained PE classifier is located at offset $y_0 \approx 0.5$ indicating the significance of the training set size on the performance of a classifier. The importance of the size and diversity of the training is also reflected in the differences in the columns referring to the experts and retrained experts classifiers of the tables 5.7 and 5.8. Figure 5.10 also shows the similarity between a PE and a non-polymer spectrum sampled both from the same artificially created specimens, which only increases the need for an exhaustive training set, that learns the random forest to separate these classes.



(a) **PSU** results for all three classifiers.



(b) **PU** in brown, purple, red and **PE** in green, orange, blue.



(c) Results for **PEEK**, showing much worse performance for the feature selected classifier.

(d) **EVOH**, contrary to (c) showing better performance for the feature selected classifier.

Figure 5.11: Analysis of mixtures of PU, PE, PSU, PEEK and EVOH with non-polymer.

Chapter 6

Conclusion

The set of automatically selected features produced in the first experiment displayed comparable results to the spectral descriptors (SPDCs) handpicked by spectroscopic experts. The location of the automatically selected features across the 609 layers of the investigated polymer particle spectra largely coincided with the positions of absorption bands, revealing spectroscopically important regions in the spectra.

In the latter parts a tool for the evaluation of the performance of a classifier was created. This tool analyses the behavior of the classifier when confronted with spectral mixtures and noise disturbed data. The features developed to measure performance were calculated from linear hyperspectral mixture images, classified by the classifier under examination. These mixture images measure 101 × 101 pixels and consist of artificially generated linear mixtures of two pure, selected spectra with varying amounts of noise added. The conducted experiments revealed differences in the individual binary classifiers of the three random forests examined. Among these random forests is one using the automatically selected features and a small training set consisting of 20 spectra per polymer and 180 non-polymer samples. The other two classifiers utilize the manually selected set of spectral descriptors. The random forest termed ”*experts classifier*” was trained on a data set, that in size far exceeds the training set mentioned above and used for the two other classifiers.

The experts classifier was analyzed in two experiments. The first covered mixture images of acrylonitrile butadiene styrene (ABS), polystyrene (PS) and polyacrylonitrile (PAN) and the latter contained mixtures of the five most produced polymers, including polyethylene (PE), polypropylene (PP), polyvinyl chloride (PVC), polyethylene terephthalate (PET) and polyurethane (PU). The initial experiment revealed more rapid deterioration of the responses of the PAN classifier when dealing with increasing amount of noise compared to the other two. Comparison of the five most produced polymers proved the classifiers for PU and PE to be almost unaffected by the noise added to the mixture images. Set apart were the PET and PVC binary random forests, which demonstrated to be adversely affected by higher noise levels. Lastly the experiment involving all three random forests and mixtures of polymer and non-polymer spectra resulted mostly in the experts classifier outperforming the others, owed to its significantly larger training set. Similarities in the spectra of ethylene vinyl alcohol (EVOH) and PE when compared to non-polymer caused classifiers using the manually selected set of SPDCs to produce worse results than the feature selected classifier.

During the evaluation the features offset y_0 and midpoint x_0 appeared most useful. They describe the equally named parameters of the logistic function that best approximates the values of a specific row in a classified mixture image.

Possible future improvements could include a greater variety of features extracted from the classified mixture images. One specific group of features for the assessment of noise induced behavior can be recommended. The left and right most columns of a mixture image hold the selected, pure spectra with added noise. The parameters of a logistic function approximating the response values in these columns could give information about the performance of the classifier with regards to noise only.

Bibliography

[Epi20] Epina GmbH. *ImageLab Help*. Version 3.15. 2020.

[SHC16] Douglas A. Skoog, F. James Holler and Stanley R. Crouch. *Principles of Instrumental Analysis*. 7th ed. Cengage Learning, 2016. ISBN: 978-1-305-57721-3.

[CDS13] Gary D. Christian, Purnendu K. Dasgupta and Kevin A. Schug. *Analytical Chemistry*. 7th ed. John Wiley & Sons, 2013. ISBN: 978-0-470-88757-8. DOI: 10.1007/s00216-014-7884-7.

[RD17] 'Characterization and Analysis of Microplastics'. In: *Comprehensive Analytical Chemistry*. Ed. by Teresa A.P. Rocha-Santos and Armando C. Duarte. Vol. 75. Comprehensive Analytical Chemistry. Elsevier, 2017. DOI: 10.1016/S0166-526X(17)30014-4.

[Sil+17] João Cajaiba Da Silva et al. 'Advances in the Application of Spectroscopic Techniques in the Biofuel Area over the Last Few Decades'. In: *Frontiers in Bioenergy and Biofuels*. Ed. by Eduardo Jacob-Lopes and Leila Queiroz Zepka. Rijeka: IntechOpen, 2017. Chap. 3. DOI: 10.5772/65552.

[LB05] Ira W. Levin and Rohit Bhargava. 'Fourier Transform Infrared Vibrational Spectroscopic Imaging: Integrating Microscopy and Molecular Recognition'. In: *Annual Review of Physical Chemistry* 56.1 (2005). PMID: 15796707, pp. 429–474. DOI: 10.1146/annurev.physchem.56.092503.141205.

[Löd+15] Martin Günter Joachim Löder et al. 'Focal plane array detector-based micro-Fourier-transform infrared imaging for the analysis of microplastics in environmental samples'. In: *Environmental Chemistry* 12.5 (2015), pp. 563–581.

[SV13] Enrique Saldívar-Guerra and Eduardo Vivaldo-Lima. *Handbook of Polymer Synthesis, Characterization, and Processing*. 2013. ISBN: 978-047063032-7. DOI: 10.1002/9781118480793.

[Epi13] Epina GmbH. *ImageLab*. Version 3.15. 2013–2020. URL: http://www.imagelab.at (visited on 25th Sept. 2019).

[Nod+07] I. Noda et al. 'Group Frequency Assignments for Major Infrared Bands Observed in Common Synthetic Polymers'. In: *Physical Properties of Polymers Handbook*. Ed. by James E. Mark. 2nd ed. 2007. ISBN: 978-0-387-31235-4. DOI: 10.1007/978-0-387-69002-5.

[Bas+09] Paul Bassan et al. 'Resonant Mie scattering in infrared spectroscopy of biological materials – understanding the 'dispersion artefact''. In: *The Analyst* 134.8 (2009), p. 1586. DOI: 10.1039/b904808a.

[GGB07] Paul Geladi, Hans F. Grahn and James E. Burger. 'Multivariate Images, Hyperspectral Imaging: Background and Equipment'. In: *Techniques and Applications of Hyperspectral Image Analysis*. Ed. by Hans F. Grahn and Paul Geladi. John Wiley & Sons, 2007. ISBN: 9780470010884. DOI: 10.1002/9780470010884.

[WZ15] Liguo Wang and Chunhui Zhao. *Hyperspectral image processing*. Springer, Berlin, Heidelberg, 2015. ISBN: 978-3-662-47456-3. DOI: 10.1007/978-3-662-47456-3.

[ES10] Gamal ElMasry and Da-Wen Sun. 'Hyperspectral Imaging for Food Quality Analysis and Control'. In: *Principles of Hyperspectral Imaging Technology*. Ed. by Da-Wen Sun. Academic Press, 2010, pp. 3–43. DOI: 10.1016/B978-0-12-374753-2.10001-2.

[DL97] M. Dash and H. Liu. 'Feature selection for classification'. In: *Intelligent Data Analysis* 1.1 (1997), pp. 131–156. ISSN: 1088-467X. DOI: 10.1016/S1088-467X(97)00008-5.

[LO14] Hans Lohninger and J. Ofner. 'Multisensor hyperspectral imaging as a versatile tool for image-based chemical structure determination'. In: *Spectroscopy Europe* 26.5 (2014). URL: https://www.spectroscopyeurope.com/article/multisensor-hyperspectral-imaging-versatile-tool-image-based-chemical-structure (visited on 26th Sept. 2019).

[ND05] J.M.P. Nascimento and J.M.B. Dias. 'Vertex component analysis: A fast algorithm to unmix hyperspectral data'. In: *IEEE Transactions on Geoscience and Remote Sensing* 43.4 (2005), pp. 898–910. DOI: 10.1109/TGRS.2005.844293.

[KM02] Nirmal Keshava and John F. Mustard. 'Spectral unmixing'. In: *IEEE Signal Processing Magazine* 19.1 (2002), pp. 44–57. DOI: 10.1109/79.974727.

[Bio+12] J.M. Bioucas-Dias et al. 'Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches'. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.2 (2012), pp. 354–379. DOI: 10.1109/JSTARS.2012.2194696.

[Bre01] Leo Breiman. 'Random Forests'. In: *Machine Learning* 45 (2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324.

[Ho95] Tin Kam Ho. 'Random Decision Forests'. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Aug. 1995, pp. 278–282. ISBN: 0-8186-7128-9. DOI: 10.1109/ICDAR.1995.601943.

[Ho98] Tin Kam Ho. 'The random subspace method for constructing decision forests'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8 (Aug. 1998), pp. 832–844. ISSN: 1939-3539. DOI: 10.1109/34.709601.

[Str+07] Carolin Strobl et al. 'Bias in random forest variable importance measures: Illustrations, sources and a solution'. In: *BMC Bioinformatics* 8 (2007). DOI: 10.1186/1471-2105-8-25.

[Str+08] Carolin Strobl et al. 'Conditional variable importance for random forests'. In: *BMC Bioinformatics* 9 (2008). DOI: 10.1186/1471-2105-9-307.

[SAP19] Science Advice for Policy by European Academies, SAPEA. *A Scientific Perspective on Microplastics in Nature and Society*. Tech. rep. Berlin, 11th Jan. 2019. DOI: 10.26356/microplastics.

[ABB09] Courtney Arthur, Joel Baker and Holly Bamford, eds. *Proceedings of the International Research Workshop on Microplastic Marine Debris* (9th–11th Sept. 2008). NOAA Technical Memorandum NOS-OR&R-30, Jan. 2009. URL: https://marinedebris.noaa.gov/proceedings-international-research-workshop-microplastic-marine-debris (visited on 10th Dec. 2019).

[GES15] GESAMP. 'Sources, fate and effects of microplastics in the marine environment: a global assessment'. (IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). In: *GESAMP Reports and Studies* 90 (2015). Ed. by P. J. Kershaw, p. 96. URL: http://www.gesamp.org/publications/reports-and-studies-no-90 (visited on 10th Dec. 2019).

[And11] Anthony L. Andrady. ‘Microplastics in the marine environment’. In: *Marine Pollution Bulletin* 62.8 (2011), pp. 1596–1605. ISSN: 0025-326X. DOI: 10.1016/j.marpolbul.2011.05.030.

[Bro15] Mark A. Browne. ‘Sources and Pathways of Microplastics to Habitats’. In: *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow and Michael Klages. 2015, pp. 229–244. DOI: 10.1007/978-3-319-16510-3_9.

[Lus15] Amy Lusher. ‘Microplastics in the marine environment: Distribution, interactions and effects’. In: *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow and Michael Klages. 2015, pp. 245–307. DOI: 10.1007/978-3-319-16510-3_10.

[GES16] GESAMP. ‘Sources, fate and effects of microplastics in the marine environment: part two of a global assessment’. (IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). In: *GESAMP Reports and Studies* 93 (2016). Ed. by P. J. Kershaw and Chelsea M. Rochman, p. 220. URL: http://www.gesamp.org/publications/microplastics-in-the-marine-environment-part-2 (visited on 10th Dec. 2019).

[KBV15] Susanne Kühn, Elisa L. Bravo Rebolledo and Jan A. Van Franeker. ‘Deleterious effects of litter on marine life’. In: *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow and Michael Klages. 2015, pp. 75–116. DOI: 10.1007/978-3-319-16510-3_4.

[VBK12] Nadia Von Moos, Patricia Burkhardt-Holm and Angela Köhler. ‘Uptake and Effects of Microplastics on Cells and Tissue of the Blue Mussel Mytilus edulis L. after an Experimental Exposure’. In: *Environmental Science & Technology* 46.20 (2012), pp. 11327–11335. DOI: 10.1021/es302332w.

[LLD11] Delilah Lithner, Åke Larsson and Göran Dave. ‘Environmental and health hazard ranking and assessment of plastic polymers based on chemical composition’. In: *Science of The Total Environment* 409.18 (2011), pp. 3309–3324. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2011.04.038.

[Roc15] Chelsea M. Rochman. ‘The complex mixture, fate and toxicity of chemicals associated with plastic debris in the marine environment’. In: ed. by Melanie Bergmann, Lars Gutow and Michael Klages. Marine Anthropogenic Litter. 2015, pp. 117–140. DOI: 10.1007/978-3-319-16510-3_5.

[Pri+18] Sebastian Primpke et al. 'Reference database design for the automated analysis of microplastic samples based on Fourier transform infrared (FTIR) spectroscopy'. In: *Analytical and Bioanalytical Chemistry* 410.21 (1st Aug. 2018), pp. 5131–5141. ISSN: 1618-2650. DOI: `10.1007/s00216-018-1156-x`.

[GJL17] Roland Geyer, Jenna R. Jambeck and Kara Lavender Law. 'Production, use, and fate of all plastics ever made'. In: *Science Advances* 3.7 (2017). DOI: `10.1126/sciadv.1700782`.