Quantifying the Amount of Lithium-Bearing Minerals in Fine Matrix

An FTIR approach based on mineral powder mixtures

T.P van der Ven

Delft University of Technology

TU Delft

Quantifying the Amount of Lithium-Bearing Minerals in Fine Matrix

An FTIR approach based on mineral powder mixtures

by

T.P van der Ven

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Tuesday October 31, 2023 at 15:00 PM.

Student number:	5315816	
Project duration:	September 4, 2023 – October 31, 2023	
Thesis committee:	Dr. T. Schmiedel	TU Delft, Supervisor
	Dr. O. M. Kamps	TU Delft, Supervisor
	Dr. K. H. A. A. Wolf	TU Delft

TU Delft

Abstract

Lithium is a highly sought-after resource currently. Among other things, its use in lithium-ion batteries used extensively in electric vehicles, mobile phones and other electronics keeps the demand high. The demand for lithium is even projected to double in the coming years. It is safe to say that lithium will continue to remain an important resource for the future. Because of this increasing demand, the supply of lithium has grow. This supply can increase by expanding current mines, or opening new ones. Lithium can be found in lithium-caesium-tantalum pegmatites, a rare subcategory of granitic pegmatites. In these pegmatites, the lithium-containing minerals spodumene, petalite and lepidolite are most commonly mined. These minerals can be found in the form of larger crystals, but in recent years more attention has come to lithologies where quartz and lithium-containing minerals have been found growing side-by-side in smaller crystals.

This thesis aims to propose two models to quantify the amount of lithium-bearing minerals in fine matrix. These models are created by using Fourier-Transform Infrared Spectroscopy (FTIR) data of powder mixtures consisting of petalite and quartz for the first model, and spodumene and quartz for the second model. This approach using powders allows for simple and fast data acquisition which aids model development. The models considered in this thesis are linear regression, ridge regression and support vector regression. Linear regression produces the best results. Both models have shown to accurately predict the respective weight percentages of petalite and spodumene in powder mixtures with quartz. The spodumene model, when applied to a rock which contains both spodumene and quartz in fine matrix, produces similar results compared to Laser-Induced Breakdown Spectroscopy (LIBS) and thin section analysis, which shows that FTIR models created using powder mixtures can be utilized for rock samples as well. To expand on the results presented in this thesis, further testing of the models presented here, and looking into models that can predict mixtures of three or more mixtures is recommended.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
FTIR	Fourier-Transfrom Infrared Spectroscopy
SQI	Spodumene-Quartz Intergrowth
BB01	Sample 01 from Bergby
PCA	Principal Component Analysis
XRD	X-ray Diffraction
LIBS	Laser-induced Breakdown Spectroscopy
ATR	Attenuated Total Reflection
SPD	Spodumene
QTZ	Quartz
IR	Infrared
KBr	Potassium Bromide

Symbols

Symbol	Definition	Unit
ρ	Density	[g·cm^{-1}]
χ	Volume fraction	[-]
R^2	Coefficient of determination	[-]

1

Introduction

1.1. Background

Demand for lithium has grown dramatically over the last ten years and is projected to increase even further (Ambrose & Kendall, 2020). In order to keep decarbonising industries, lithium-ion batteries will continue to be needed for years to come. In order to meet this growing demand for lithium, new deposits need to be found and exploited. In Europe, the focus for lithium deposits lies on pegmatite systems (Gourcerol et al., 2019). These pegmatites can contain macroscopic lithium-containing minerals which are easily identified in the field. However, microscopic lithium-bearing minerals can also occur in fine matrix. These occurrences are often overlooked but can host significant amounts of lithium. Ways of quantifying these microscopic lithium occurrences can lead to more deposits being exploited in Europe by not discarding pegmatites solely on their macroscopic features. More exploited pegmatites aids the European autonomy on resources, as the European union currently relies too greatly on importing lithium (Dunlap & Riquito, 2023) and other rare elements which are also often rare elements often found in lithium-bearing pegmatites (Bradley & McCauley, 2013). It is therefore worthwhile to look into developing techniques to quantify pegmatites on their microscopic features.

1.2. Aim

This thesis aims to develop a method to quantify microscopic occurrences of lithium in fine matrix. This is achieved by preparing different mixtures of finely ground powders of different minerals occurring in lithium-bearing pegmatites, and scanning these mixtures with a Fourier-Transform Infrared (FTIR) spectrometer and developing a way to quantify the amount of lithium contained in a sample based on the results.
This thesis has multiple goals. These goals are illustrated below.

- To collect FTIR spectra of selected minerals present in lithium-bearing pegmatites.
- To describe the spectral signatures of these minerals in order to better understand their features and how to tell them apart.
- To develop one or more models that are able to distinguish between a lithium-bearing mineral and a non-lithium-bearing mineral based on mixtures of powdered forms of these minerals.
- To link these models to measurements of rock samples and to see if the models are usable for making predictions on actual rocks.

Briefly put, these aims can be summarized in one research questions, which is: *What are the characteristic features of minerals in lithium-bearing pegmatites and can these be used to develop a model that predicts the lithium-bearing mineral content in fine crystalline matrix?*

1.3. Scope

This thesis proposes models that distinguish between pegmatite minerals, and briefly discusses the infrared spectral features these minerals exhibit. It is beyond the scope of this thesis to go deep into the theory of why these spectral features occur. This thesis also does not discuss the theory of building machine-learning models or the genesis of lithium-bearing pegmatites. It is mostly focused on if it is possible to make predictions using models that use the spectral features of pegmatite minerals.

2

Methodology

2.1. FTIR Data Acquisition

Fourier-transform infrared spectroscopy (FTIR) is a type of infrared spectroscopy that is capable of measuring molecular vibrations. Measuring molecular vibrations using FTIR makes it possible to determine which substance is present. This is usually done by spectral fingerprinting, in which a measured infrared spectrum is compared against a library (Peak, 2005). FTIR spectroscopy is a non-destructive technique, which means that it does not alter or destroy a sample. The infrared spectrum typically consists of the wavelengths from around 700 nm up until $10^6 nm$ or until 10^9 nm, including the microwave region. (Khan et al., 2018)

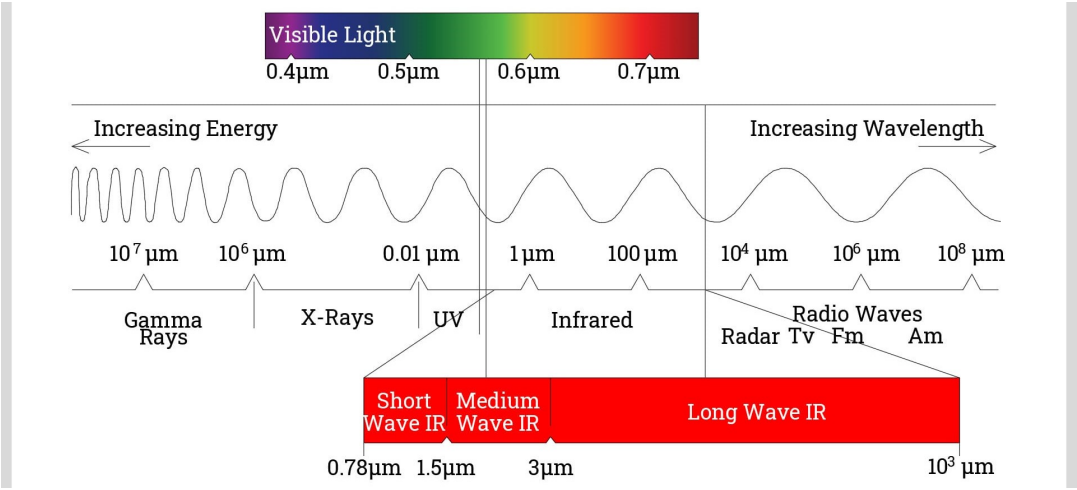


Figure 2.1: Electromagnetic spectrum. Highlighted in the bottom is the region of wavelengths attributed to infrared radiation. (Herschel Infrared, 2023)

All samples were scanned with an Agilent handheld FTIR spectrometer that is capable of simultaneously measuring wavelengths of 5200 to 650 cm^{-1}, or 2 to 15 μm, with a spectral resolution of 4-16 cm^{-1} (Agilent Technologies, 2023). The spectrometer was equipped with a diffuse reflectance attachment. Each scan with this device generally takes less than a minute. After a set time or number of scans, the device has to re-calibrate itself. This was done using a coarse silver attachment included by the manufacturer. The device being handheld also contributes to opportunities to use it in the field, which makes models that can be used to quickly identify samples interesting. The samples that were scanned were mixtures of mineral powders, a lithium-bearing quartz rock, and a range of pegmatite minerals.

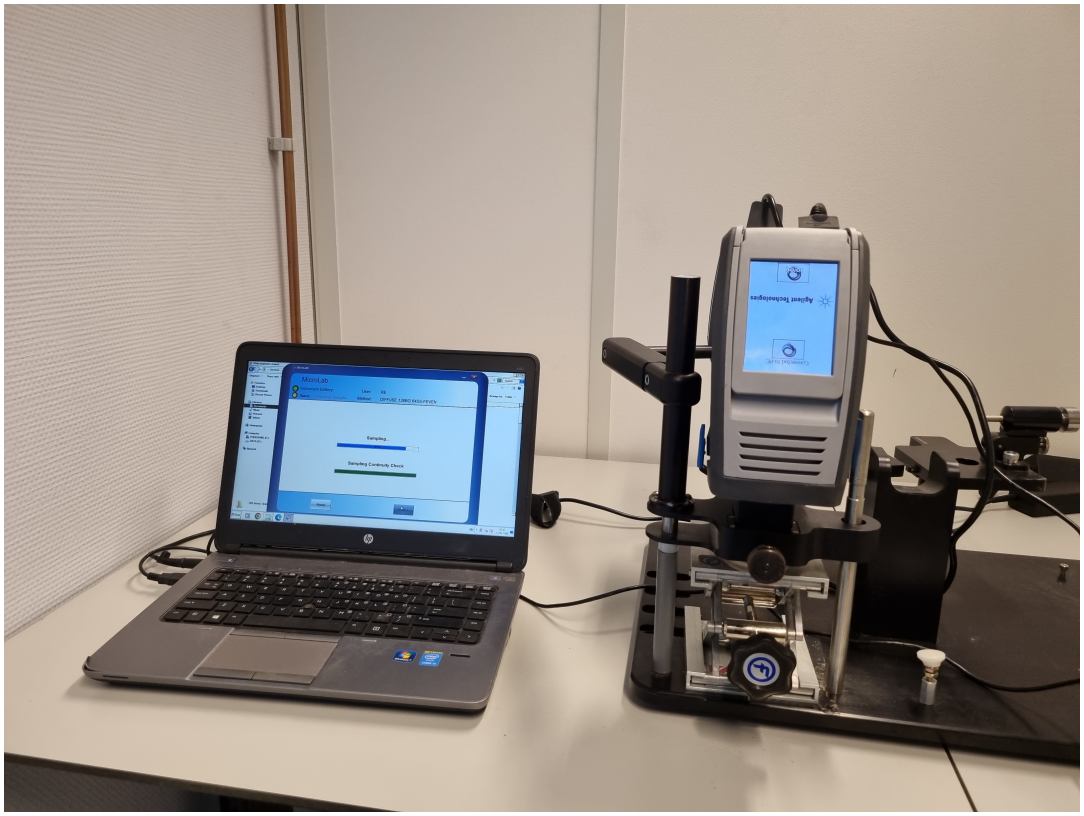


Figure 2.2: FTIR setup. Right: Agilent 4300 handheld FTIR with adjustable sample platform below, left: computer running the corresponding FTIR software.

2.1.1. Powder Measurements

Two sets of mixtures were prepared. Mixture set one consisted of quartz and petalite, and mixture set two consisted of quartz and spodumene. The powders that were used were fine ceramic-grade powders acquired from a commercial source. For each sample, ten grams of material was prepared by unit of weight percentage. The exact concentrations can be seen in Table 2.1 and Table 2.2. The samples were prepared using a scale that was capable of weighing with an accuracy of 0.01 grams. Once a powder mixture had been prepared, it was mixed by shaking it thoroughly in a sealed container. This was then transferred to a sample holder, where a couple of grams of material was packed into a circular depression and smoothed out, leaving a smooth circular surface of powder. This sample could then be placed underneath the spectrometer and raised towards the lens so that it touched slightly. Starting out, ten repetitions were performed, but it was found that the difference between measurements was very small. The different measurements had exactly the same features at the same wavelengths and the reflectivity was largely the same as well. This is why it was deemed unnecessary to scan the powders this many times. In the end it was decided to repeat this procedure three times for each mixture to still have some redundancy in the data.

Sample	Petalite [Wt%]	Quartz [Wt%]
1	100	0
2	95	5
3	90	10
4	85	15
5	80	20
6	75	25
7	70	30
8	65	35
9	60	40
10	55	45
11	50	50
12	45	55
13	40	60
14	35	65
15	30	70
16	25	75
17	20	80
18	15	85
19	10	90
20	5	95
21	0	100

Table 2.1: Overview of petalite-quartz mixtures

Sample	Spodumene [Wt%]	Quartz [Wt%]
1	100	0
2	90	10
3	80	20
4	70	30
5	60	40
6	50	50
7	40	60
8	30	70
9	20	80
10	10	90
11	0	100

Table 2.2: Overview of spodumene-quartz mixtures

2.1.2. Target Rock Measurements

The end goal is to use the models created using the powder mixtures on actual rocks. To test the model, one rock in the ore collection of the department of Geosciences and Engineering at Delft University of Technology which is suitable. This rock was taken from a system of system of lithium-bearing pegmatites in Bergby, Sweden (Leijd et al., 2023). From hereafter this rock will be called BB01, after Bergby, stop 01, for convenience. Lithium occurs mainly in a subgroup of pegmatites called Lithium-Caesium-Tantalum (LCT) pegmatites. These pegmatites are formed from the last part of a granitic melt to crystallize, which causes the pegmatite to be enriched in rare elements. (Bradley & McCauley, 2013) The system of pegmatites at Bergby is such an LCT-pegmatite system (Leijd et al., 2023). BB01 shows intergrowth of quartz with spodumene. Intergrowth of spodumene and quartz has also been identified in other pegmatites (Thomas et al., 1994). The term spodumene-quartz intergrowth (SQI) is used for this type of texture (Lima & Dias, 2019). The rock has been cut in advance so that it has a flat surface. BB01 was scanned using the FTIR spectrometer in handheld mode, holding the lens as tightly as possible to the rock. Two types of measurements were performed. In one measurement a grid of 18 points was measured in regular intervals along the face of the rock. In the other measurement a line was measured along the top of the sample near the side in regular intervals from top to bottom.

Figure 2.3: Top and bottom cut faces of sample BB01.

2.1.3. Mineral Measurements

In order to better understand the powder measurements, the measurements on BB01 and the link between them, samples of spodumene, quartz and petalite minerals were scanned as well. These minerals were taken as pure as possible from the ore collection of the department of Geosciences and Engineering at Delft University of Technology. They were scanned in the same way as BB01, by placing the lens of the FTIR spectrometer as tightly as possible on the mineral in multiple different locations. Table 2.3 Outlines gives an overview of which minerals were scanned. Figure 2.4 gives a visual overview of the samples listed in Table 2.3.

Mineral	Origin	ID	Filename
Spodumene (Kunzite)	Unknown	#1	SpodumeneCrystal_1
Spodumene	Unknown	#2	SpodumeneCrystal_2
Spodumene	Zimbabwe	12124	SpodumeneCrystal_3
Spodumene	Zimbabwe	16922	SpodumeneCrystal_4
Spodumene	Zimbabwe	16485	SpodumeneCrystal_5
Spodumene	USA, Huntington, MA	13509	SpodumeneCrystal_5
Spodumene	Unknown	27087	SpodumeneCrystal_6
Quartz	Unknown	5656	QuartzCrystal_1
Quartz (rose)	Unknown	14674	QuartzCrystal_2
Petalite	Norway, Kragerö	16817	PetaliteCrystal_1
Petalite	Zimbabwe	16488	PetaliteCrystal_2

Table 2.3: Tabular overview of the scanned minerals

Figure 2.4: Visual overview of some of the scanned minerals. top left: spodumene (6 samples), right: petalite (2 samples), mid bottom: quartz (2 samples).

2.2. Data Processing and Model

The data was processed in four steps. These processing steps are meant to improve the signal-to-noise ratio and to remove any unnecessary data. The data was processed using python. All of the code that was used can be found in Appendix A.

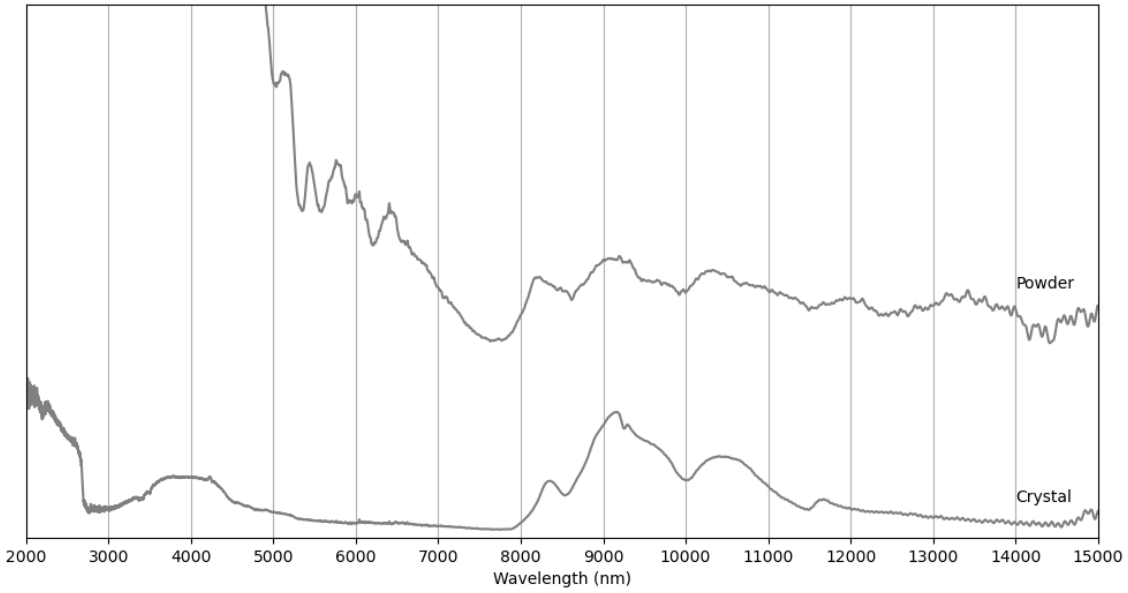


Figure 2.5: Comparison of crystalline and powdered FTIR measurements of spodumene

Figure 2.5 shows the difference in spectral signature between a powdered spodumene sample and a crystalline one. Below 7000 nm the powdered sample is a lot more reflective than the crystalline sample. This effect is very large and not desirable since the spectral signatures should be as similar as possible to make sure the model works on both rocks and powders. Because of this dissimilarity these shorter wavelengths were simply removed. The reason for the high reflectivity in the short wavelengths could

be a result of the measuring method. Similar measurements in literature are done using either attenuated total reflection (ATR) attachments for the FTIR spectrometer (Silva et al., 2009) or by preparing a small amount of powder together with KBr, which is transparent to infrared, and making a pellet out of that. (Chen et al., 2014)
The mean value of each wavelength was subtracted from the data. This was done to try to prevent the model from training on the mean of the data, and on the features instead. A rolling average of n data points was applied to each measurement, in order to improve signal-to-noise ratio. To determine the number of points in the window, a range of numbers from 10 to around 1000 was iterated over. For every iteration the rolling average was taken using the current window width. The model was then retrained and the test score was recorded. The window width for which the test score was highest was chosen. Using this method, windows of around 100 data points proved to be an optimal number based on the test score. Lastly a principal component analysis (PCA) was applied to reduce the dimensionality of the data. This algorithm reduces the dimensionality of a data set. The PCA algorithm aims to maximize the amount of variance in the data set using n components, where n can be defined by the user (Jolliffe & Cadima, 2016). To determine n, the amount of variance explained for each n-component PCA from $n = 1$ to $n_{max} = n_{samples}$. These variances were plotted, and the amount of components for which the variance did not increase much more was chosen as the number of components. This method is often called the ”Elbow method” and is often used in machine learning, specifically in clustering. (Cui et al., 2020).

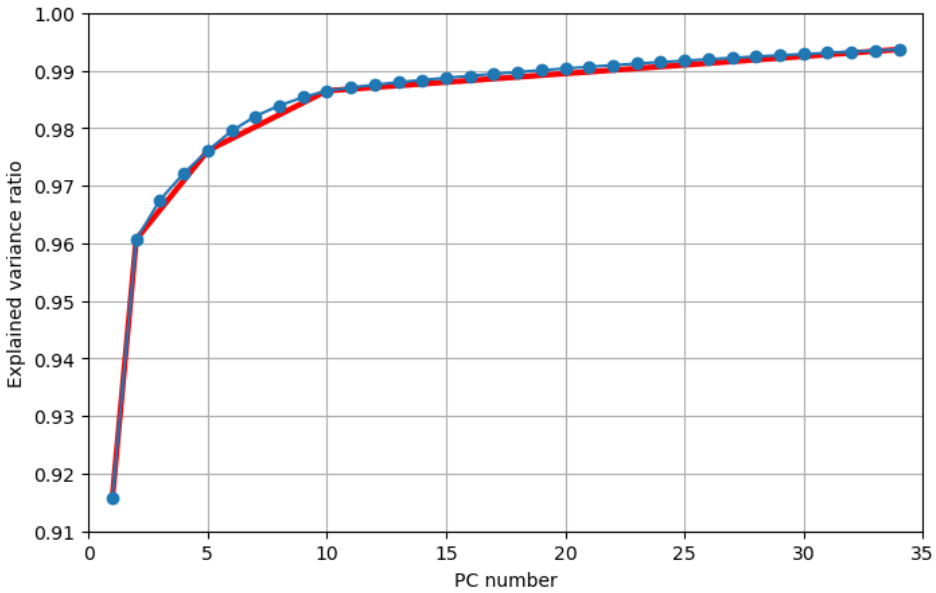


Figure 2.6: Explained variance per n-component PCA.

This number was equal to 10 as can be seen in Figure 2.6. Note that there are ”elbows” at $n = 1$, $n = 5$ and $n = 10$. $n = 10$ was chosen as this explained the most variance by being the final elbow. Thus, originally, every measurement consisted of 18674 individual data points, and PCA reduces this to 10 data points while retaining more than 95% of the original variance. This simplifies both computation and storage of the data, at the cost of interpretability of which wavelengths make up which components. Whereas in the original data each feature was a certain wavelength, after PCA it is unclear what each feature means because of the combination and removal of features. To summarise, the following processing chain was applied to the data: Remove the wavelengths shorter than 7000 nm, Subtract the mean, smooth out each measurement using a rolling average of 100 data points, reduce the dimensionality by applying a principal component analysis of 10 components. Notably, the parameters of the processing chain

remained the same for every measurement fed into the model, which includes the unknown measurements on which the model was later applied. This makes sure every FTIR measurement is treated in the exact same way. An example of this is the PCA. The PCA is instantiated on the training data, which means that the algorithm computes the optimal way to maximize the amount of information retained from the training data with only 10 components. If a new PCA is instantiated on the data that the model is applied to, these 10 components might be calculated differently. The way to overcome this is to use the PCA that was instantiated on the training data to transform any other data as well.
The processed data was used to fit a regression model. The models that were tried were linear regression, ridge regression and support vector regression. Linear regression estimates coefficients for each feature of the data set. These coefficients multiplied with the value for the feature and summed together give the value of the regression. This is analogous to a straight line through the feature space (IBM, 2023). Ridge regression is used often when multiple features are highly correlated. This makes it suitable in theory for FTIR, since wavelengths near each other are often correlated (van Wieringen, 2015). Support vector regression uses support vectors to plot a straight line through the data set. A support vector is a vector from the regression line to a data point. The algorithm tries to adjust the line so the size of the support vectors is greatest. A way of visualizing this is imagining all support vectors as springs pushing the regression line. The springs all try to be as relaxed as possible, pushing the regression line as far away from the points, maximizing the margin between the regression line and the data point. Notably, different kernels can be specified for support vector regression, which does not limit the regression to being linear (Basak et al., 2007). Ridge regression and support vector regression have multiple parameters that can be optimized, which is why a grid search was performed in order to find the best scoring combination of parameters. To determine the best model, some spectra were deliberately left out during training. This data could then be used to assess the models performance. This prevents a phenomenon called overfitting, which would occur if the model would be chosen based on the performance on the training set. This was implemented a little differently for each model, as there were less spodumene mixtures. For the petalite model, the second sample set offset by 5 was used as a testing set. For the spodumene model the model was retrained multiple times, each time leaving one sample out and predicting the spodumene content in that sample. This was done for every sample to create a representative test set. The metric used to evaluate the accuracy of the model is called the coefficient of determination, or R^2 (Chicco et al., 2021) and can be seen below. In the end linear regression performed the best on the unseen data and was chosen to make the models with.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{2.1}$$

Where y_i is the i-th observed value, $\bar{y}$ is the mean of the data, and f_i is the i-th theoretical value. following from this equation, the optimal R^2 score would be 1.

2.3. Model verification

The concentration of lithium-bearing minerals in BB01 is unknown. To know if the values that the model predicts for this rock are valid, verification is needed. This was done using two different methods: LIBS spectroscopy, which uses a laser to determine the mineral content in a sample, and thin sections, which can be used to determine mineral distributions in a sample.

2.3.1. LIBS Measurements

To validate the model's performance on BB01, Laser-Induced Breakdown Spectroscopy (LIBS) measurements were performed. In this method of spectroscopy, a small laser pulse of high intensity is launched towards the rock. This laser pulse causes a small part of the rock to turn into a plasma. This plasma emits a certain spectrum which can be analysed to determine the element concentration in the rock (Haider et al., 2012). The LIBS spectroscope that was used was a Keyence EA300 Laser-based elemental analyser (Keyence, 2023). This spectroscope can fire multiple laser bundles in a grid and average them to get a more representative reading. This setting was used where possible. If the area in focus was large enough, a 5x5 grid of 25 points was used. If the area was smaller, a 3x3 grid of 9 points was used. The weight percentages per element for each measurement were output to a file, which could then be used to calculate the weight percentages of spodumene in the samples, given it has a known formula and all the lithium in the sample is attributed to spodumene. The amount of moles could be calculated by dividing each element percentage by its molar weight. The chemical formula for spodumene is $LiAlSi_2O_6$. If the amount of moles for lithium is equal to n and all the lithium in the measurement is attributed to spodumene this number n can be multiplied by the molar weight of spodumene to get the total weight of spodumene in the sample. If the amount of moles used in spodumene is subtracted from the total, e.g. n for lithium, $6n$ for oxygen, any moles left can be attributed to other minerals that were present in the sample. These moles per element can be multiplied again by their respective molar masses, which, when summed together gives the total weight of other material in the sample. Lastly, the weight of spodumene can be divided by the sum of the two masses, resulting in the weight percentage of spodumene. Note that even though the amount of moles was calculated, this was not actually the amount of moles measured. This amount of moles is purely a virtual amount, just as the weights, but this virtual weight is divided by in the end, ending up with a percentage again.

2.3.2. Thin Sections

To verify the lithium-bearing mineral content in BB01, thin sections were made and analysed. A thin section of a representative part of the rock was placed under a polarizing microscope. These thin sections are thin enough to allow light to pass through, which makes it possible to identify the minerals based on the colors that it produces under polarizing light, the cleavage of the crystals and the general shape of the crystals. Pictures were taken of the sample under polarizing light. Following a regular grid that totalled 130 points, points were counted to be either spodumene or quartz, to get an estimate of the spodumene content in the source rock. However, the ratio calculated using this method is not sufficient, as this ratio is based on an arbitrary area of the thin section. The unit we are after is Wt%. That is the ratio has to be normalized based on the densities of spodumene and quartz, according to the following formula.

$$\frac{\chi_x \cdot \rho_x}{\chi_{spd} \cdot \rho_{spd} + \chi_{qtz} \cdot \rho_{qtz}} \quad \rho_{qtz} = 2.65\ g \cdot cm^{-1}, \quad \rho_{spd} = 3.18\ g \cdot cm^{-1} \tag{2.2}$$

Where χ is the fraction of space the mineral x occupies in the thin section, ρ_{spd} is the density of spodumene and ρ_{qtz} is the density of quartz.

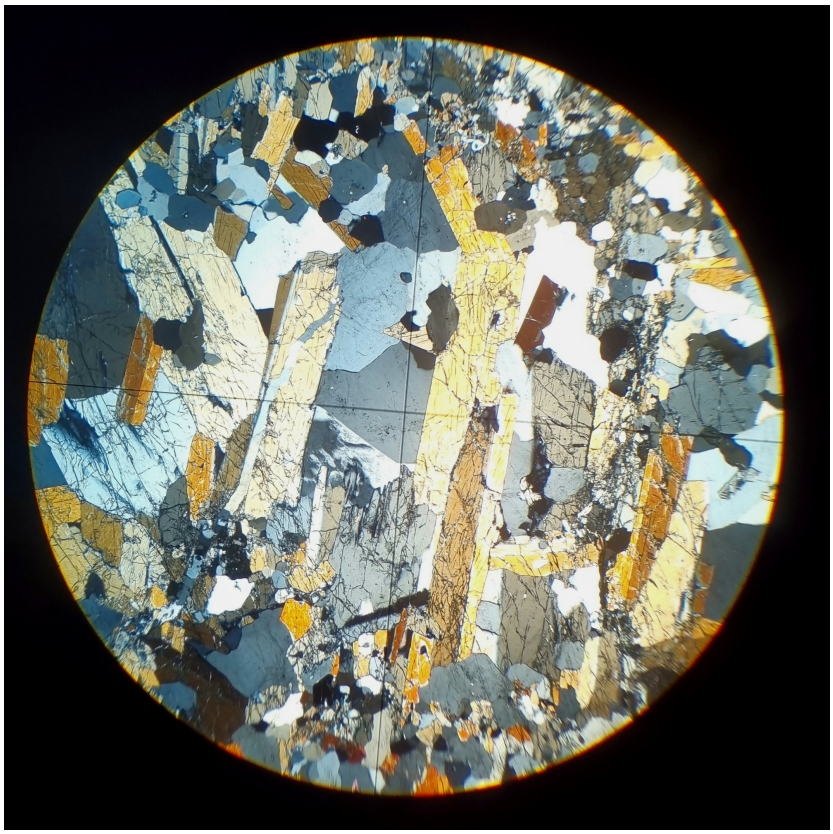


Figure 2.7: Photo of the thin section. Elongated orange to brown minerals that show cleavage are spodumene, White to black minerals are quartz.

3

Results

3.1. FTIR Spectral Features of Lithium-bearing Pegmatite Minerals

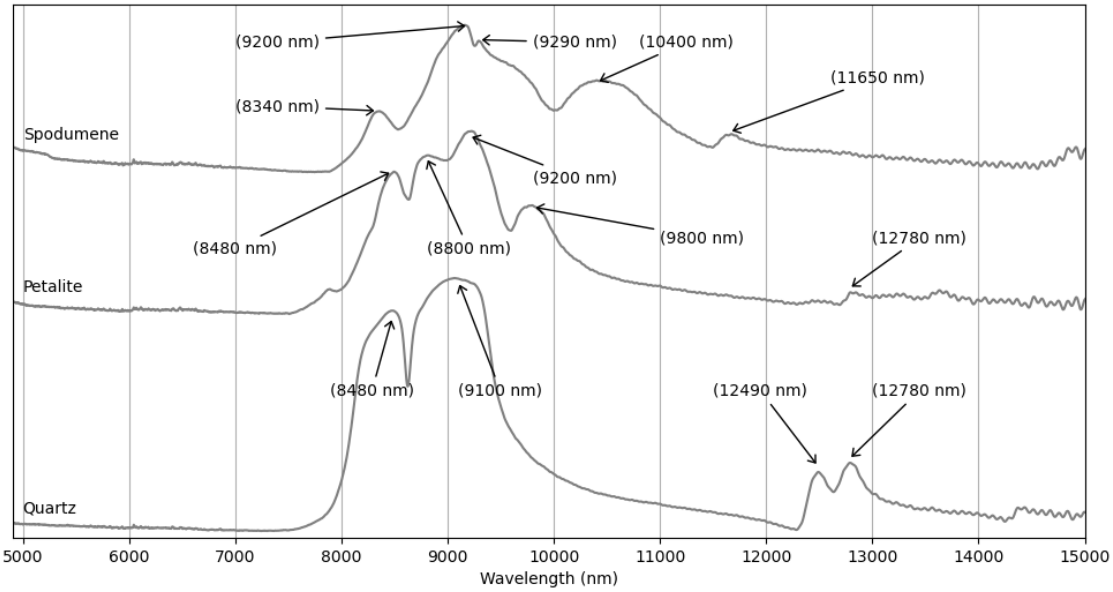
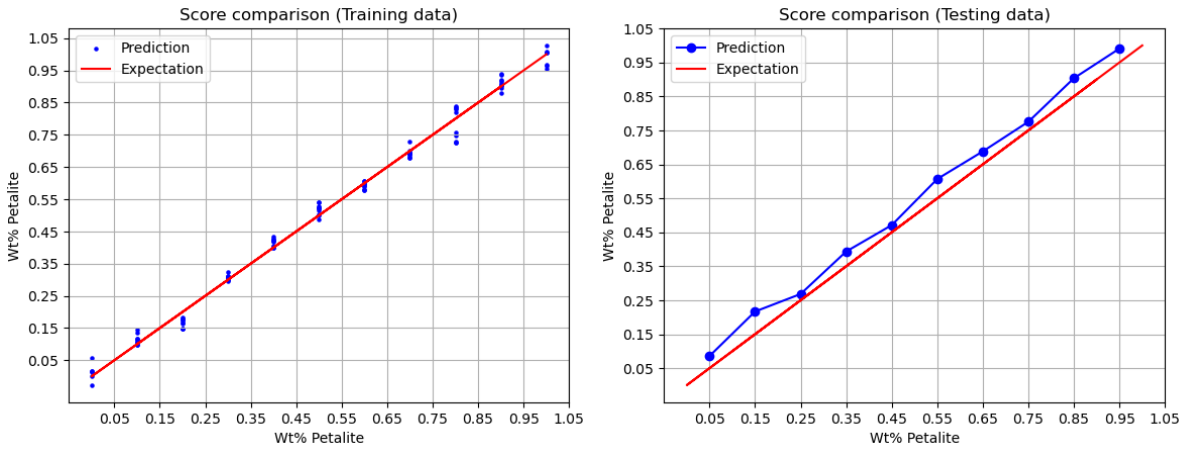


Figure 3.1: Spectral overview of the examined minerals

As outlined in section 2.1, the above are the results of the minerals scanned. In quartz, the main reflectance features can be seen at 8480, 9100, 12490, and 12780 nm. In petalite, the main features are located at 8480, 8800, 9200, 9800 and 12780 nm. In spodumene, the main features are at 8340, 9200, 9290, 10400, and 11650 nm. The next section will present the models created for distinguishing between quartz and petalite or quartz and spodumene.

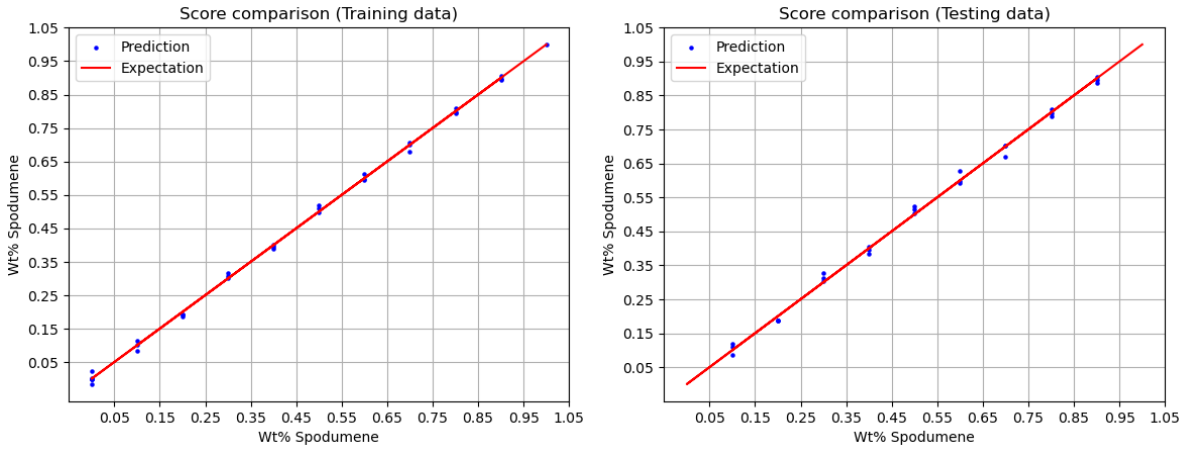
3.2. Linear regression models

As outlined in chapter 2, the model that performs best based on the data is a linear regression model. The performance of this model, built with spectral signatures of the powder mixtures described in section 2.1, is shown in Figure 3.2 and Figure 3.3 below.



(a) **(b)**

Figure 3.2: Differences between the ideal outcome and model predictions of the a) training set, b) testing set of the petalite model



(a) **(b)**

Figure 3.3: Differences between the ideal outcome and model predictions of the a) training set, b) testing set of the spodumene model

The mean difference between the predictions on the test set for petalite and the theoretical values is 0.0400 or 4.00%, with a standard deviation of 0.0150 or 1.50%. For spodumene, this difference is 0.0113 or 1.13% with a standard deviation of 0.0140 or 1.4%. This shows, along with Figure 3.2, that the predictions of the model are not far off compared to the actual value of each point. The R^2 scores for the models are 0.978 for petalite and 0.997 for spodumene.

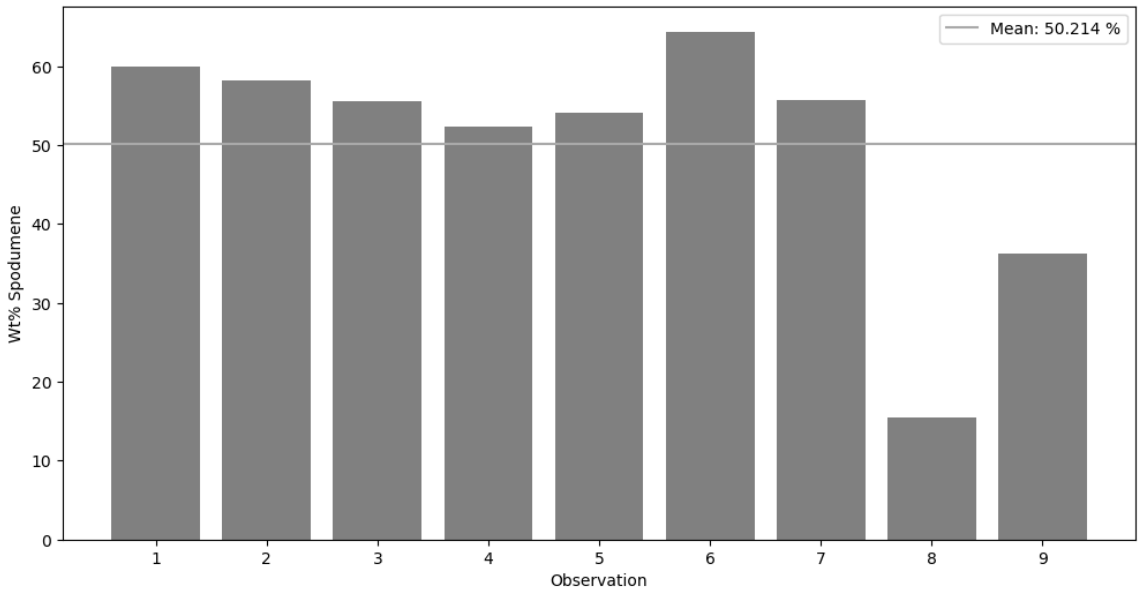


Figure 3.4: Model predictions along a line on the sample

The first seven predictions the model makes on the measurements along the line have a small variation as can be seen in Figure 3.4. Most of the measurements are around 0.5 or 50% spodumene. The standard deviation for the first seven measurements is 0.0376 or 3.76%, which shows that the model predicts a rather homogeneous spodumene content along the profile of the rock for these measurements. For the last two measurements, taken near the bottom of the sample, the model predicts a significantly lower amount of spodumene, at only 15.4% and 36.2% respectively. With these two samples included, the standard deviation is much higher, at 14.31%.

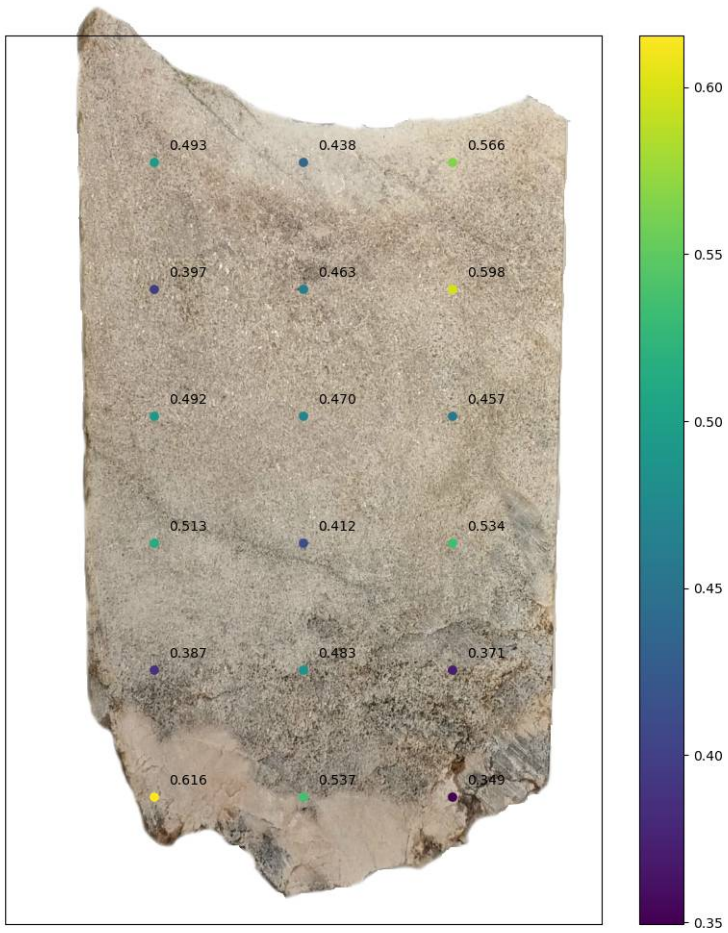


Figure 3.5: Model predictions in units of weight fraction of spodumene, following a grid of measurements on the sample, as an overlay on the picture to show where the measurements were taken.

The predictions of the spodumene model on the grid measurements taken of BB01 can be seen in Figure 3.5. As stated in chapter 2, 18 points were sampled in a regular grid on the rock. In the figure, these predictions are placed at the location where the measurements were taken . The numbers indicate the weight fraction of spodumene that the model predicts for the given point. The bottom rightmost point is the lowest prediction, at 0.349 or 34.9% spodumene. The bottom leftmost point is the highest prediction, at 0.616 or 61.6% spodumene. The mean value of spodumene of all points in the grid is 0.476 or 47.6%. The standard deviation is 0.074 or 7.4%.

3.3. LIBS and thin section validation

Measurement	Number	Spodumene content [Wt%]	
LIBS	1	37.53	
	2	42.90	
	3	42.90	
	4	42.90	
Thin section	1	49.93	

Table 3.1: LIBS and thin section results for BB01

Outlined in Table 3.1 are the calculated spodumene contents of the LIBS method and the thin section method. The LIBS measurements were taken with 5x5 point grids where possible, resulting in four measurements which averages 41.56%. For the thin section, a 130 point count of a grid overlayed on the image of the thin section as outlined in chapter 2 results in 49.93% after being corrected for mineral densities as explained in subsection 2.3.2.

4

Discussion

4.1. Spectral features of lithium-bearing pegmatite minerals

In the quartz spectrum, the peaks at 8300-8500 nm and at 9100-9200 nm strongly resemble the quartz reststrahlen bands (Laukamp et al., 2021). The peaks at 12490 nm and 12780 nm could well be the secondary quartz reststrahlen bands, most notably visible in the quartz itself, less so in the other minerals. These bands are attributed to fundamental stretching vibrations of Si-O bonds (Laukamp et al., 2021). The restrahlen bands can be recognized by an area of high reflectance that has a dip in reflectance in the middle, producing an M-shape (Spitzer & Kleinman, 1961). In petalite, the peak seen around 9800 nm is a feature observed in minerals such as biotite and Muscovite (USGS Database, 2023) but not in the quartz and spodumene scanned in this thesis. Biotite and muscovite are both phyllosilicates. Petalite is also a phyllosilicate. this feature is thus most likely a phyllosilicate features. In spodumene, the peaks observed at 10400 nm and 11650 nm can also be observed in other minerals such as diopside and augite. (RRUFF Database, 2023). These minerals are pyroxene minerals, which spodumene is as well. This and the fact that these peaks are not present in quartz and petalite leads to the conclusion that these are pyroxene features.

4.2. Linear regression models

As seen in chapter 3 the model test predictions are accurate to around 4% with a standard deviation of 1.5% for the spodumene model, and close to 1% with a standard deviation of around 0.1% for the petalite model. This shows that the model works well on the powder mixtures.

The powder mixtures used in this thesis have a certain grain size distribution. This distribution is unknown at this time, but it is very fine. As seen in Figure 2.5, the texture and grain size of a sample can cause a big difference in the FTIR spectral signatures between two samples, even though they are identical in composition. The fact that the model works on this grain size distribution thus does not mean that it works well on powders that have other grain size distributions, or rock samples that have other textures or grain sizes.

As stated in chapter 2, the powders were acquired from a commercial source. They were labelled as petalite, quartz and spodumene powders, but the degree of purity of these powders is unknown at this time. This will have introduced a systematic error in the models, as the assumption was made that these powders were pure. The golden rule of machine learning is that ”Garbage in = Garbage out” (Rose & Fischer, 2011), which states that if the quality of the input data used to build a model is poor, the predictions it will make will be poor as well. Because the exact composition of the powders is unknown at this time, impurities can have had a negative effect on the model.

4.3. LIBS and Thin Section Validation

The results show that the model predicts an average spodumene content of 50.2% for the line measurement of BB01 and 47.6% for the grid measurement. In addition, the LIBS measurements average a weight percentage of 41.6% and the thin section analysis results in a weight percentage of 49.9% for BB01. Thus, these two methods of validating the spodumene content that the spodumene model predicted are around the same values that the model presented here predicts. Thus, the model results look promising. Of course these methods are not fully accurate themselves; thin sections and LIBS measurements analyze a very small area of rock which may not be representative for the whole sample.

4.4. Errors

In section 2.2, a big difference can be seen in reflectivity in the FTIR spectra of the powder and crystal measurements. This difference can be explained by two things. One way this difference can be explained is by the way the infrared radiation interacts with the samples. When the sample is powdered, its surface is not smooth at a microscopic level. This rough surface causes scattering of the infrared radiation. This scattering bounces more radiation back to the detector, which causes an extremely high level of reflectance being measured. In (Mayer et al., 2022) powders are scanned using FTIR as well, but there is no mention of this high degree of reflectance. This paper uses the KBr method outlined in (Chen et al., 2014). This method thus causes less scattering of IR radiation, because there is less sample for the radiation to reflect on. A second way the difference can be explained is because of defects in the machine. The FTIR spectrometer used here may have a number of issues that contribute to the high reflectivity measured. When measuring the powders, it repeatedly measured reflectivity values of over 100%. In some cases the reflectivity even measured >200%. In other measurements not involving powders it has measured reflectivity values of over 100% as well. This could signify that the machine has to be returned to the manufacturer for inspection. On the other hand, this does not influence the results as the peaks of samples have been correctly measured because they match up with databases and literature. The only thing that might be wrong are the reflectivity values.

5

Conclusion

This thesis presents the results of experimenting with models which aim to respectively predict the weight percentages of petalite and spodumene where these minerals are mixed with quartz, from a given FTIR measurement. The results of this model were verified using laser-induced breakdown spectroscopy and thin section analysis. The main results of this thesis are the following:

1. It is possible to distinguish between quartz, spodumene and petalite by looking at their FTIR spectral signatures, because each mineral features at least one characteristic peak not present in the other two minerals and has a general characteristic spectral signature that is not similar to the other two minerals.
2. The results show that the petalite model is able to accurately predict the weight percentage of petalite from an FTIR measurement of a petalite-quartz powder mixture, and the spodumene model is able to accurately predict the weight percentage of spodumene from an FTIR measurement of a spodumene-quartz powder mixture.
3. The spodumene model works when applied to an FTIR measurement of a pegmatite sample that shows intergrowth of spodumene and quartz. The model predicts a weight percentage of spodumene that is in line with other types of measurement such as LIBS and thin section analysis.

In short, this thesis shows that it is possible to distinguish between two minerals by looking at their spectral signatures. It also shows that a linear regression model trained on powder samples can accurately distinguish between two minerals and predict the ratio between the two minerals. The results show that the model is also valid when used on rocks instead of powders and that these results agree with other types of measurement.

6

Recommendations

This thesis created two linear regression models which distinguish between two minerals each. As discussed, these models are not perfect and can be improved. The points below illustrate the main recommendations that arose during research and writing that can be explored to advance the topic discussed in this thesis.

Sample standards have the potential to improve the model and the validation. Standards are samples with a independently verified content of, in this case, lithium-bearing minerals (Xu et al., 2001). Having access to a number of standards can serve as a ”ground truth” for the models, which makes it possible to indisputably assess the performances of the models. Such standards can also be used for LIBS validation, as they should give the same known element concentration

In addition to the method used here, literature suggests different ways of scanning powders using FTIR. One of these methods is mixing a tiny amount of sample together with potassium bromide (KBr), and pressing a pellet out of this mixture (Chen et al., 2014). This works because KBr is transparent to infrared radiation and thus has no influence on the FTIR measurement. A second option is using an attenuated total reflectance (ATR) attachment for the FTIR spectrometer. This attachment makes it possible to directly measure a powder sample, without pressing it into a pellet. (Silva et al., 2009) These methods of measuring powders could be tried to improve the similarity between powder measurements and diffuse reflectance rock measurements.

A higher degree of similarity between the two could in turn improve model performance. This thesis demonstrated that the model works for mixtures of two minerals. A logical next step would be to try to expand this to mixtures of three or more minerals. For example, the rock sample used in this thesis contains feldspar, which the models can not account for or recognize. Adding feldspar as a third member could improve the usability of the model, because this opens up the possibility to use the model on more samples.

The powders that were used to build the models presented in this thesis were acquired from a commercial source. These powders might contain impurities which lead to inaccuracies in the model . A next possibility is to analyze the degree of purity of these powders and to acquire purer powders if necessary.

Only one rock sample was scanned. The measurements of this rock were ran through the spodumene model. It is shown that this spodumene model performs in line with other types of measurement. However, a logical next step is to test out both models on more rock samples and to see if they work on other samples as well.

References

Agilent Technologies. (2023). *Agilent technologies*. Retrieved October 25, 2023, from https://www.agilent.com/en/product/molecular-spectroscopy/ftir-spectroscopy/ftir-compact-portable-systems/4300-handheld-ftir

Ambrose, H., & Kendall, A. (2020). Understanding the future of lithium: Part 1, resource model. *Journal of Industrial Ecology*, *24*(1), 80–89.

Basak, D., Pal, S., & Patranabis, D. (2007). Support vector regression. *Neural Information Processing – Letters and Reviews*, *11*.

Bradley, D., & McCauley, A. (2013). *A preliminary deposit model for lithium-cesium-tantalum (lct) pegmatites* (tech. rep.). US Geological Survey.

Chen, Y., Furmann, A., Mastalerz, M., & Schimmelmann, A. (2014). Quantitative analysis of shales by kbr-ftir and micro-ftir. *Fuel*, *116*, 538–549.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623.

Cui, M., et al. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, *1*(1), 5–8.

Dunlap, A., & Riquito, M. (2023). Social warfare for lithium extraction? open-pit lithium mining, counterinsurgency tactics and enforcing green extractivism in northern portugal. *Energy Research & Social Science*, *95*, 102912.

Gourcerol, B., Gloaguen, E., Melleton, J., Tuduri, J., & Galiegue, X. (2019). Re-assessing the european lithium resource potential–a review of hard-rock resources and metallogeny. *Ore Geology Reviews*, *109*, 494–519.

Haider, Z., Munajat, Y., & Raja Ibrahim, R. (2012). Review: Laser-induced breakdown spectroscopy (libs) a promising technique, its limitations and a proposed method. *Jurnal Teknologi (Sciences and Engineering)*, *57*, 45–56.

Herschel Infrared. (2023). *Herschel infrared*. Retrieved October 25, 2023, from https://www.herschel-infrared.co.uk/how-do-infrared-heaters-work/types-of-infrared-heat/

IBM. (2023). *Ibm*. Retrieved October 25, 2023, from https://www.ibm.com/topics/linear-regression#:~:text=Linear%20regression%20analysis%20is%20used,is%20called%20the%20independent%20variable.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Keyence. (2023). *Keyence*. Retrieved October 26, 2023, from https://www.keyence.eu/nlnl/products/microscope/elemental-analyzer/ea-300/

Khan, S. A., Khan, S. B., Khan, L. U., Farooq, A., Akhtar, K., & Asiri, A. M. (2018). Fourier transform infrared spectroscopy: Fundamentals and application in functional groups and nanomaterials characterization. *Handbook of materials characterization*, 317–344.

Laukamp, C., Rodger, A., LeGras, M., Lampinen, H., Lau, I. C., Pejcic, B., Stromberg, J., Francis, N., & Ramanaidou, E. (2021). Mineral physicochemistry underlying feature-based extraction of mineral abundance and composition from shortwave, mid and thermal infrared reflectance spectra. *Minerals*, *11*(4), 347.

Leijd, M., Högdahl, K., Jonsson, E., & Zetterqvist, A. (2023, April). *Field trip guide to the bergby lct-pegmatite field, greenpeg international conference 230426*.

Lima, A., & Dias, F. (2019). Spodumene and quartz intergrowth–textural and genesis point of view. *Geophysical Research Abstracts*, *21*.

Mayer, J. M., Abraham, J. A., Surhigh, B., Kinzer, B., & Chandran, R. B. (2022). Temperature-dependent diffuse reflectance measurements of ceramic powders in the near-and mid-infrared spectra. *Solar Energy*, *245*, 193–210.

Peak, D. (2005). Fourier transform infrared spectroscopy. In D. Hillel (Ed.), *Encyclopedia of soils in the environment* (pp. 80–85). Elsevier. https://doi.org/https://doi.org/10.1016/B0-12-348530-4/00174-0

Rose, L. T., & Fischer, K. W. (2011). Garbage in, garbage out: Having useful data is everything. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 222–226.

RRUFF Database. (2023). rruff.info

Silva, F. E., Ferrão, M. F., Parisotto, G., Müller, E. I., & Flores, E. M. (2009). Simultaneous determination of sulphamethoxazole and trimethoprim in powder mixtures by attenuated total reflection-fourier transform infrared and multivariate calibration. *Journal of Pharmaceutical and Biomedical Analysis*, *49*(3), 800–805.

Spitzer, W., & Kleinman, D. (1961). Infrared lattice bands of quartz. *Physical Review*, *121*(5), 1324.

Thomas, R., Bühmann, D., Bullen, W., Scogings, A., & De Bruin, D. (1994). Unusual spodumene pegmatites from the late kibaran of southern natal, south africa. *Ore Geology Reviews*, *9*(2), 161–182.

USGS Database. (2023). https://crustal.usgs.gov/speclab/QueryAll07a.php

van Wieringen, W. N. (2015). Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*.

Xu, Z., Cornilsen, B. C., Popko, D. C., Pennington, W. D., Wood, J. R., & Hwang, J.-Y. (2001). Quantitative mineral analysis by ftir spectroscopy. *Internet J. Vib. Spectrosc*, *5*(4).

A

Source Code

In this appendix, some important pieces of code are presented. However, to put the full code base here would result in a lot of added pages, which is why it can be found in an added .zip file.

A.1. Code for reading .asp files

```
import os
import pandas as pd
import numpy as np
import glob
import re

def FTIR_Thom(folder_path):
    """Generate wavelength array, FTIR dataframe and sample list for samples with
        known concentration

    Args:
        folder_path (str): the path to the folder in which the .asp files are
            located

    Returns:
        wavelength: array
        df_FTIR: pd.DataFrame
        sample_list: pd.DataFrame
    """
    file_nr = 0

    ratio_dict = {
        "100-0": 0.00,
        "95-05": 0.05,
        "90-10": 0.10,
        "85-15": 0.15,
        "80-20": 0.20,
        "75-25": 0.25,
        "70-30": 0.30,
        "65-35": 0.35,
        "60-40": 0.40,
        "55-45": 0.45,
        "50-50": 0.50,
        "45-55": 0.55,
        "40-60": 0.60,
        "35-65": 0.65,
        "30-70": 0.70,
```

```
        "25-75": 0.75,
        "20-80": 0.80,
        "15-85": 0.85,
        "10-90": 0.90,
        "05-95": 0.95,
        "0-100": 1.00
    }

    sample_list = pd.DataFrame()
    data_folder = os.path.join(folder_path,'*.asp')
    for f in glob.glob(data_folder):

        if file_nr == 0:
            ref_spec=pd.read_csv(f, index_col=None, header=None)

            points = ref_spec[0][0]
            wavi = ref_spec[0][1]
            wavf = ref_spec[0][2]
            res = ref_spec[0][3]

            # Calculate the wavelengths
            step = ((wavi - wavf) / (points) * (-1))
            wavenum =np.arange(wavi,wavf,step)
            wavelength = (10000/wavenum)

            # wavelength = wavenum
            df_FTIR = pd.DataFrame(wavelength,columns=['Wavelength'])

            file_nr += 1

        filename = os.path.basename(f)
        # print("filename:", filename, "\n")

        # Split the filename into (sample / ratio / repetition) and remove ".asp"
        filename_splitted = filename[0:-4].split('_')
        sample = filename_splitted[0]
        ratio = filename_splitted[1]
        repetition = filename_splitted[2]

        # print("sample:", sample)
        # print("ratio:", ratio)
        # print("repetition:", repetition, "\n")

        # Skip first 6 rows but why???
        file_dataframe = pd.read_csv(f, index_col=None, header=None, skiprows=6,
            names=[sample])

        df_FTIR = pd.concat([df_FTIR,file_dataframe],axis=1)

        data=[[sample, repetition, ratio]]
        data = pd.DataFrame(data,columns=['Sample','Repetition','Ratio'])

        sample_list = pd.concat([sample_list,data])

    sample_list["wt% Li"] = [ratio_dict[key] for key in sample_list["Ratio"]]
    df_FTIR=df_FTIR.set_index('Wavelength')
    df_FTIR = df_FTIR.transpose()
    df_FTIR.index.name = 'Sample'

    return wavelength, df_FTIR, sample_list

```

```
def FTIR_Thom_testsamples(folder_path):
    """Generate wavelength array, FTIR dataframe and sample list for samples with
        unknown concentration

    Args:
        folder_path (str): the path to the folder in which the .asp files are
            located

    Returns:
        wavelength: array
        df_FTIR: pd.DataFrame
        sample_list: pd.DataFrame
    """
    file_nr = 0

    sample_list = pd.DataFrame()
    data_folder = os.path.join(folder_path,'*.asp')
    for f in glob.glob(data_folder):

        if file_nr == 0:
            ref_spec=pd.read_csv(f, index_col=None, header=None)

            points = ref_spec[0][0]
            wavi = ref_spec[0][1]
            wavf = ref_spec[0][2]
            res = ref_spec[0][3]

            # Calculate the wavelengths
            step = ((wavi - wavf) / (points) * (-1))
            wavenum =np.arange(wavi,wavf,step)
            wavelength = (10000/wavenum)

            # wavelength = wavenum
            df_FTIR = pd.DataFrame(wavelength,columns=['Wavelength'])

            file_nr += 1

        filename = os.path.basename(f)

        # Split the filename into (sample / ratio / repetition) and remove ".asp"
        filename_splitted = filename[0:-4].split('_')
        sample = filename_splitted[0]
        repetition = filename_splitted[1]

        file_dataframe = pd.read_csv(f, index_col=None, header=None, skiprows=6,
            names=[sample])

        df_FTIR = pd.concat([df_FTIR,file_dataframe],axis=1)

        data=[[sample, repetition]]
        data = pd.DataFrame(data,columns=['Sample','Repetition'])

        sample_list = pd.concat([sample_list,data])

    df_FTIR=df_FTIR.set_index('Wavelength')
    df_FTIR = df_FTIR.transpose()
    df_FTIR.index.name = 'Sample'

    return wavelength, df_FTIR, sample_list
```

A.2. code used to create the model

```
# random state
random_state=42

# Make X and y data
X_train = df_FTIR_train.to_numpy()
X_test = df_FTIR_test.to_numpy()
y_train = sample_list_train["wt% Li"].to_numpy()
y_test = sample_list_test["wt% Li"].to_numpy()

# Remove wavelengths < 2.75 um from index 5854 onwards
# Remove wavelenghts < 7.00 um from index 15332 onwards
X_train = X_train[:, 15332:]
X_test = X_test[:, 15332:]

# Standardize
scaler = StandardScaler(with_std=False).fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)

# Smooth
window = 100
df_train = pd.DataFrame(X_train)
df_test = pd.DataFrame(X_test)
df_train_smoothed = df_train.rolling(window=window, axis=1).mean()
df_test_smoothed = df_test.rolling(window=window, axis=1).mean()

# Cut off frequencies that turned to NaN due to smoothing
X_train = df_train_smoothed.to_numpy()[:,window-1:]
X_test = df_test_smoothed.to_numpy()[:,window-1:]

# PCA
n_components = 10
pca = PCA(n_components=n_components).fit(X_train)
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)

regression = LinearRegression().fit(X_train, y_train)
score_training = regression.score(X_train, y_train)
score_testing = regression.score(X_test, y_test)

print(f'Training score: {score_training}')
print(f'Testing score: {score_testing}')
```