APPLICABILITY OF HANDHELD RAMAN SPECTROMETER FOR PLANT HEALTH MONITORING

Lappeenranta-Lahti University of Technology LUT

Bachelor's Program in Computational Engineering, Bachelor's Thesis

2023

Matias Heikkinen

Examiner: Prof. Satu-Pia Reinikainen
Zina-Sabrina Duma
Tuomas Sihvonen

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Computational Engineering

Matias Heikkinen

Applicability of Handheld Raman Spectrometer for Plant Health Monitoring

Bachelor's thesis
2023
40 pages, 17 figures, 2 tables, 4 appendices
Examiners: Prof. Satu-Pia Reinikainen, Zina-Sabrina Duma and Tuomas Sihvonen

Keywords: Raman, Basil, spectroscopy, spectra, pollutant, plant

In this study, Basil plants are monitored with a handheld Raman spectrometer. The objective is to find out if a handheld Raman spectrometer can be used as a monitoring tool for plant pollution and suggest a methodology to treat the spectra for maximum information gain. For monitoring plant health, the experimental part is designed to observe different quantities of pollution by performing quantitative tests, qualitative tests, and evolution tests. The Basil plants are polluted with four different pollutants, sulfate salts of Zn, Co, Fe, and Cu, metals that represent either micronutrients found in fertilizers or metals found in mining wastewater. The effects of sulfate salts were able to be detected visually already after 24 hours of exposing the samples to the pollutants. The differences could also be seen from the spectrum making the handheld Raman spectrometer a great tool for monitoring plant health and for detecting possible plant diseases early on. From spectral effects, the samples polluted with $CuSO_4$ differ the most from the healthy plant, based on the captured variation of the first two principal components in the control model. The observed intensities of $FeSO_4$ were found to be higher than the intensities observed from the healthy plants. However, the plant was least affected when polluted with $ZnSO_4$. The visual observations and PCA analysis support the conclusion. This study proposes that the best methodology to separate the effects of the pollutants from the healthy plants is by removing the background from the spectra by using statistical methods and then applying multivariate analysis to visualize the differences.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
Teknillinen korkeakoulu
Laskennallinen tekniikka

Matias Heikkinen

Käsikäyttöisen Raman-spektrometrin soveltuvuus kasvien terveyden seurantaan

Kandidaatin työ
2023
40 sivua, 17 kuvaa, 2 taulukkoa, 4 liitettä
Tarkastajat: Prof. Satu-Pia Reinikainen ja Zina-Sabrina Duma ja Tuomas Sihvonen

Hakusanat: Raman, Basilika, spektroskopia, spektri, saaste, kasvi

Tässä tutkimuksessa seurataan muutoksia basilikassa Raman spektrometrillä. Tutkimuksen tavoitteena on selvittää, kuinka muutokset näkyvät mitatussa spektrissä ja ehdottaa menetelmä niiden käsittelemiseksi. Kokeellinen osa on suunniteltu havainnoimaan spektrin eri muutoksia käyttämällä kvantitatiivisia testejä, kvalitatiivisia testejä ja evoluutiotestejä. Kokeellista osaa varten basilika altistettiin neljän eri metallin Zn, Co, Fe ja Cu sulfaattisuoloille, jotka edustavat joko lannoitteiden hivenravinteita tai kaivosjätevesien metalleja. Sulfaattisuolojen vaikutukset pystyttiin havaitsemaan visuaalisesti jo 24 tunnin kuluttua siitä, kun näytteet oli altistettu niille. Erot näkyivät myös spektristä, jonka perusteella kädessä pidettävää Raman-spektrometriä voidaan pitää varteenotettavana työkaluna kasvien terveydentilan seurantaan ja mahdollisten kasvitautien varhaiseen havaitsemiseen. Vertaillessa sulfiittisuolojen vaikutuksia basilikan spektriin, $CuSO_4$:lla saastuneet näytteet eroavat eniten terveestä kasvista. $FeSO_4$:lla saastutettujen näytteiden havaitut Raman-sironta intensiteetit olivat korkeammat kuin terveistä kasveista havaitut intensiteetit. Silti kasvi kesti parhaiten $ZnSO_4$ altistamisen. Suoritettu PCA-analyysi ja visuaaliset havainnot vahvistavat johtopäätöksen. Kasvin spektri muutosten erottamiseen tämä tutkimus pitää parhaana menetelmänä tapaa, jossa kasvin spektristä poistetaan ensin ylimääräinen taustan vaikutus käyttämällä tilastollisia menetelmiä, jonka jälkeen kasvin spektrien erot visualisoidaan pääkomponentti analyysillä niiden erottamiseksi.

CONTENTS

Appendix 5: Raw Spectra

1 INTRODUCTION

Raman spectrometry is a relatively recent subject in research. Although C.V. Raman discovered Raman scattering in 1928, it took decades for laser technology to evolve to a level high enough for obtaining accurate enough spectra for Raman spectroscopy [1,2]. A search made in the Scopus Elsevier database reveals that the number of articles published about Raman spectrometry annually has multiplied by seven in the past 20 years as can be observed from Figure 1. There are various fields in which Raman spectroscopy can be applied, but the most recent research has focused mostly on agriculture and surface analysis.

Humanity has found itself in a situation where there is a growing demand for field areas to cover food insecurity. The global population is growing [3–5] which leads to a growing need for food supplies to be able to feed everyone [6]. Carbon sinks cannot be lost. The best option for securing food adequacy for everyone with the current resources available is to mitigate yield loss. Researchers have shown interest in using Raman spectroscopy as a precision tool in agriculture [7,8]. Ideally, farmers could monitor yield with a handheld Raman spectrometer [4]. With the data gained from the spectra, the farmer could take action to respond to plant stress [7]. Yield losses have been roughly estimated to be even 20 to 40% of global agriculture production [3].

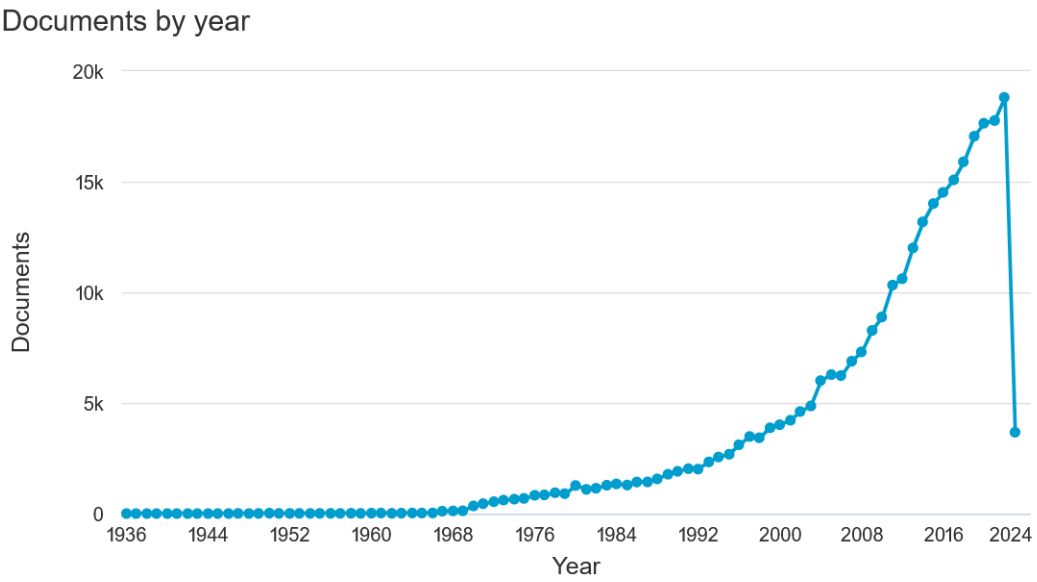

Figure 1. Growing interest in Raman Spectroscopy over the years, reflected in the number of published articles in 3/2023

1.1 Research Objective

The focus of the research is to determine if a handheld Raman spectrometer can be used as a monitoring tool for plant pollution. The research has two objectives: to assess Raman spectroscopy's reliability in examining plant health and to propose a methodology for examining spectral changes in plant experiments. The research objectives can be seen as answers to the following research questions. The first research question is:

1. *Can a handheld Raman spectrometer be used to monitor plant health?*

In this research objective, the interest is how reproducible are the measurements, what is the ratio signal-to-noise of the measurement, how difficult it is to create reliable and reproducible measurements, and if the measured plant is affected by the measurement. For analyzing the effects of pollutants more. The second research question was formed:

2. *What are the spectral effects of pollutants in Basil?*

This research aims to determine if there are visible spectral differences between healthy and polluted plants.

3. *What are the mathematical tools that we can apply to Raman spectra to observe pollutant differences?*

Different tools exist for spectral handling. This study aims to propose an optimal pretreatment procedure, spectral modeling techniques, and diagnostic tools for evaluating plant pollution.

1.2 Research Design

The research is divided into an experimental part and a mathematical modeling part. For the experimental part, a Basil plant, *Ocimum basilicum*, has been studied for its interest in many fields of research [9]. To monitor plant health, the Basil plant is polluted with four different pollutants. The pollutants are sulfate salts of Zn, Co, Fe, and Cu, metals that represent either micronutrients found in fertilizers or metals found in mining wastewater. The measurements are taken at 24 and 48h intervals, for each pollutant, with an additional unpolluted control sample being monitored.

To validate that the measurements are reliable for plant health monitoring and to answer the first research question, firstly heterogeneity tests are applied, which reveal if different locations of a healthy plant give consistent spectra. Secondly, intensity tests are taken, to see if the plant is affected by the measurement, and to see which laser intensity gives the most consistent spectra with a good signal-to-noise ratio.

For answering the second research question on the spectral effects of pollutants, quantitative tests, qualitative tests, and evolution tests are applied. Quantitative tests are for revealing how different amounts of pollutants affect the Raman spectra. In qualitative tests, it is evaluated how different pollutants change the plant's spectra. Ultimately the effect of exposure times on the spectra is assessed with evolution tests.

To answer the third research question and develop a mathematical procedure for handling and interpreting spectral differences, spectral pretreatment methods are tested, along with multivariate analysis methods used in literature for Raman monitoring and other spectral monitoring methodologies.

To validate the discoveries, the practical observations are compared with existing literature on plant health monitoring in Raman spectroscopy. The reliability of measurements is defined by comparing the difference between the control samples formed out of experimental measurements and by comparing the differences in effects of pollutants between experimental measurements to the ones in the literature.

2 THEORY

2.1 Raman Spectroscopy

In Raman spectroscopy, changes in molecules' polarization ellipsoid are observed and converted to Raman shifts [2,10,11]. Raman shift is a standard scale of measurement. One of the best features of Raman spectroscopy is that the sample does not need to be prepared. It is also one of the reasons the popularity of Raman spectroscopy has increased rapidly among many areas of research Figure 1. Unlike lasers with longer wavelengths, lasers used in Raman spectroscopy function in visible light wavelengths or in short infrared wavelengths. Because of that, they can pass through common sources allowing the user to take spectra from samples that are in the water or behind glass [2].

2.1.1 Raman Scattering

Raman scattering is used to identify specific molecule structures of material. In Raman scattering, different molecules emit a photon with a certain wavelength depending on the wavelength of the incoming radiation [2,10]. Every molecule has its own unique identifier - the Raman shift [11]. The same molecule might have multiple different unique Raman shifts, but certain Raman possible shifts can always be expected if the wavelength of the incoming radiation is known.

Raman scattering is a two-photon event [2]. Incoming radiation changes the molecule's energy balance temporarily boosting the molecule to a higher virtual energy state [2, 10]. Excess energy is stored as a molecule's vibrational motion before energy is released either entirely or partly. If the emitted energy is as much as the absorbed energy, the phenomenon is called Rayleigh scattering, but if the emitted energy is more or less than the molecule's absorbed energy, it is Raman scattering. The chance for a photon to Raman scatter is one in a million [2].

2.1.2 Quantum Theory in the Functionality of Raman Spectrometer

Raman spectrometers measure the difference in wavelength of transmitted and incoming radiation. Because the wavelength of the laser's radiation is known, the difference can be calculated by subtracting the incoming radiation's wavelength from the transmitted

wavelength [2]. In terms of energy, because of the conservation of energy, the difference in energy must be as much as the difference between the energy of transmitted radiation E_0 of the spectrometer and energy incoming radiation E_1 of the spectrometer:

$$\Delta E = E_1 - E_0 \tag{1}$$

Energy can be described as a function of wavelength λ, where c is the speed of light and h is Planck's constant.

$$\frac{hc}{\Delta\lambda} = \frac{hc}{\lambda_1} - \frac{hc}{\lambda_0} \tag{2}$$

When Planck's constant and speed of light can be divided from both sides, there are only terms of wavelength left [2, 10, 11]. The terms of $\frac{1}{\lambda}$ are describing Raman shifts, and are subscribed as ν so the final equation is formed as:

$$\Delta\nu = \nu_1 - \nu_0 \tag{3}$$

If the Raman shift of transmitted radiation ν_1 is bigger than the Raman shift of incoming radiation measured by the spectrometer ν_0, the difference between them $\Delta\nu$ is negative. If the difference is negative, the molecule has received energy and the phenomenon is Stokes scattering. If the difference is positive, then the molecule has released energy, and the phenomenon is anti-Stokes scattering.

The intensity I of observed Raman shifts can be predicted by using Equation 4 [12].

$$I \sim \nu^4 I_0 N (\frac{\partial a}{\partial q})^2 \tag{4}$$

In the Equation 4 ν is laser power, I_0 is the intensity of the excitation laser, N is the number of scattered molecules, and $\frac{\partial a}{\partial q}$ is the change in polarization.

2.1.3 Spectral Monitoring of Basil in Raman Spectroscopy

There can be found articles of Basil in the context of Raman spectroscopy but in the category of plant health monitoring, Basil has not been studied yet. The effects of fertilizers like Zn [7] and Cu [8] on Basil and its essential oil products can be found in studies conducted during past years as well as Basil essential oil quality control in industry [13]. The essential oil of Basil has been studied on a deeper level to characterize the chemical composition of the oil [13–15]. Most of the studies in the field of Raman spectroscopy

focus on observing the essential oil of Basil instead of detecting features of Basil plants.

Some of the literature includes visualization of area normalized spectra of a healthy Basil. To be able to compare the differences spectra-wise, Figure 2 was created by displaying the spikes of the Raman shifts with Gaussian distributions fitted to the Raman shifts and intensities found from the articles.

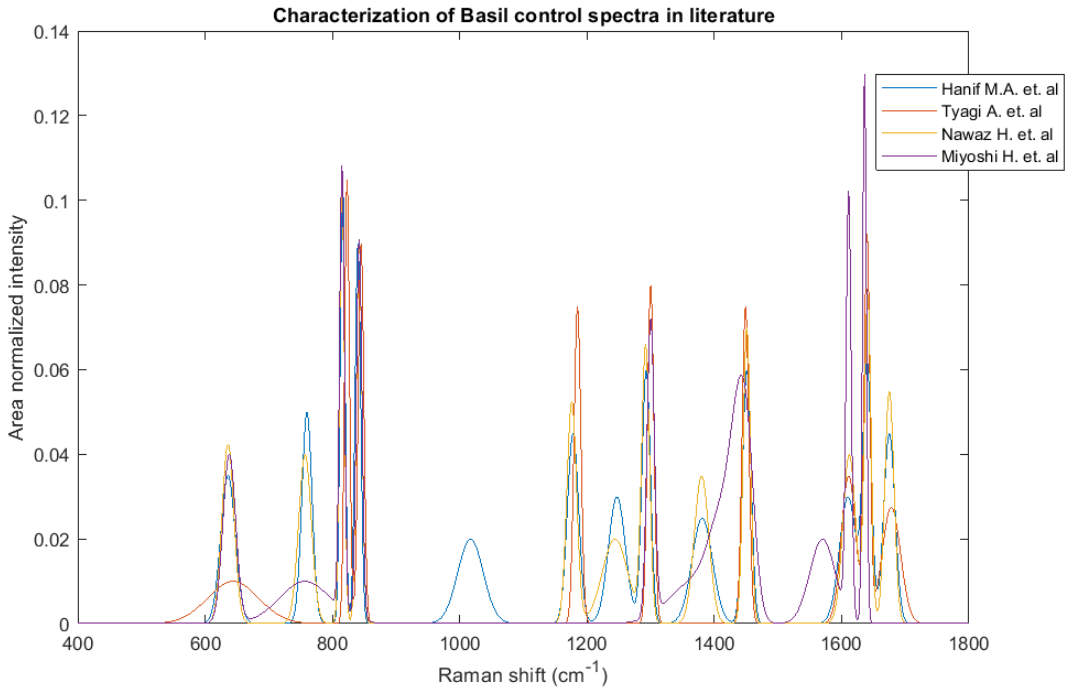

Figure 2. Characterization of most important spikes of area normalized control spectra from literature: Hanif et. al [7], Tyagi et. al. [13], Nawaz et. al. [8], Miyoshi et. al. [14]

Figure 2 shows that the locations of the spikes of the spectra of the essential oil of Basil should be reasonably consistent. There are still some differences between the spectra of area-normalized healthy Basil. Hussain et. al. suggests that there are changes in Basil's chemical composition based on seasonal and geographical variations [16].

The area-normalized spectrum of Miyoshi H. et. al differs the most from the rest. It has a spike on 1570 Raman shifts that cannot be found from any other spectrum (Figure 2). It is also missing the spikes at 1180 and 1675 Raman shifts that can be found in every other spectrum. The intensity discovered by Miyoshi et. al. at the 1611 Raman shifts is much higher than in the other spectrum, although it also has a spike at the 1636 Raman shift also higher than in the other spectra.

Hanif et. al. have discovered a spike at 1017 Raman shifts that cannot be found in other spectra. The spectrum of Hanif et. al. has also a spike at 1245 Raman shifts, and at 1382 Raman shifts with Nawaz et. al. spectrum that cannot be found in Miyoshi et. al or Tyagi et. al. spectrum (Figure 2). Overall, the spectra of Basil in different research can be considered consistent.

2.2 Mathematical Methods

2.2.1 Savitzky-Golay Filtering

A Savitzky-Golay filter is a tool used in signal processing to the signal. The Savitzky-Golay filter is a convolution filter and it is based on local least-squares fitting [17]. In Savitzky-Golay filtering, as shown in Figure 3, polynomials are fitted in subsets of the data by the linear least squares method. If the data points are equally spaced, a kernel matrix to map the subset into a noise-reduced signal can be derived analytically [17]. The kernel matrix also solves differences in the system, but in Savitzky-Golay filtering only the coefficient of the polynomial is used. The Savitzky-Golay filter is overall a very well-functioning method, but some of the first and last points of the signal's chain will be lost in the process, or for those, there needs to be some separate filter.

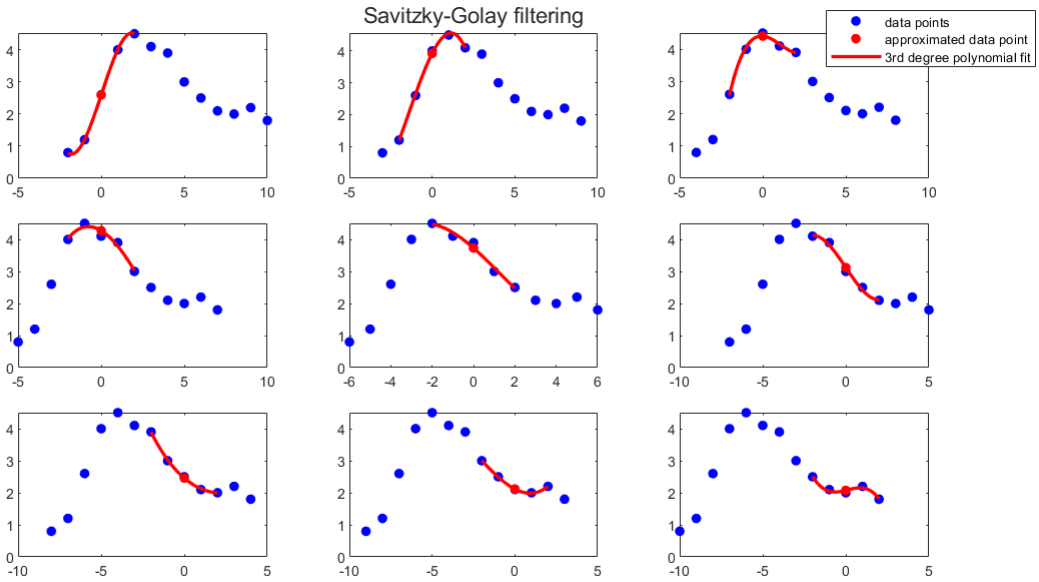

Figure 3. Savitzky-Golay 3-degree polynomial filtering: the filtered value is represented as the middle point of the filter

To form a Savitzky-Golay filter, there must be decided the degree of the fitted polynomial and the length of the subset where the polynomial is to be fitted. The length of the subset has to be an odd number and it cannot be bigger than the length of the signal itself. Then the polynomial formed to describe the subset of data points $\mathbf{y}$, $\mathbf{y}_m \subset \mathbf{y}$, can be represented as:

$$c_0 + c_1\mathbf{x}_m + c_2\mathbf{x}_m^2 + ... + c_{n-1}\mathbf{x}_m^{n-1} + c_n\mathbf{x}_m^n = \mathbf{x}_m \tag{5}$$

In Equation 5, n is the degree of the polynomial and m is the length of the subset. $\mathbf{x}_m$ is a subset of $\mathbf{x}$ that has been centered to the y-axis (Figure 3) by subtracting the center value of the subset. The equation 5 can be derived into a form of linear combination:

$$\mathbf{X}_{m*n}\mathbf{c}_{n*1} = \mathbf{y}_{m*1} \tag{6}$$

$$\mathbf{X} := \begin{pmatrix} 1 & x_0 & x_0^2 & ... & x_0^{n-1} & x_0^n \\ 1 & x_1 & x_1^2 & ... & x_1^{n-1} & x_1^n \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_{m-1} & x_{m-1}^2 & ... & x_{m-1}^{n-1} & x_{m-1}^n \end{pmatrix} \mathbf{c} := \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{pmatrix}$$

Because Savitzky-Golay finds the filtered value to the center point of the subset $\mathbf{x}_m$, the coefficient c_0 will then be the filtered value for the point. The approximative solution for the coefficients can be solved by the linear least squares method from equation 6 by using orthogonal projection:

$$\mathbf{c}_{n*1} \sim (\mathbf{X}_{n*m}^T\mathbf{X}_{m*n})^{-1}\mathbf{X}_{n*m}^T\mathbf{y}_{m*1} \tag{7}$$

Other coefficients are filtered values for derivatives of the process.

Moving average smoothing is a special case of the Savitzky-Golay filter. It responds Savitzky-Golay executed with a zero-degree polynomial. The reason moving average filtering is a special case of Savitzky-Golay is that; unlike other fitted polynomials of

Savitzky-Golay, the moving average is not dependent on the x-values of the signal.

In the moving average smoothing, the filtered value c_0 is described as the average of the center value of subset $\mathbf{y}_m$. Where $\mathbf{y}_m$ is a subset of data points $\mathbf{y}$, $\mathbf{y}_m \subset \mathbf{y}$, and m is an odd number subscribing to the length of the subset.

$$c_0 = \frac{\sum_m y_m}{m} \tag{8}$$

The same equation is possible to derive from zero-order Savitzky-Golay:

$$\begin{aligned} \mathbf{X}_{m*1} c_{1*1} &= \mathbf{y}_{m*1} \\ c_0 &\sim (\mathbf{X}_{1*m}^T \mathbf{X}_{m*1})^{-1} \mathbf{X}_{1*m}^T \mathbf{y}_{m*1} \\ \mathbf{X} &:= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \end{aligned}$$

Because in zero order case, the matrix $\mathbf{X}$ is, in reality, a vector of m pieces of ones, the term $(\mathbf{X}_{1*m}^T \mathbf{X}_{m*1})^{-1}$ is the same as $\frac{1}{m}$. The term $\mathbf{X}_{1*m}^T \mathbf{y}_{m*1}$ can be also described as product between $\mathbf{X}$ and $\mathbf{y}_m$ which will lead to $\sum_m y_m$. Therefore c_0 can be also presented as Equation 8.

2.2.2 Baseline Correction

Baseline correction is a signal pre-processing technique, where background effects are removed from the measured signal [18]. In signal analysis, baseline correction is used to separate the real signal from the measurement. Measurements might be affected by background effects for many different reasons. One of the typical reasons in Raman spectroscopy is that the spectrometer captures also Raman shifts from the material where the sample is held. When the signal is background corrected, it can be expected to explain mostly the measured material and white noise.

The method of baseline correction implemented in this study is based on moving average smoothing [19]. The moving average smoothing is fitted iteratively to the spectrum 4.

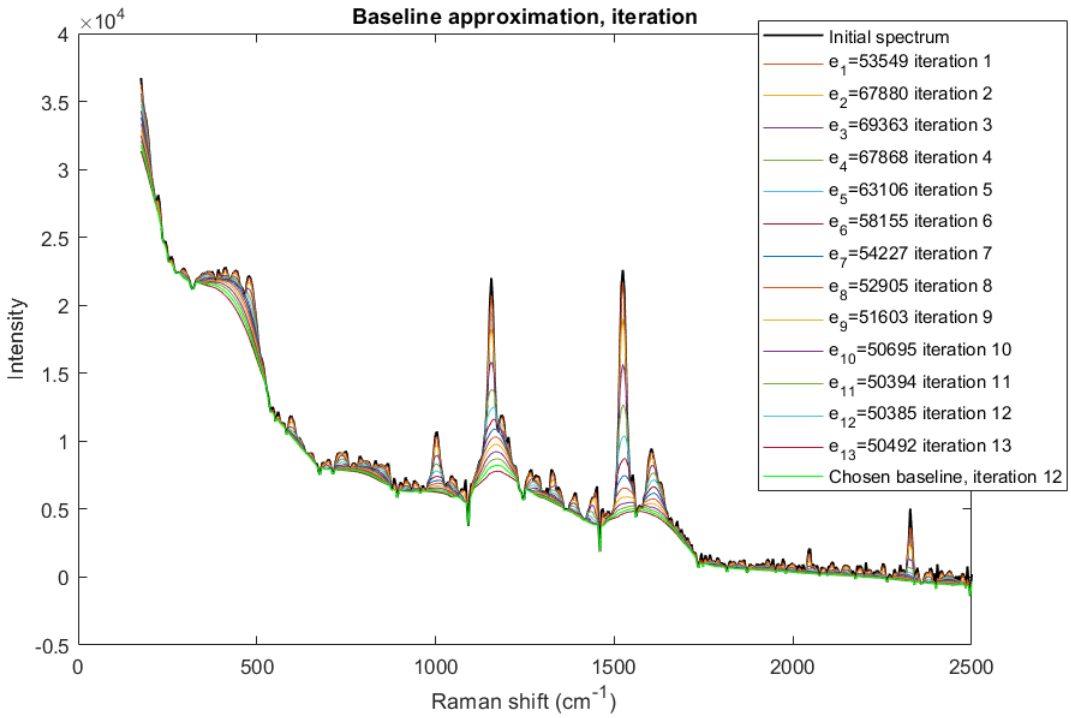

Figure 4. Iterative solution to finding the baseline

The length of subsets, which the moving average is calculated off, is increased every iteration. The new moving average smoothing is fitted into the previous baseline approximation, which is derived as minimum values of corresponding values between the initial spectrum and moving average smoothing. It can be done because the previous approximated baseline represents the shape of the gradually smoothing initial spectrum.

Fitting the moving average smoothing into the previous approximated baseline allows the length of the subset to remain lower and it to converge faster to the optimal baseline state. The optimal baseline represents the background of the initial spectrum. It is chosen by finding the minimal summed absolute difference $\mathbf{e}_i$ between two consecutive approximations of the baseline, $\mathbf{B}_{i-1}$ and $\mathbf{B}_i$.

$$e_i = \sum |\mathbf{B}_{i-1} - \mathbf{B}_i| \tag{9}$$

2.2.3 Normalization

Normalization is used to transform spectra into the same scale. It is needed when the spectrum is changing its shape or the peaks are shifting on the wavelength scale. With different normalizations, there is a possibility to observe different effects in the spectra. Normalizations can also be applied to smooth out features from between different measurements. In spectroscopy, one common feature is laser power, which is sometimes needed to be changed between measurements to be able to observe the best spectra possible.

A very common normalization method in spectroscopy is area normalization, the integration of the spectrum is fixed to be 1 [20]. In a continuous case, it would mean that the spectrum would need to be divided by its integral, and in a continuous case divided by the sum of the rooted squares of the spikes. Another common method is to divide every spectrum of spectra with the same spike in terms of wavelength. The forming spectrum is then represented with respect to the spike in the wavelength.

Autoscaling is a normalization method, where the mean of the spectrum is first subtracted from the spectrum itself and then divided by its standard deviation [21]. Autoscaling scales all the spectra to zero means, and to have a standard deviation of 1. In spectra, the intensity differences depending on laser power transforms can be fixed to the same scale by autoscaling. It is still not a method that should be used too lightly, because by scaling the intensities at the same level, it can also distort the observations. Naturally, if the laser power needs to be reduced, the intensity of the observed Raman shifts is lower - even if lower laser power would not result in fewer photons to Raman scatter of the transmitted radiation. Instead, in the spectrum, the relative changes of the spikes in the same spectrum are highlighted after autoscaling and thus the relative changes in spikes are easier to detect.

2.2.4 Principal Component Analysis

PCA (Principal Component Analysis) is commonly used to reduce noise from samples by reducing dimension. It can be derived from SVD (Singular Value Decomposition). Principal components describe the direction of variance in a matrix. From SVD the best principal components are the ones, which have the largest singular values and they describe the original matrix the best.

To find the principal components, the data matrix $\mathbf{X}$ must be first centered on the origin by

subtracting the mean from the data. If the spectra are in data matrix column space and the Raman shift intensity observations are in data matrix row space, the data matrix should be transposed. From the data matrix that has its spectra in row space and the Raman shift intensity observations in column space, the mean should be taken and subtracted from every single column to center the observed intensities of each wavelength.

$$\bar{\mathbf{X}} = \mathbf{X} - \mu_X \tag{10}$$

The centered data $\bar{\mathbf{X}}$ can be decomposed for matrices **U**, **S**, and V^T like every R^{n*m} matrix.

$$\bar{\mathbf{X}}_{n*m} = \mathbf{U}_{n*n}\mathbf{S}_{n*m}\mathbf{V}^T_{m*m} \tag{11}$$

The column vectors of the matrix **V** are row space singular vectors of centered data matrix $\bar{\mathbf{X}}$, which are orthonormal. The best singular vector with the biggest singular value describes the matrix $\bar{\mathbf{X}}$ row space the best. Because the spectra are also at row space, column vectors of **V** are the vectors that explain the row space of $\bar{\mathbf{X}}$ the best. Thus, column vectors of V are loadings of principal components of matrix $\bar{\mathbf{X}}$.

$$\mathbf{P}_{m*m} = \mathbf{V}_{m*m} \tag{12}$$

The matrix **U** column vectors are the column space singular vectors of matrix $\bar{\mathbf{X}}$. They are also orthonormal, and the singular vector with the biggest singular value describes $\bar{\mathbf{X}}$ column space the best. The matrix **S** has the singular values of matrix $\bar{\mathbf{X}}$ on its diagonal and all other components of **S** are zeros. The singular values are the same as square-rooted eigenvalues of matrix $\bar{\mathbf{X}}$. R^2 value of a principal component k can be formed as:

$$R^2_k = \frac{\mathbf{S}_{k,k}}{\Sigma S} \tag{13}$$

The linear combination of matrices **U** and **S** is referred to as scores **T**. In the matrix, **T** column space is the centered spectra $\bar{\mathbf{X}}$ coordinates in a space formed from principal components.

$$\mathbf{T} = \mathbf{US} \tag{14}$$

For applications only the k principal components that describe the centered data the most are wanted. Those are found by looking at singular values or R^2 values. It is considered that only the biggest singular values are related to the real phenomenon and those which are converging to zero are only related to the noise. To get the best k principal components, only the first k columns are taken from matrix $\mathbf{V}$.

$$\mathbf{P}_{m*k} = \mathbf{V}_{m*k} \tag{15}$$

The score of the best k principal components can be formed by taking only the first k columns of coordinates of projected space.

$$\mathbf{T}_{n*k} = \mathbf{U}_{n*k}\mathbf{S}_{k*k} \tag{16}$$

SVD is a common technique for computing PCA, along with eigenvalue decomposition. The data matrix $\mathbf{X}$ is represented in PCA space as in Equation. 17, where $\mathbf{T}$ is the score matrix, $\mathbf{P}$ is the loadings matrix and $\mathbf{E}$ the model residual.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \tag{17}$$

Thus, the noise-reduced data could be formed as:

$$\hat{\mathbf{X}}_{n*m} = \mathbf{T}_{n*k}\mathbf{P}^T_{k*n} \tag{18}$$

3 EXPERIMENTAL METHODS

3.1 Introduction of Instrument

In the experimental part, the used spectrometer is BWTEK's TacticID-1064 Handheld Raman System Figure 5. It uses a 1064 nm wavelength which reduces fluorescence [11]. The spectrometer was mostly used with an angular lens and the spectra were analyzed using MATLAB.

Figure 5. The handheld Raman spectrometer

The plant chosen for the experimental part is Basil, *Ocimum basilicum*. Basil was chosen based on its features [9] and its great supply even in winter in Finland. For its various features, research from basil interests many in culinary [13, 14] and agriculture [7, 8] which was one of the motivations for choosing Basil to plant pot experimental part. The measured plant, *Ocimum basilicum*, was acquired from a Finnish grocery store in a pot.

3.2 Instrumental Set-up Testing

The objective of the instrumental setup testing was to learn to optimize the measurement-taking process and which kinds of spectra are obtained from the spectrometer before any measurements for actual research will be taken. The spectra were taken with different

spectrometer lenses, background bases, and setup adjustments to gain knowledge of how the best spectra were detected with the device.

The spectrometer lenses available were a point-and-shoot lens and a right-angle lens. The background bases represent the stand that the leaves are placed on. The materials tested for bases were glass and aluminum. The top layer is a material placed between the sample and the lens. For the top layer, both glasses with different thicknesses and no cover scenarios were tested. The last parameter of the experimental setup tested was the laser power, which varied from 84 to 378 mW.

3.3 Heterogenity Testing

The intention of heterogeneity measurements was to find out if separate leaves of the same healthy plant at about the same age are alike and if separate visually healthy plants are similar. For heterogeneity testing, the samples were taken from two different plants, in two experimental set-ups: plants in pots and plants in water. From the plants, leaves of different ages were ripped and placed on a glass plate for taking the spectrum. Baby leaves were not used because the spectrometer was much easier to be pointed to older bigger leaves.

The heterogeneity was tested by calibrating a PCA model on the samples in the two plants. If there are clear groups either between the plants or between acquisition areas of the plant, the plant is considered heterogeneous, and more careful sampling is taken into account for the pollutant plants measurements. If there are no visible clusters formed between the two plants or the measurement areas on the leaf, the plant is considered homogeneous and the samples from the polluted plants can be taken from any plant area. Also, the number of samplings per experiment is decided based on heterogeneity testing. If the plant measurements are homogeneous, every sample is representative of the whole plant, thus fewer measurements are needed.

3.4 Polluted Plants Measurements

In pollutant measurements, measuring techniques from past sampling were applied. In total 30 spectra from four healthy plants in pots and 10 spectra from leaves of Basil branches in water were taken to form a control sample of 40 spectra. From the plants, old

and young leaves were ripped and used in the process of taking spectra. For pollutants, in total four samples with four branches were made. Each sample had 200 ml water and dissolved salts; 0.5g of $FeSO_4$, $CuSO_4$, $CoSO_4$ or $ZnSO_4$ Table 1.

Table 1. Pollutants and concentrations

Sample	Pollutant	m_{salt} [g]	C_{M+} [mol/L]
1.	$FeSO_4$ $1.5H_20$	0.5	0.014
2.	$CuSO_4$	0.5	0.02
3.	$CoSO_4$	0.5	0.016
4.	$ZnSO_4$ $7H_20$	0.5	0.009

To understand the changes based on the duration of the pollutant effect on Basil, the measurement of polluted Basil were taken 24 hours and 48 hours after exposing the branches to the pollutants. The changes in the samples were observed visually Appendix 1 and thus the Raman spectrometer. The samples were held under the same environment the whole observation time. The samples were on the windowsill in their own solutions in measuring glasses.

For each pollutant measurement, two of the branches were used completely, excluding the very dry leaves. Ten successful spectra were tried to be taken with each pollutant on both occasions to guarantee enough data points. Also, notes of visual changes between individual samples and between the sample and a healthy plant were taken.

To measure the effects of the plants, two approaches were taken. Firstly, a PCA model is calibrated on the control samples, and the polluted samples are projected into the model. Tools such as score plots evaluate if there is any clustering of polluted samples. Also, clustering between different exposure times is desired. Secondly, individual PCA models are calibrated for each experiment. The PCA loadings are compared for the PC explaining most variation, and differences between the main variance profile are evaluated before and after exposure to the pollutant.

4 RESULTS & DISCUSSION

The experimental part consists of heterogeneity measurements and pollutant monitoring measurements. Nearly 300 spectra were taken from the Basil plants overall. The changes in plants were monitored with a handheld Raman spectrometer and observed visually to understand if the effects of the pollutants could be assumed to be observed from the spectra based on the physical changes in the plants.

4.1 Measurement Technique

Different setups were tested to determine which setup would give the spectrum with the best spectra-to-noise ratio. Practice learned from setup testing Table 2 was to use a thick glass plate as a background base. Also, instead of single leaves, stashes of leaves seemed to perform better. Single leaves might have been too thin, and radiation penetrated the leaf completely and thus forming no spectrum.

Also, spectra were much worse when single or multiple 1 mm thin glass plates were used on top of the sample Figure 6 Scenario 2. The problem with the glass plates on top was that the fingerprint of the glass was reflected in the area of 500 to 1200 Raman shifts when important spikes of Basil were detected at the area of 800 to 1250 Raman shits. Thus the fingerprint of glass would have been distorting the information of spectra of Basil.

The laser power was chosen to be 378 mW because it was found to give the best signal-to-noise ratio. It can be concluded from Figure 6 by comparing Scenario 7 where the laser power was 252 mW and Scenario 9 where it was 378 mW.

Table 2. Experimental and instrumental parameters tested

Scenario	Leaves		Lens		Base		Top			Laser power [mW]			Comment
	A ripped leaf	Multiple ripped leaves	Right angle adaptor	Point & shoot	Al base	Glass plate [3 mm]	Glass plate [3 mm]	Glass plate [1 mm]	No top	378	252	84	
1	x			x		x	x			x			No spectra was captured.
2	x			x		x		x		x			No spectra was captured.
3	x		x			x	x			x			Low-intensity spectra, with glass fingerprint.
4	x		x			x		x		x			Low-intensity spectra, with glass fingerprint.
5	x		x			x			x	x			Medium intensity with no interference.
6	x		x		x				x	x			Medium intensity with no interference.
7		x	x			x			x	x			High-intensity spectra with high signal-to-noise.
8		x	x			x			x			x	Low signal-to-noise.
9		x	x			x			x		x		Low signal-to-noise.

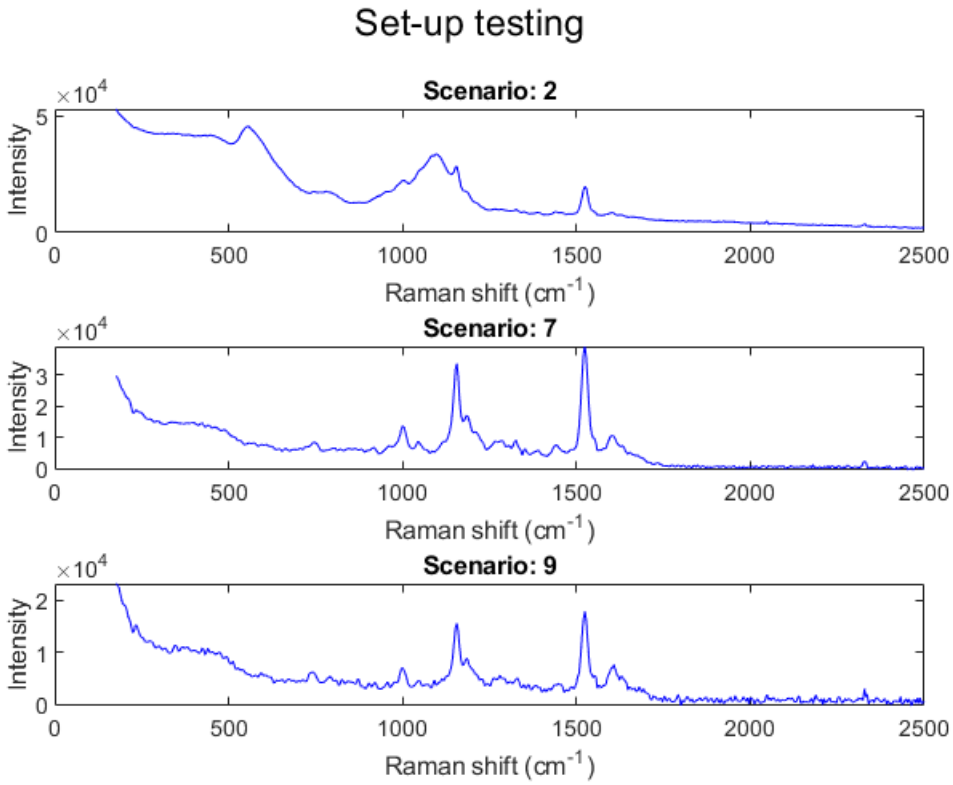

Figure 6. Different spectrum from set-up testing of scenarios in Table 2

The right-angle lens was the most workable of the lenses. It gave better spectra compared to point-and-shoot. One of the problems with the point-and-shoot lens might have been that it needs to be held stationary very close to the sample without touching it. The final measurement setup can be observed from Figure 7, where instead of one leaf, multiple leaves were stashed on top of each other.

Figure 7. Schematic of a measurement setup, instrument fitted with the right angle adapter.

4.2 Visual Observations

The physical appearance of Basil's branches changed 24 hours after the samples were exposed to the pollutants Appendix 2. Based on visual changes in the branches, Basil resisted $FeSO_4$ and $ZnSO_4$ the most and $CuSO_4$ the least. The phenomena were confirmed with PCA scatter plot in Chapter 4.5, as $FeSO_4$ and $ZnSO_4$ are closest to the control sample and $CuSO_4$ furthest away. The differences in observed intensities can be seen in Figure 10, which confirms the conclusion as $CuSO_4$ enfeeble the plant the most spectra-wise.

After 48 hours of exposing the branches to the pollutants, all the samples were withered badly. The sample of $ZnSO_4$ was the only one that visually looked alive Appendix A1.2. Measurements without a significant noise ratio were difficult to be taken. Especially the spectra of $CuSO_4$ were very noisy. The reason might be, that the sample was dried so badly, that in the process of taking spectrum, the laser burned the leaves the much that little sparks could have been seen, even though the power of the laser was reduced from 378 mW to 84 mW.

4.3 Pretreatment of Spectra

The baseline correction was applied to the spectra and outliers were removed. After baseline correction and outlier removal, the heterogeneity of the spectra was tested. After the treatment of the spectra Figure 8, the spectra were autoscaled spectral-wise, which amplified the noise, which was then removed by applying moving average smoothing. Autoscaling was applied to remove the uncertainty based on the change of laser power between healthy and polluted samples.

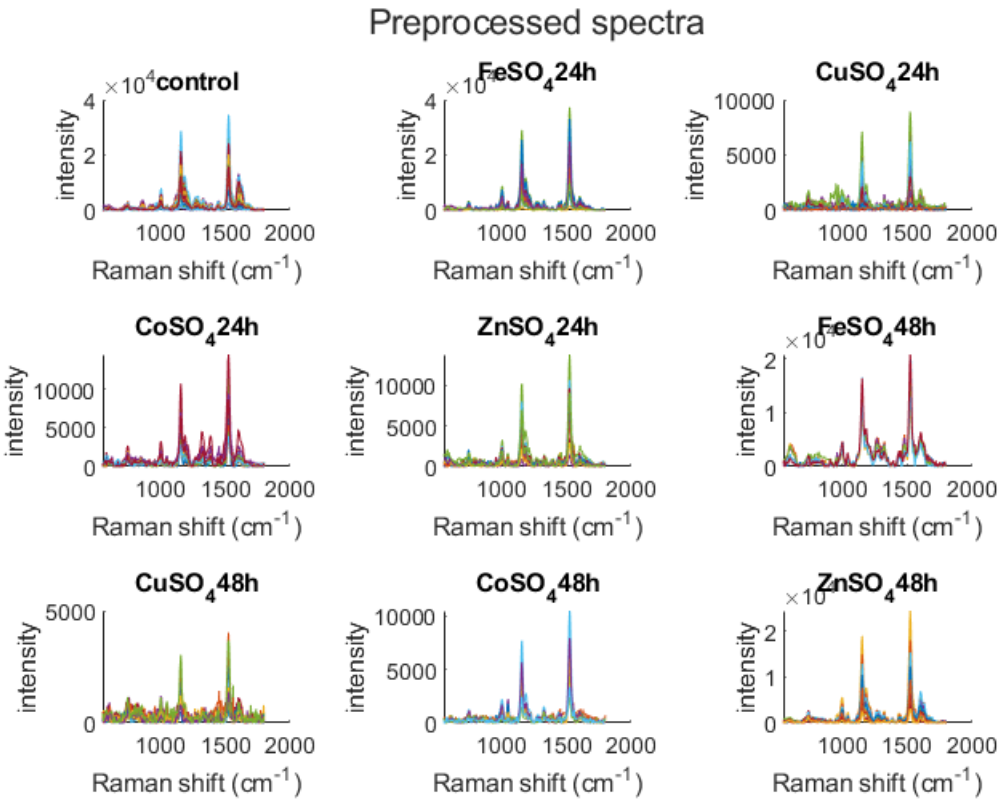

Figure 8. Preprocessed spectra

4.4 Heterogeneity Testing

To ensure that the untreated plant samples are heterogeneous, spectra from multiple plants were taken, having different ages and different acquisition spots. These *healthy* plant observations have been modeled with PCA. The analysis revealed that most of the variation is captured by the first principal component, which confirms the homogeneity of the measurements. There were no clear outliers in the observations, which confirms that any

acquisition spot on the plant is representative of the whole plant. As seen in Figure 9, for the biggest amount of variation, the different plants give similar spectra, with no clear outliers in the scatter plot. The ripped plant kept in the water glass for 24h distinguishes itself from the two plants in the pot, but the difference is not significant.

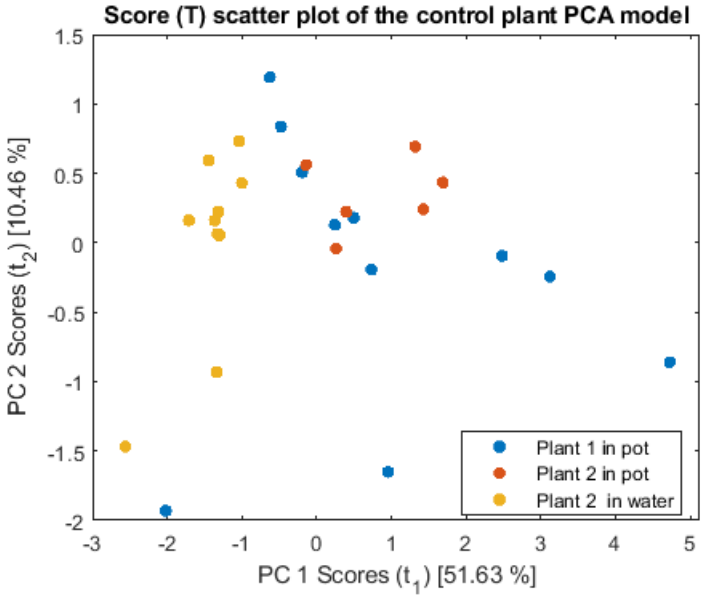

Figure 9. Score plot for the healthy sample model. The blue and red dots represent two different plants, with measurements taken from different parts of the leaves. The yellow group represents the control plant that has been ripped of, and placed in a glass with water for 24h.

4.5 Spectral Analysis

From Figure 10, an interesting finding is that after 48 hours of Basil being exposed to $FeSO_4$, some major changes have started to happen. It can also be seen from Figure A4.1, which regards better the change in laser power between 24-hour and 48-hour samples. Unlike other samples, the absorbed intensities of molecule vibrations of $FeSO_4$ are increased in some areas.

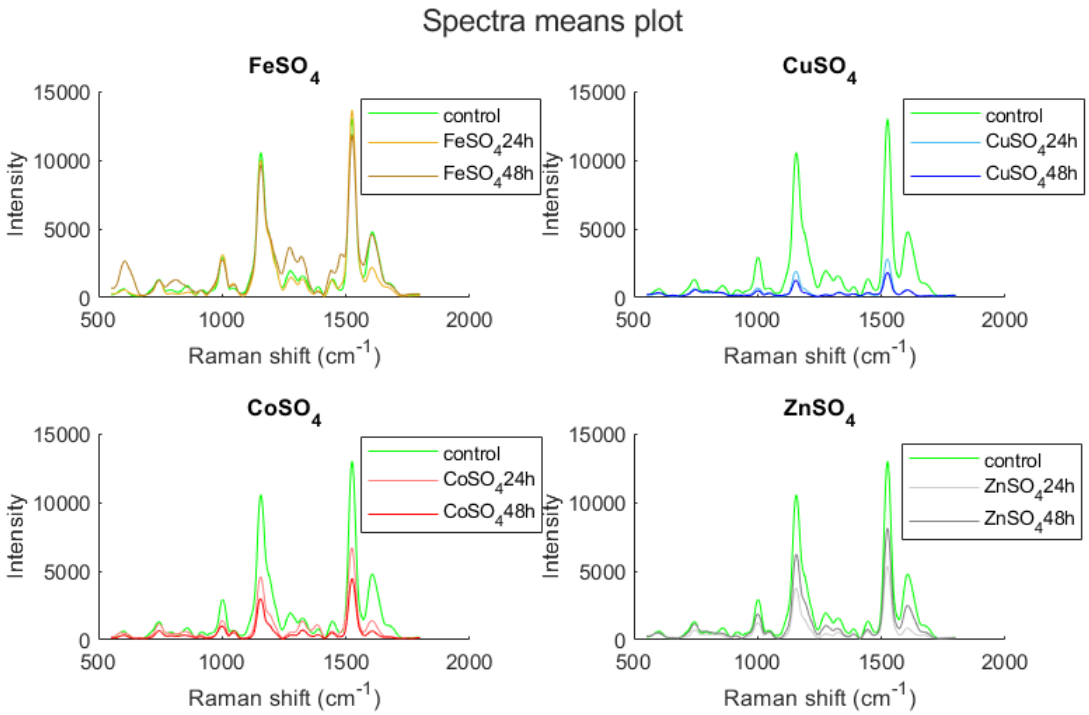

Figure 10. Comparison of Moving Average filtered spectra

In $FeSO_4$ 24-hour sample the spikes at 1524 Raman shifts have increased over the spike of the control sample Figure 10. After 48 hours the spike at 1524 Raman shift is decreased but spikes at 608, 1272, 1320, 1480, and 1604 Raman shifts can be clearly defined. The shoulder at 1480 cannot be observed in any other sample and it might be also the reason the spike at 1440 can be detected strongly. The shoulder could be a result of the decrease of the spike at 1524 Raman shifts because of the close location. The significant increase of spikes at 1272 and 1320 might be due to a double effect. Another shoulder can be detected at 1188 Raman shits Figure 11. The shoulder can not be detected from Figure 10, because the MA-smoothing filters it off. The shoulder keeps its form in every single pollutant except $CuSO_4$, where it disappears compared to the spike at 1004 Raman shifts.

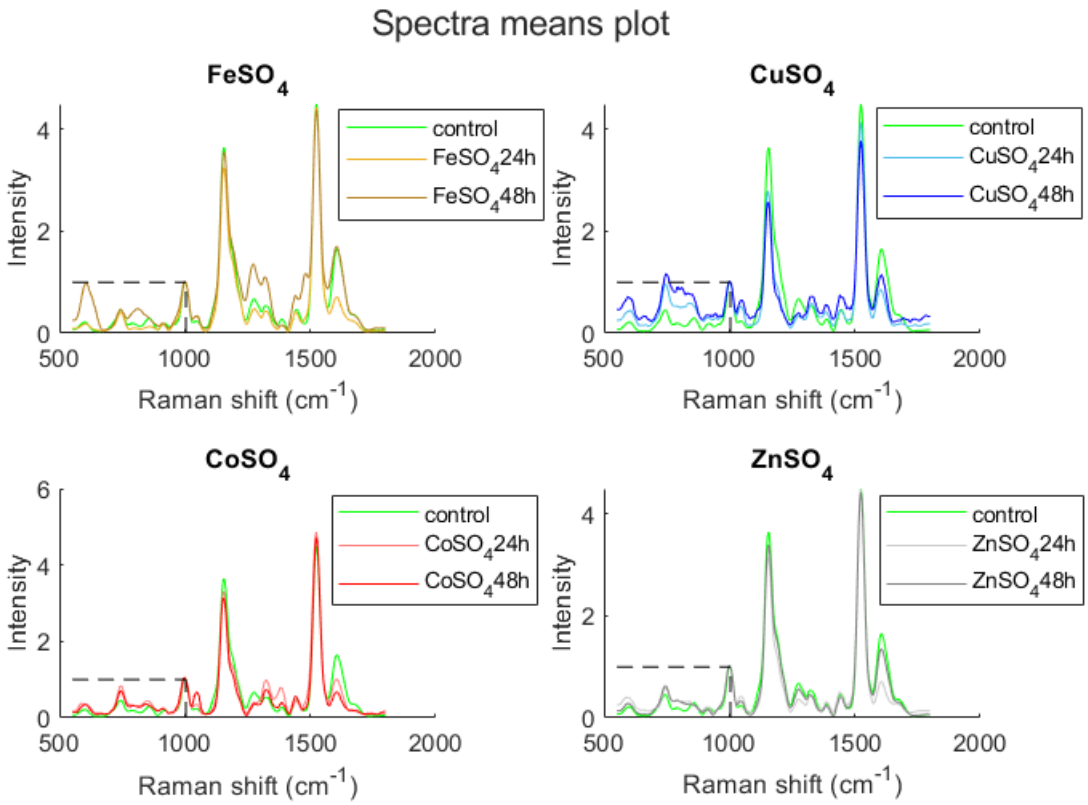

Figure 11. Comparison of mean spectra, which have been shift normalized by fixing the 1004 Raman shift to 1

Another notable change in Figure 10 is that the intensity of $ZnSO_4$ 48-hour spectra is higher than it is in the 24-hour sample. The most notable change can be found from the 1608 Raman shifts spike since it is relatively returned the most to the level of the control sample in Figure 10. The loadings of the $ZnSO_4$ represent the control sample the best in Figure 12 and in Figure 13, $ZnSO_4$ is closest to the control sample.

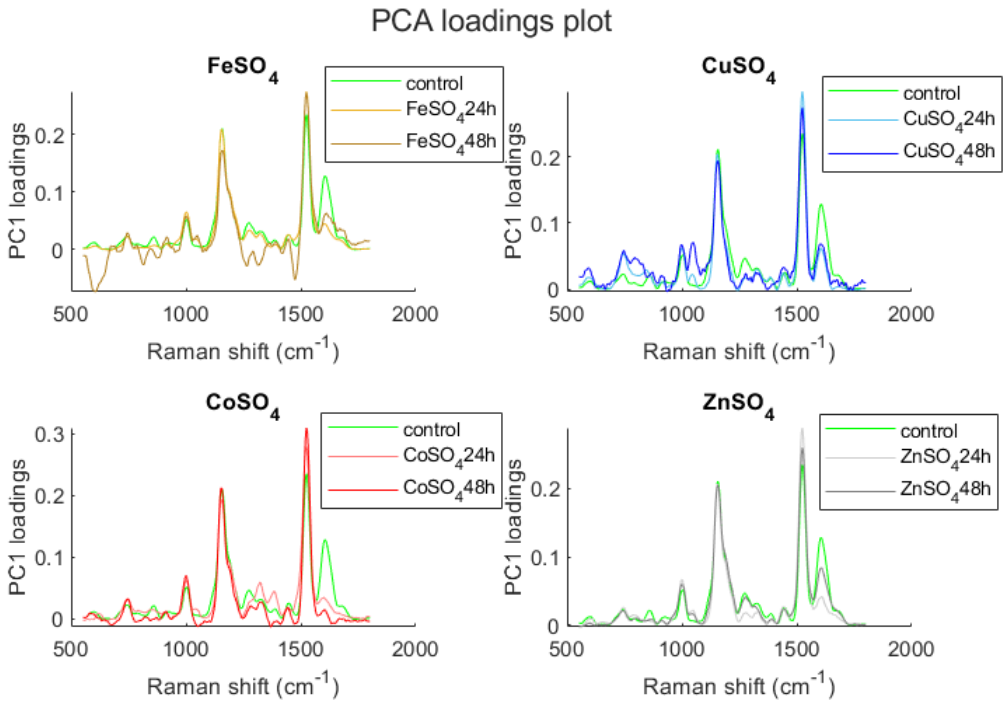

Figure 12. Loadings of individual PCA models for each experiment. Individual spectrum has been moving-mean-averaged with a window of 5.

The most notable changes from the healthy plant in terms of the first loadings can be seen from the $FeSO_4$ 48-hour sample in Figure 12. Between Raman shifts of 1472 and 1524, there is a clear negative correlation that cannot be seen in any other sample. The spike of 600 Raman shifts can also be clearly seen in the first loading. The last interesting observation is the change of correlation between $FeSO_4$ 24-hour and 48-hour samples from positive to negative between the spike of 1156 Raman shifts and the double spike at 1250 to 1350 area 12.

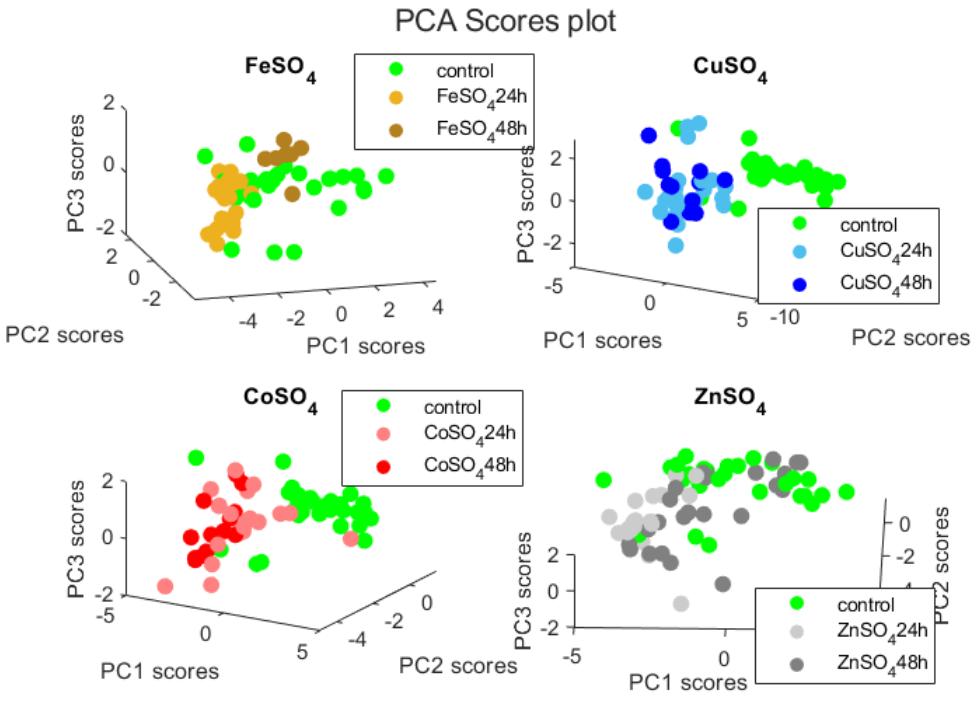

Figure 13. PCA score plots: visualization of changes after exposing plants for 24 h and 48 h.

Figure 13 is from a model that has been calibrated for the control samples and the polluted samples have been projected into it. The most similar to the control samples are the samples exposed to $ZnSO_4$. Also, it can be observed that the samples exposed to $CoSO_4$ and $CuSO_4$ differ clearly from control samples but altering time did not have a clear effect on grouping Figure 13. However, when a combined healthy-polluted PCA model has been calibrated Figure 14, the groups of $CuSO_4$ and $CoSO_4$ can be separated based on the altering time.

In Figure 14 the effect of pollutants can be observed. The further from the control sample the polluted sample has shifted the more effective the pollutant has been. Along the PC with the most variation, there can be a differentiation between both pollutants and exposure times. Generally, the more time a basil plant is exposed to the pollutant, the farther away it is in the score plot from the control sample group. Along the PC2, $CoSO_4$ and $FeSO_4$ have the highest variation. When considering both PCs, Zn seems to be most similar to the control samples.

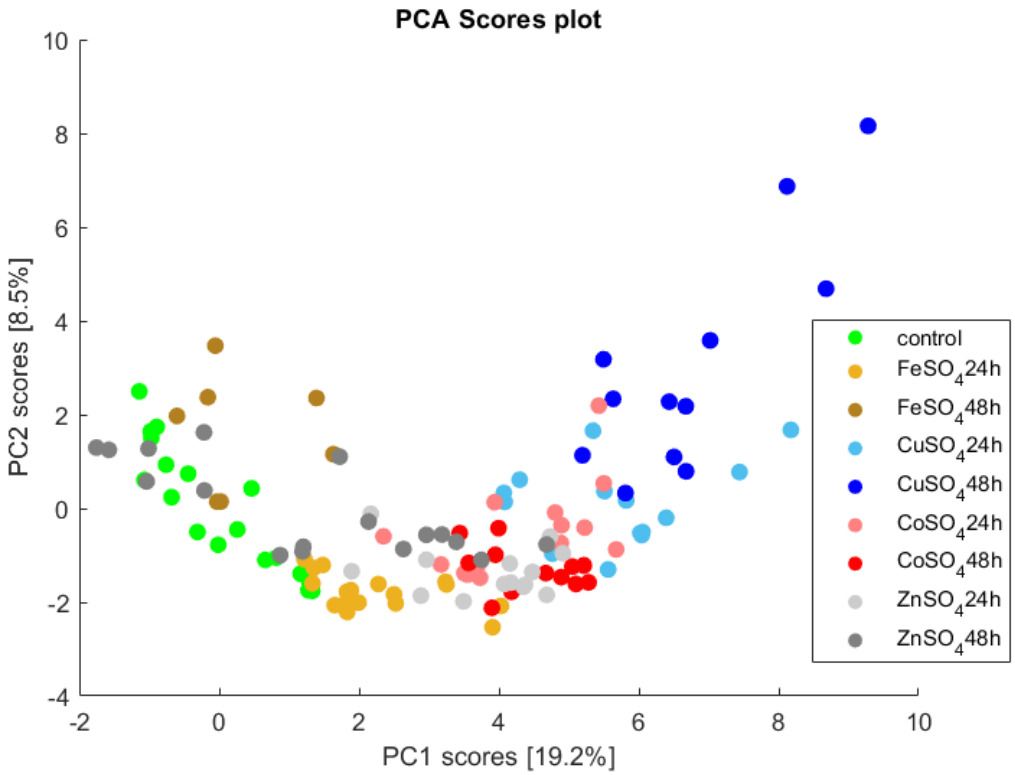

Figure 14. PCA score plots: visualization of changes of all plants

From Figure 14 it can be observed that the $ZnSO_4$ has had the least effect and $CuSO_4$ the most on plant health based on the shift. Also, the samples of $FeSO_4$ are near the healthy plant after 24 hours of exposing the plant to the pollutant. However, the effect of altering time shifts the sample of $FeSO_4$ towards the third principal component Appendix A3.1. $FeSO_4$ seems to be the only pollutant that affects the healthy plant in a way that can be seen from the third principal component in 48-hour exposure time.

Spectral-wise, from Figure 14, can be observed that PC1 has high loadings at around 1180 Raman shift Figure 15. $FeSO_4$ and the healthy samples have higher than average values for this peak, according to their placement on high PC1 scores. PC2 has high positive loadings at 1604 Raman shifts where one of the main peaks is located. The higher the $FeSO_4$ exposure, the higher the value in this peak Figure 15. Also, a strong variation included in PC2 is the negative correlation between the peak at 1520 and 1604. The elements that present this variation are mostly $CuSO_4$ and $FeSO_4$ polluted samples.

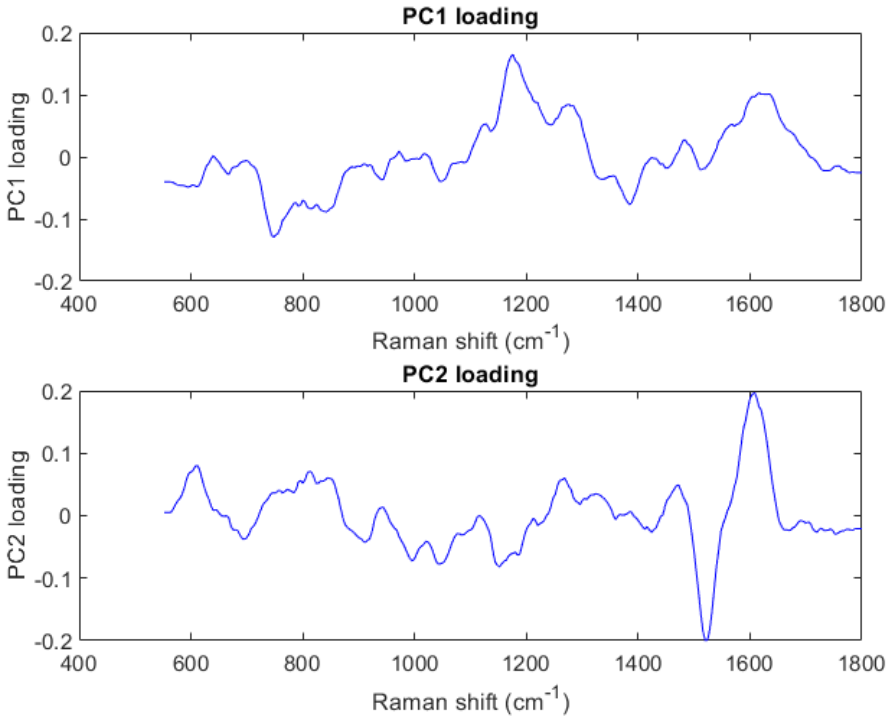

Figure 15. PCA loadings for the common healthy/polluted model, with corresponding PCA scores in Figure 14

5 CONCLUSIONS

A handheld Raman spectrometer can be used to monitor plants' health. A plant's health can be observed by comparing its spectra to the spectra of a healthy plant. The change can be seen from the change of variation in spectra, from the principal components scores. The shift in variation between polluted samples is not as clear as compared to a healthy plant. From pollutants, only $FeSO_4$ has a clear change in spectra when comparing the 24-hour to 48-hour measurement in all the different methods applied. The change in spectra can be observed by visualizing the score of principal components. Other pollutants are hard to separate clearly from each other but after autoscaling also they cluster nicely.

Different pollutants have different effects on plants' spectra. From tested pollutants, $CuSO_4$ seems to be the most effective by killing the plant completely. It can be seen based on visual observation of plant health and by comparing the spectra of the polluted plant to the spectra of the healthy plant. From the mean spectra, it can be seen that the received intensities of molecule vibrations of pollutants were lower than the intensity of healthy plants among all pollutants, except for $FeSO_4$. The spectra of the plant polluted with $FeSO_4$ stayed the most similar to the spectra of the healthy plant spectral-wise based on the mean spectra but based on PCA analysis, the plant was least affected by $ZnSO_4$ that was also supported by the visual observations. In conclusion, the evidence supports that the plant resisted $ZnSO_4$ the best as the branches in its solutions were the only ones standing after 48 hours after exposure.

To be able to analyze spectra, the spectra need to be pre-treated. The raw spectra are noisy and they may have background effects of different baseline trends. One of the methods that is used to remove the baseline is taking a spectrum of background material and removing it from the spectra. That technique cannot be used when taking spectra of organic material with a handheld spectrometer. The raw spectra have too many variations of background and it needs to be treated by defining the baseline for each spectrum individually. For that reason, fitting some iterative baseline correction method that changes depending on the spectrum is ideal.

Before and after baseline removal, the spectra will most likely be noisy and include outliers. The treatment of outliers is best to be handled before and after the baseline correction because it will most likely boost the noise of some already noisy spectrum. The outliers can be treated by excluding all the spectra whose intensity is higher or lower than some boundaries of defined Raman shifts. After excluding outliers it is good to smooth the spectra. The moving average smoothing has shown to be very effective in that.

For monitoring plant health, the multivariate analysis methods are the easiest for visualizing the information. From score plots, it can be clearly observed if the polluted plants are shifted from healthy plants and if time has an effect on the change.

6 FUTURE RESEARCH

There is no doubt that the statistical approach to remove the baseline, which is used in this study, would not be better than removing the baseline based on some variation of the measured background spectrum and visual observation of the best fit [19]. The baselines that are removed from the spectra in this study are overfitting from 1150 Raman shifts to 1750 Raman shifts Figure 16. For that area, a better fit would be even a straight line. By overfitting the baseline, the problem is that some of the information is either lost or distorted. Further research would involve optimization of the cost function Equation 9.

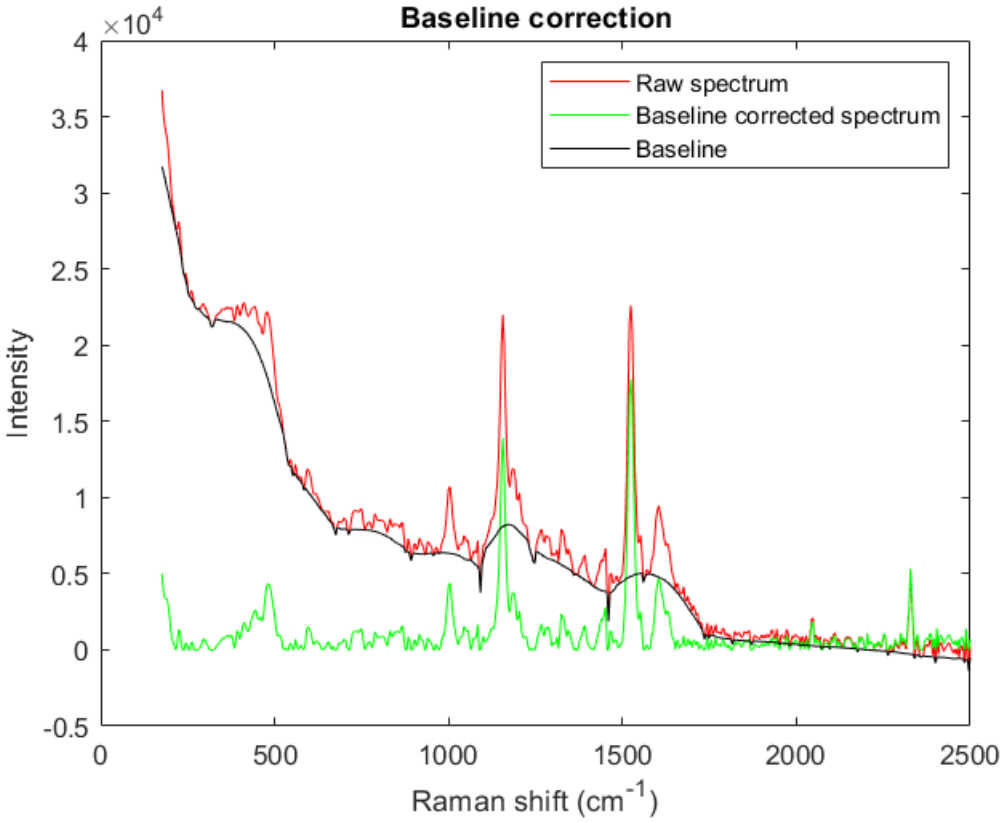

Figure 16. Baseline correction of the raw spectrum

The spectra of healthy Basil from this research do not match with the spectra in the literature Figure 17. The combined spectrum of Figure 17 was made by summing the characterized spectra of Chapter 2.1.3 and dividing it by four as the spectrum consists of four different researches. Both of the spectra were scaled by dividing the spectrum by its own maximum value. The comparison is not easy since all the reference research spectra are from Basil oil and thus most likely have a different composition. An example of future re-

search to avoid the limitations is a parallel analysis of leaves and basil oil and the creation of reference spectra for both forms.

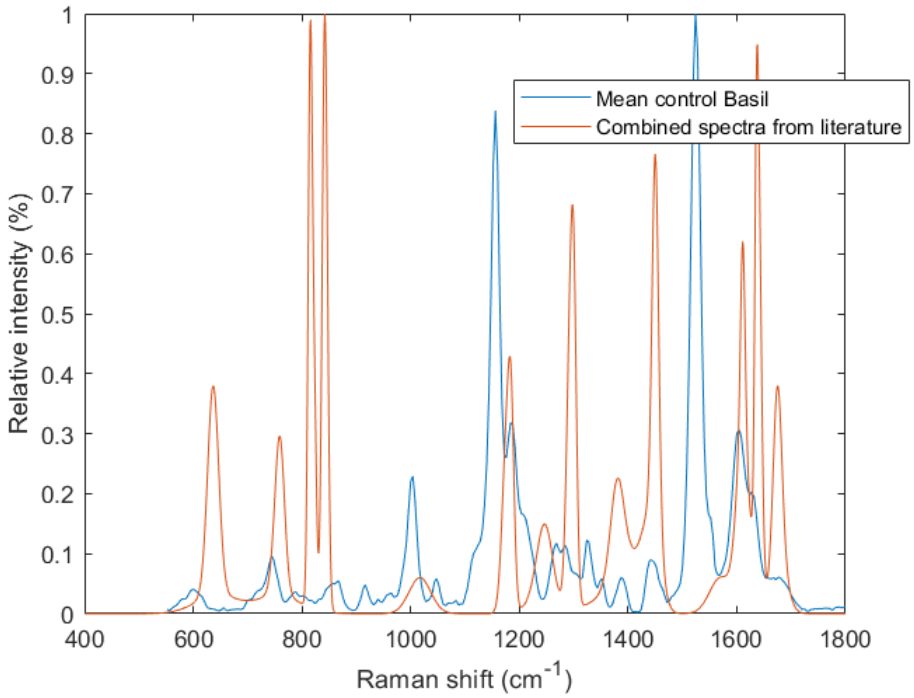

Figure 17. Visulization of combined characterization of reference spectra compared to the mean spectrum of this research

Even though there are differences in the spectrum there are some similarities also. At 1550 to 1700 Raman shifts zone the spectra do not necessarily correspond to each other but from the mean control spectrum of this research, there can be similar shapes found that the reference spectrum has. The only real spike at 1604 Raman shifts is very close to the spike in 1611 Raman shifts of reference spectrum and close to where the reference spectrum has a spike at 1637 Raman shifts the mean spectrum of this research has a shoulder at 1629 Raman shifts Figure 17. The spike at 1188 of the mean spectrum and the spike at 1181 in the reference spectrum have the closest fit overall. From the raw spectrum, it can be discovered that also in the spectra of this research, there is some variation between the top points of spikes. The variation seems to be four Raman shifts in both directions.

The reasons for difference in the spectra may be caused by various factors. Hussain et. al. suggests that Basil can be a little bit different depending on the time of the season and geographical location [16]. Also, in the genus of *Ocimum basilicum*, there are over a hundred

different species that have their own variation in composition [9, 16]. Based on that Basil in this research has been grown in Finland in a greenhouse and the reference researches on the other hand are made in Brazil, India, and Pakistan and are measuring the oil of Basil, it is safe to say that all the different factors, different seasonal and geographical conditions and different species have applied in making the spectra different.

The factors represented may be enough to explain the differences between the spectra. The spectra of this research are close enough to the reference spectra that the differences could be explained as environmental differences in plant proteins. The change in plant protein would be a valid explanation to change in Raman shifts within under 100 Raman shifts because different variations of the plant proteins cannot change the region of the functional group [22]. Future research may involve basil plants gathered from different geological locations and originating from different seasons to be able to understand the possible differences in spectra.

REFERENCES

[1] David L Elliott. *Ultraviolet Laser Technology and applications*. Academic Press, San Diego, CA, June 2014.

[2] Ian R Lewis and Howell Edwards. *Handbook of Raman spectroscopy*. Practical Spectroscopy. CRC Press, Boca Raton, FL, August 2001.

[3] Serge Savary, Andrea Ficke, Jean-Noël Aubertot, and Clayton Hollier. Crop losses due to diseases and their implications for global food production losses and food security. *Food Security*, 4(4):519–537, July 2012.

[4] Shilpi Gupta, Chung Hao Huang, Gajendra Pratap Singh, Bong Soo Park, Nam-Hai Chua, and Rajeev J. Ram. Portable raman leaf-clip sensor for rapid detection of plant stress. *Scientific Reports*, 10(1), November 2020.

[5] Samir Suweis, Joel A. Carr, Amos Maritan, Andrea Rinaldo, and Paolo D'Odorico. Resilience and reactivity of global food security. *Proceedings of the National Academy of Sciences*, 112(22):6902–6907, May 2015.

[6] Ultan Mc Carthy, Ismail Uysal, Ricardo Badia-Melis, Samuel Mercier, Colm O'Donnell, and Anastasia Ktenioudaki. Global food security – issues, challenges and technological solutions. *Trends in Food Science &amp Technology*, 77:11–20, July 2018.

[7] Muhammad Asif Hanif, Haq Nawaz, Muhammad Adnan Ayub, Nayla Tabassum, Nazish Kanwal, Nosheen Rashid, Muhammad Saleem, and Mushtaq Ahmad. Evaluation of the effects of zinc on the chemical composition and biological activity of basil essential oil by using raman spectroscopy. *Industrial Crops and Products*, 96:91–101, February 2017.

[8] Haq Nawaz, Muhammad Asif Hanif, Muhammad Adnan Ayub, Faiqa Ishtiaq, Nazish Kanwal, Nosheen Rashid, Muhammad Saleem, and Mushtaq Ahmad. Raman spectroscopy for the evaluation of the effects of different concentrations of copper on the chemical composition and biological activity of basil essential oil. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 185:130–138, October 2017.

[9] Balakrishnan Purushothaman, Ramalingam Prasanna Srinivasan, Purushothaman Suganthi, Balu Ranganathan, Jolius Gimbun, Kumaran Shanmugam1, , , , , , , and and. A comprehensive review on iocimum basilicum/i. *Journal of Natural Remedies*, 18(3):71–85, December 2018.

[10] Paul Rostron, Safa Gaber, and Dina Gaber. Raman spectroscopy, review. *laser*, 21:24, 2016.

[11] Shehara Gunawardana, Chinthika Gunasekara, and NMS Sirimuthu. Raman spectroscopy in phytochemical analysis. *Sri Lankan Journal of Applied Sciences*, 1(01):01–10, 2022.

[12] D Ambrose. CHROMATOGRAPHY. In *The Characterization of Chemical Purity*, pages 81–103. Elsevier, 1971.

[13] Archasvi Tyagi, Anil K. Yadav, Akanksha Yadav, Lalita Saini, Vivek Kumar, Pooja Jain, Inam Mohammad, Mohammad Javed Ansari, Hesham Ali El Enshasy, Fagr Kh. Abdel-Gawad, Sami Al Obaid, Shahida Anusha Siddiqui, and Vijai Malik. Vibrational spectroscopic methods for the identification and distinction of essential oils in genus ocimum l.: A chemometric approach. *Journal of King Saud University - Science*, 34(8):102355, November 2022.

[14] Juliana H. Miyoshi, Juliana C. Castro, Vanderson C. Fenelon, Francielle P. Garcia, Celso V. Nakamura, Ana C. Nogueira, Tania Ueda-Nakamura, Hâmara M. de Souza, Camila S. Mangolim, Gislaine F. Moura-Costa, and Graciette Matioli. Essential oil characterization of ocimum basilicum and syzygium aromaticum free and complexed with -cyclodextrin. determination of its antioxidant, antimicrobial, and antitumoral activities. *Journal of Inclusion Phenomena and Macrocyclic Chemistry*, 102(1-2):117–132, September 2021.

[15] R. Holghoomi, S. Hosseini Sarghein, J. Khara, and B. Hosseini. Effect of functionalized-carbon nanotube on growth indices in ocimum basilicum l. grown in vitro. *Russian Journal of Plant Physiology*, 68(5):958–972, August 2021.

[16] Abdullah Ijaz Hussain, Farooq Anwar, Syed Tufail Hussain Sherazi, and Roman Przybylski. Chemical composition, antioxidant and antimicrobial activities of basil (ocimum basilicum) essential oils depends on seasonal variations. *Food Chemistry*, 108(3):986–995, 2008.

[17] Michael Schmid, David Rath, and Ulrike Diebold. Why and how savitzky–golay filters should be replaced. *ACS Measurement Science Au*, 2(2):185–196, February 2022.

[18] Kristian Hovde Liland, Elling-Olav Rukke, Elisabeth Fjærvoll Olsen, and Tomas Isaksson. Customized baseline correction. *Chemometrics and Intelligent Laboratory Systems*, 109(1):51–56, November 2011.

[19] Kristian Hovde Liland, Trygve Almøy, and Bjørn-Helge Mevik. Optimal choice of baseline correction for multivariate calibration of spectra. *Applied Spectroscopy*, 64(9):1007–1016, September 2010.

[20] Kannan Ramamurthy, Karunakaran Ponnusamy, and Selvaraju Chellappan. Excitation-resolved area-normalized emission spectroscopy: a rapid and simple steady-state technique for the analysis of heterogeneous fluorescence. *RSC Advances*, 10(2):998–1006, 2020.

[21] Helen M Parsons, Christian Ludwig, Ulrich L Günther, and Mark R Viant. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics*, 8(1), July 2007.

[22] Giuseppe Pezzotti, Wenliang Zhu, Haruna Chikaguchi, Elia Marin, Francesco Boschetto, Takehiro Masumura, Yo-Ichiro Sato, and Tetsuya Nakazaki. Raman molecular fingerprints of rice nutritional quality and the concept of raman barcode. *Frontiers in Nutrition*, 8, June 2021.

Appendix 1. Plant Images

Figure A1.1. Visual observation of pollutants 24 hours after exposing the healthy plant to the pollutants: in the blue solution there is $CuSO_4$ as a pollutant, in the yellow solution there is $FeSO_4$ as a pollutant, in the blue solution there is $CoSO_4$ as a pollutant, in the blue solution there is $ZnSO_4$ as a pollutant

Figure A1.2. Visual observation of pollutants 48 hours after exposing the healthy plant to the pollutants: in the blue solution there is $CuSO_4$ as a pollutant, in the yellow solution there is $FeSO_4$ as a pollutant, in the blue solution there is $CoSO_4$ as a pollutant, in the blue solution there is $ZnSO_4$ as a pollutant

Appendix 2. Pollutants Visual Observations

Figure A2.1. Visual observations of healthy Basil

Appendix 2. (continued)

Figure A2.2. Visual observations of $FeSO_4$ effect to Basil: Above 24-hour sample, below 48-hour sample

Appendix 2. (continued)

Figure A2.3. Visual observations of $CuSO_4$ effect to Basil: Above 24-hour sample, below 48-hour sample

Appendix 2. (continued)

Figure A2.4. Visual observations of $CoSO_4$ effect to Basil: Above 24-hour sample, below 48-hour sample

Appendix 2. (continued)

Figure A2.5. Visual observations of $ZnSO_4$ effect to Basil: Above 24-hour sample, below 48-hour sample

Appendix 3. PCA Scores Scatterplot

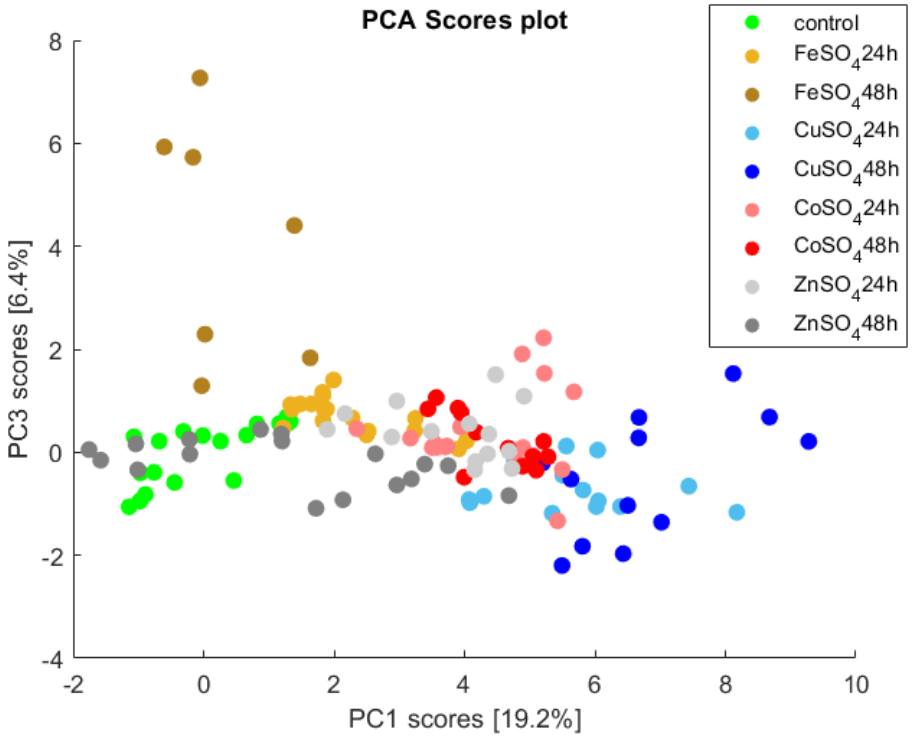

Figure A3.1. PC1 vs. PC3 score scatter plot, for the combined healthy-polluted PCA model in Figure 14

Appendix 3. (continued)

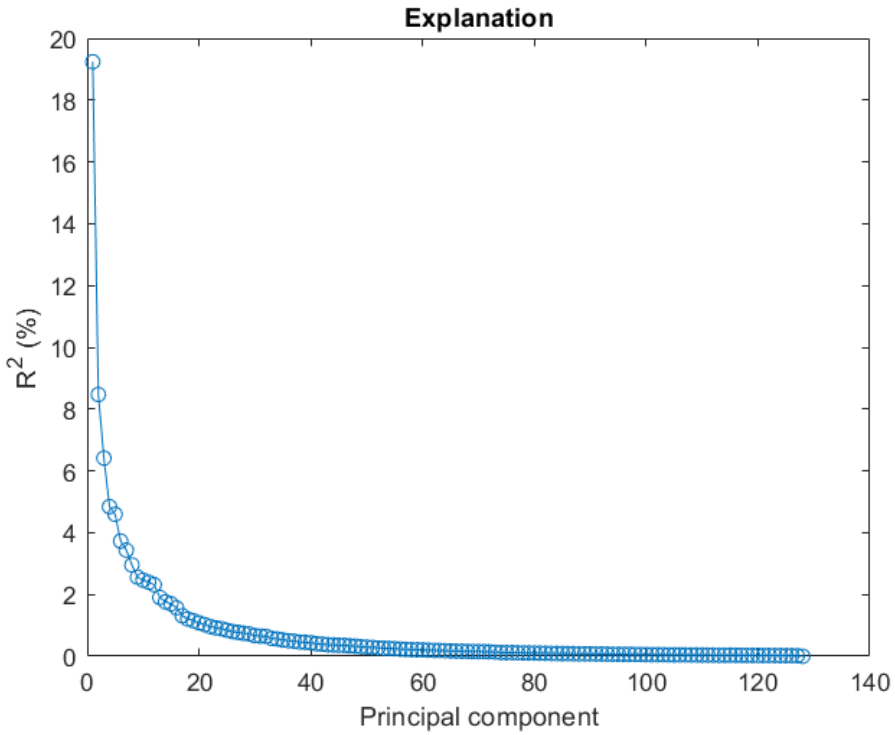

Figure A3.2. Explained variance for PCs in the combined healthy-polluted PCA model in Figure 14

Appendix 3. (continued)

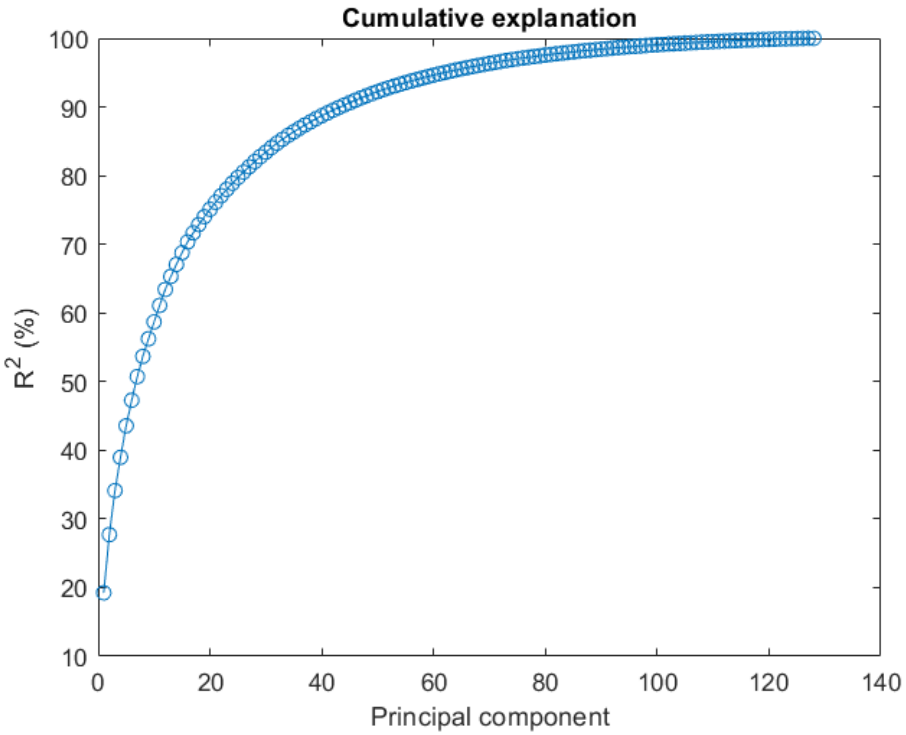

Figure A3.3. Cumulative explained variance for PCs in the combined healthy-polluted PCA model in Figure 14

Appendix 4. Mean Plot

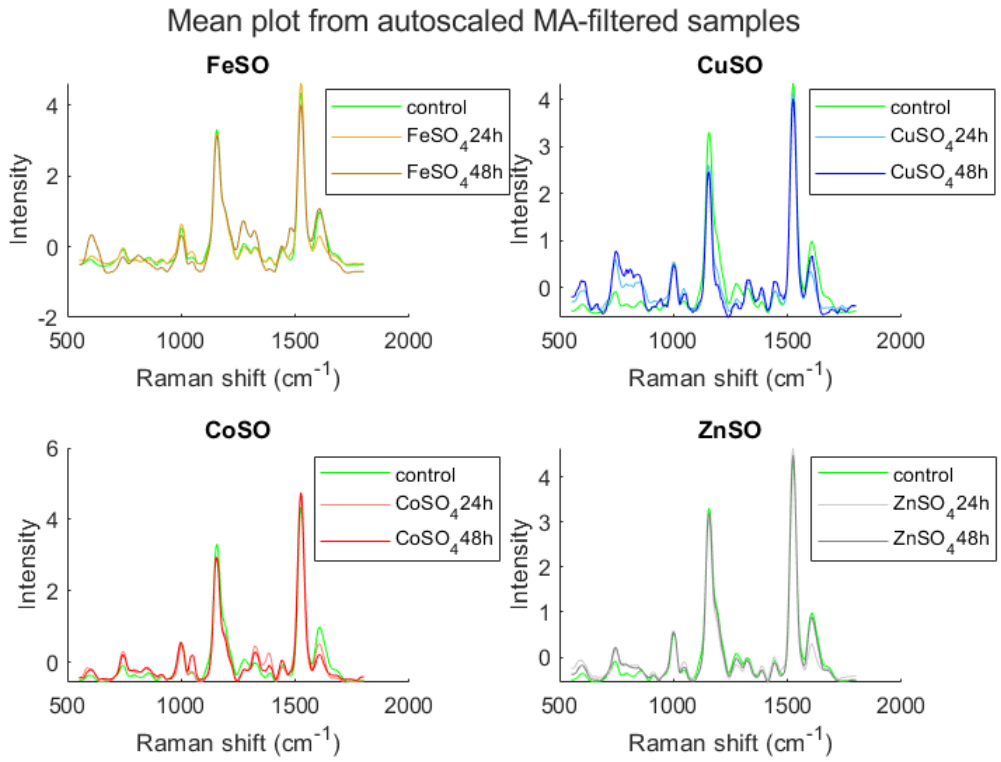

Figure A4.1. Mean spectra from 10 measurements of each experiment. Individual spectrum has been moving-mean-averaged with a window of 5.

Appendix 5. Raw Spectra

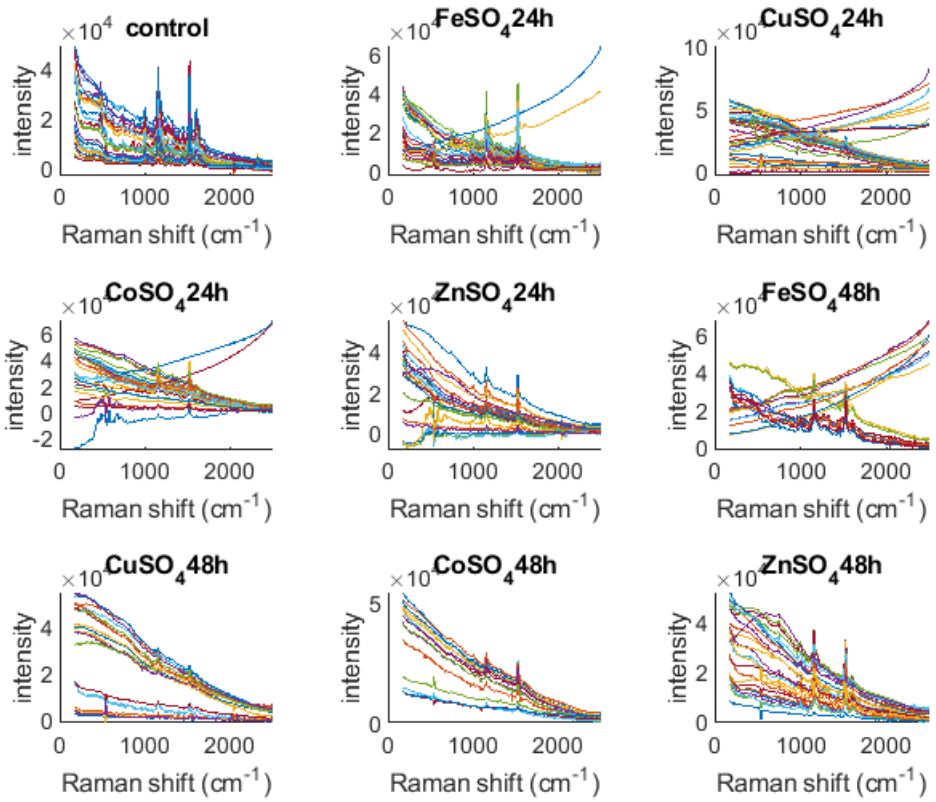

Figure A5.1. Raw spectra