CHALMERS
UNIVERSITY OF TECHNOLOGY

Multivariate linear regression of LIBS spectra

Master's thesis in Engineering Mathematics and Computational Science

KRISTER EKSTRÖM

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020:NN

Multivariate linear regression of LIBS spectra

KRISTER EKSTRÖM

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden 2020

Multivariate linear regression of LIBS spectra

Krister Ekström

Supervisors:
Moritz Schauer, Department of Mathematical Sciences, Chalmers and Gothenburg University
Jonas Petersson, Swerim AB

Examiner:
Aila Särkkä, Department of Mathematical Sciences, Chalmers

Master's Thesis 2020
Department of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2020

Abstract

Laser induced breakdown spectroscopy (LIBS) is a spectroscopic technique for chemical analysis. LIBS can be used in rough environments and measurements can often be made without any sample preparation. These properties make LIBS interesting for in-situ measurements in industrial settings. However this physical robustness and flexibility come at a cost: There are processes involved in a LIBS measurement that are difficult to model. Therefore statistical methods constitute the best choice for analysis of LIBS spectra. The accuracy and robustness obtained with current methods have not been sufficient for widespread industrial adoption of LIBS. Careful statistical analysis is required to further develop LIBS analysis and reach a level of robustness and accuracy that enables widespread industrial adoption. This thesis aims at contributing to that development by developing a method for multivariate linear regression of LIBS spectra. The first part uses multivariate linear regression on a calibration set to estimate the spectra of the elements in the sample, referred to as the pure spectra. These pure spectra are then used to predict the concentrations of new samples. The second part is a Bayesian continuation, utilizing a model from CF-LIBS to create a constraining prior for the regression. The method constitutes a change in perspective from previous multivariate attempts where the concentrations are inferred from the spectrum using methods such as partial least squares, principle component regression or multiple linear regression. The method presented in this thesis builds on the view of the spectrum as a multivariate response to the concentrations. A relatively simple model is suggested for this response. This model builds on common assumptions made when analyzing LIBS spectra using the conventional univariate approach. As a result of this model-based approach, the method is not only more interpretable and easier to develop but perhaps more importantly the degrees of freedom are decoupled from the number of pixels in the spectrum. Instead the degrees of freedom are determined by the number of elements in the analysis with two orders of magnitude improvement.

Contents

1 Introduction

This work is aimed at developing quantitative analysis of data obtained with Laser-induced Breakdown spectroscopy (LIBS), which is a spectroscopic method for chemical analysis. By focusing a pulsed laser on a material, a small part of the material is ablated and excited to a plasma. In this state the molecules have broken down to their constituent elements and the atoms are brought to higher energy levels (i.e. excited levels). The emission resulting from the atoms returning to their ground states contains information on the chemical composition of the material. The emission is detected with a spectrometer and its spectrum is analyzed. The analysis could be either qualitative or quantitative. As a qualitative tool it can be used for a wide range of applications such as to separate different aluminum alloys in a recycling plant or to check the claimed origin of red wines [1][2][3]. As a tool for quantitative analysis however it is still lacking in accuracy. There is a strong interest in developing improved methods to process LIBS data, and this thesis will investigate new approaches that have so far not been presented in literature.

The thesis connects to ongoing LIBS-work at Swerim, where the two major projects are CONSENSO and AUSOM.[1] The aim of CONSENSO is to enable in-situ measurements of the chemical composition of slag in metallurgical processes. These measurements could then be used to optimize the production processes. The characteristics of slag vary depending on its chemical composition and with better quality control it has potential to become a usable byproduct instead of a waste stream. The project AUSOM aims at making sorting equipment based on LIBS-sensors. One goal is to perform analysis so that a decision on the sorting fractions can be made based on the chemical content. Quantitative analysis allows for the operator to adapt the classification or sorting procedure to changing requirements.

1.1 Outline of thesis

This thesis represents an attempt at creating a method for multivariate regression of LIBS spectra and concentration. To effectively present this, the thesis is split into six sections. The introduction presents the background and aim for the project. Additionally, an introduction to the properties of LIBS and some existing quantitative methods of analysis is included. The introduction section is followed by a theory section where the necessary theory for subsequent chapters is presented. This is followed by the model section, where the conventional univariate method serves as the basis for the development of two multivariate approaches. The two approaches are tested on a dataset of aluminum samples and the result is presented in the subsequent result chapter. The approaches and the results are then discussed in the discussion section and finally a conclusion is presented in the conclusion section.

[1]both CONSENSO and AUSOM are financed via EIT RawMaterials under the Horizon 2020 EU framework program.

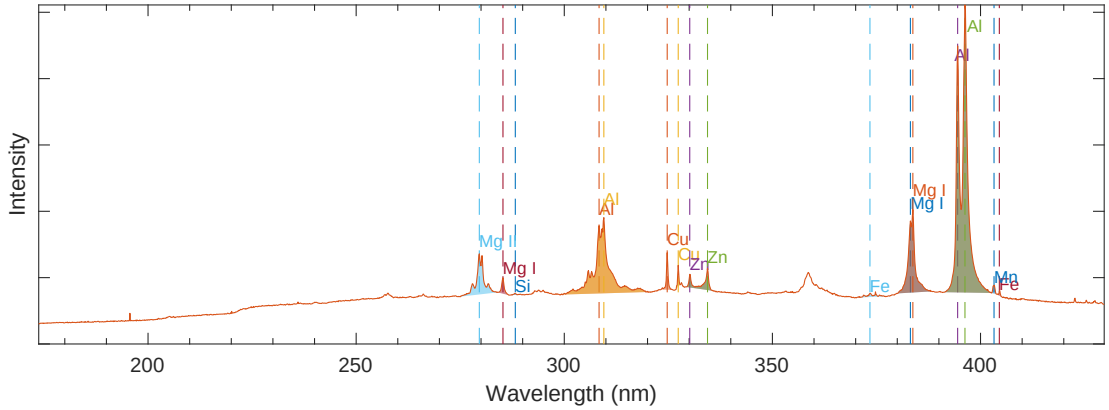

Figure 1: A LIBS spectrum of an aluminum alloy with some of the peaks marked with their corresponding element.

1.2 Properties of LIBS spectra

The classic approach to quantitative LIBS analysis is a univariate method where a function, a calibration curve, is fitted, for each element, to a calibration dataset. For each element, the analyst selects a peak at a wavelength where the element is known to emit. This peak should ideally follow an affine relationship with the concentration of the element and have no interference from other elements in the reference samples. Affine functions can then, in principle, be fitted to infer the concentrations of elements to the maximum intensity or area of the corresponding peak.

A typical LIBS spectrum is seen in figure 1. The spectrum consists of signal from a number, J, of pixels. Each pixel, indexed with j, registers light in a small range of wavelengths. Put together with the wavelengths of the pixels on the horizontal axis and the signal on the vertical axis, a LIBS spectrum is formed. As seen, the various constituting elements have several emission peaks and with varying peak intensities. The emission intensity is different from each element and also the signal strength typically varies a lot from one measurement to the other. It is therefore customary to normalize the spectra. Total intensity can be used, i.e. the sum of the intensities of all the pixels, but as some elements have far more peaks than others this can introduce bias. Another method for normalizing spectra is by dividing each spectrum by the intensity of a reference peak [4]. This peak should belong to one of the main elements of the samples, have no (or very little) interference from other elements and follow a linear relationship with the concentration of its corresponding element.

After normalizing with respect to a reference peak the spectra are negatively biased to the reference element as a higher concentration of this element would result in a stronger intensity of the peak and thus a larger denominator in the normalization. However, this can be corrected by dividing the concentrations with the concentration of the reference element. A problem arises when considering inference of the reference element itself as the relative concentration for this element is constant.

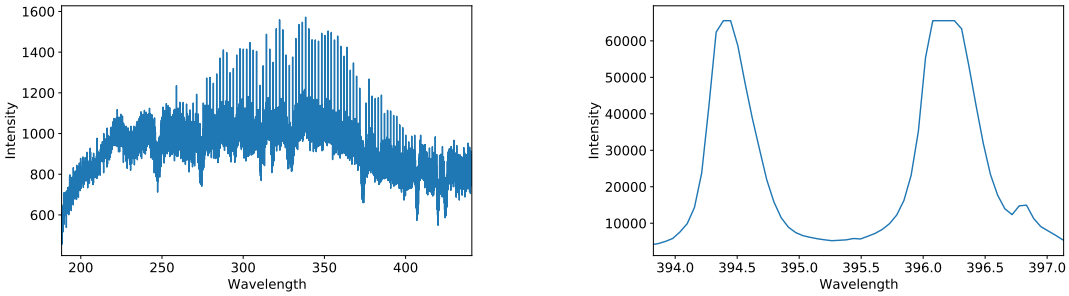

Figure 2: A weak spectrum (left) that should be discarded and part of a too strong spectrum (right) with intensities surpassing the limit of the spectrometer resulting in flat tops.

To get around this, an assumption is made that the P investigated elements are the only ones present in the sample so that the concentrations must sum to one. The concentration of the reference element can then be calculated indirectly by the relative concentrations of the other elements.

$$\sum_{p=1}^{P} x_p = 1,$$

$$x_{ref} = \left(\sum_{p=1}^{P} \tilde{x}_p \right)^{-1}, \tag{1}$$

where x_p is the concentration of element p and $\tilde{x}_p$ is the relative concentration, $\tilde{x}_p = x_p x_{ref}^{-1}$, that is estimated from the regression.

In addition to normalizing the spectra, bad spectra are identified and discarded in the analysis. These are typically spectra from weak measurements where the noise to signal ratio is unfavorable or strong measurements that exceed the maximum signal strength which the spectrometer can register, see figure 2. One method of filtering out bad spectra is to consider the maximum intensity of each spectrum and the intensity of the reference peak. If the maximum intensity is close to the limit of the spectrometer, the spectrum is discarded. Likewise, if the maximum intensity of the reference peak is below some threshold, determined ad hoc, the spectrum is deemed too weak and discarded.

After the spectra of the measurements have been selected and normalized, they are averaged for each sample so that there is one spectrum per sample consisting of the average of the spectra from all measurements of that sample. This accomplishes two objectives: It reduces the number of spectra to be analyzed in upcoming steps and it reduces the impact of heterogeneity and local contamination on the sample. The focal point of the laser is typically 50µm diameter and the measurement is therefore very local. Local variations in the sample composition thus causes the LIBS spectrum to vary. In addition the amount of material that is analyzed in a

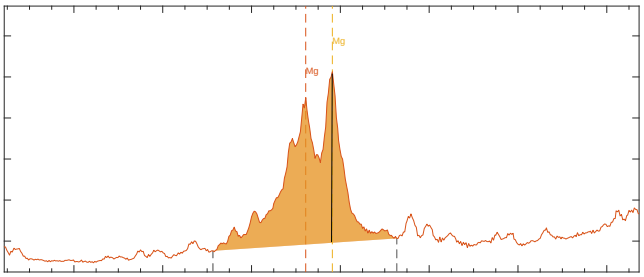

Figure 3: A baseline can be used when the peak does not start from zero either due to overlap with nearby peaks or continuous emission.

measurement is small so even a small contamination can have a large impact on the resulting spectrum.

If there is a large offset under the base of the peak, either due to continuous emission or overlap from neighboring peaks, the analyst can define a baseline for the peak, see figure 3. The peak's intensity is then measured from the baseline instead of from zero.

Once the spectra have been normalized, filtered and averaged; peaks have been selected for each element and their baselines have been corrected; a function can be fitted to link peak intensity and concentration. A good peak of good spectra with good pre-processing should be affine and simple linear regression can be used. However sometimes self-absorption is involved, and a quadratic term is added to the regression. It is not unusual either to not be able to find peaks that are independent of the other elements, so that even after baseline correction there is interference from another element. To alleviate this another term can be added, related to the interfering element. Let $\tilde{y}_p$ and $\tilde{y}_q$ be the pre-processed signal of peaks belonging to the element and the interfering element, respectively, the relative concentration $\tilde{x}_p$ of element p could then be estimated by

$$\hat{\tilde{x}}_p = \hat{\beta}_0 + \tilde{y}_p\hat{\beta}_1 + \tilde{y}_p^2\hat{\beta}_2 + \tilde{y}_q\hat{\beta}_3, \tag{2}$$

where the β are estimated using linear regression on the calibration set.

Low accuracy can be a limiting factor to the adoption of LIBS in industrial settings. One potential way of increasing the accuracy is to use multivariate methods instead of the classical univariate calibration curves. By using more than one peak in the analysis, the impact of noise and interference from other elements can potentially be reduced. An additional benefit of using the entire captured spectrum in the analysis is that the variation between predictions from different analysts, on the same data, can be reduced. Such inter-analyst variations have been identified as an issue in analysis of LIBS spectra and peak selection is a contributing part [5].

Several multivariate statistical methods have been tried, with varying degrees of success, for analyzing LIBS spectra. Two promising methods are partial least squares

(PLS) and principal component regression (PCR) [6]. While both have been shown to perform better than the univariate approach in some cases, results vary and neither have risen to become a standard approach in the analysis of LIBS spectra [6]. A big issue with full-spectra analysis is that the number of features (registered pixels in the spectra) is typically much larger than the available samples. A typical calibration set might have a few dozens of samples and the spectrometer might have thousands of pixels.

Another alternative approach for a quantitative analysis of LIBS spectra is to use calibration free methods, CF-LIBS. In such methods, first principle knowledge of the emitting plasma and databases with various physical properties are combined to estimate the chemical composition from a LIBS spectrum. This was first suggested in 1999 by Ciucci et al. and the method has been further developed thereafter by several researchers [7][8][9].

In CF-LIBS, the integrated peak of species[2] p is assumed to follow the following model:

$$I_p^{rq} = F x_p A_{rq} \frac{g_r e^{-(E_r/k_B T)}}{U_p(T)}, \tag{3}$$

where I_p^{rq} is the integrated line intensity corresponding to the transition from energy state E_r to E_q for the emitting species p, F is a spectrum-wide scaling constant that accounts for experimental factors, x_p is the concentration of the emitting species p, A_{rq} is the transition probability, g_r is the r level degeneracy, k_B is the Boltzmann constant, T is the temperature and $U_p(T)$ is the partition function for the emitting species. The energy states E_r, transition probabilities A_{rq} and the degeneracy of the states g_r, can be found in atomic spectra databases such as the NIST database [10].

To find the concentrations of the sample that generated the spectrum, the analyst selects a couple of good peaks just as in the univariate method. These peaks are then integrated to determine the integrated line intensities I_p^{rq} of their corresponding emission lines. The emission lines are characterized by their transition energy levels, E_r and E_q, and the difference between these determine the wavelength of the emission line. These integrated line intensities are then used to solve for the concentrations by linear regression of the logarithm of equation (3) [7]. The advantage of CF-LIBS is that no matrix-specific[3] calibration is needed to analyze a spectrum.

2 Theory

The two regression methods that are used in this work are multivariate linear regression and a Bayesian method. In this chapter the principles of the these methods are presented in general terms. Their application to the LIBS data follow in chapter 3.

[2] In CF-LIBS it is necessary to separate between not only elements but also ionic charge, hence *species* is used in place of *element*.

[3] Here matrix refers to material type.

2.1 Linear regression

Linear regression is a common tool in both science and industry. It is sometimes taught to students as early as in secondary education and is often thought of in terms of ordinary least squares regression (OLS). In OLS the parameters that minimize the sum of squared residuals are used as estimates for the unknown regression parameters. In this thesis linear regression is performed on a model with multiple dependent variables. Linear regression on multiple dependent variables is called multivariate linear regression, a term that is often mistakenly used for multiple linear regression which involves a single dependent variable but multiple independent variables. This section presents the concept of the *best linear unbiased estimator* and the Gauss-Markov theorem. This will serve as a preamble for the following presentation of multivariate linear regression.

Assume there are N observations of a variable, arranged in the vector $y \in \mathbb{R}^N$, and an additional N observations of P non-random variables, arranged in the matrix $X \in \mathbb{R}^{N \times P}$. Further let X have full rank. Consider the linear model

$$y = X\beta + \epsilon, \tag{4}$$

where $\beta \in \mathbb{R}^P$ is a vector of unknown but fixed parameters and $\epsilon \in \mathbb{R}^N$ is a vector of random errors. Make the following assumptions for the error random variables ϵ_i, $i = 1, ..., N$:

1 The errors ϵ_i are homoscedastic, i.e. $\mathrm{Var}(\epsilon_i) = \sigma^2$ for $i = 1, ..., N$.

2 The errors are uncorrelated, $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$.

Definition 1. A linear estimator of β is an estimator such that:

$$\hat{\beta}_j = \sum_{i=1}^{N} c_{ij} y_i,$$

where the coefficients c_{ij} do not depend on the underlying parameter β_j.

Definition 2. An unbiased estimator has the true parameter β as its expected value,

$$\mathbb{E}[\hat{\beta}] = \beta.$$

Definition 3. A linear estimator $\hat{\beta}$ is the *best linear unbiased estimator* or BLUE if it is an unbiased estimator and for any $\rho \in \mathbb{R}^N$ and any other linear unbiased estimator $\tilde{\beta}$

$$\mathrm{Var}(\rho^T X\hat{\beta}) \leq \mathrm{Var}(\rho^T X\tilde{\beta}).$$

There are many forms of linear regression, however the following theorem states that the OLS estimate is the BLUE for the linear model under the assumptions made above. Recall that the OLS estimator of equation (4) is

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T y. \tag{5}$$

The proof of the theorem follows presentation on the Wikipedia page for the Gauss-Markov theorem with some small adjustments [11].

Theorem 1 (Gauss-Markov theorem)**.** *The ordinary least squares estimator, equation* (5)*, is the BLUE for the linear model, equation* (4)*.*

Proof. Let $\tilde{\beta} = Cy$ be another linear unbiased estimator of β. For some non-zero $D \in \mathbb{R}^{N\times P}$, $C = \left(X^T X\right)^{-1} X^T + D$.

$$\begin{aligned}
\mathbb{E}[\tilde{\beta}] &= \mathbb{E}[Cy] \\
&= \mathbb{E}[\left(\left(X^T X\right)^{-1} X^T + D\right)(X\beta + \epsilon)] \\
&= \left(X^T X\right)^{-1} X^T X\beta + DX\beta + \left(\left(X^T X\right)^{-1} X^T + D\right)\mathbb{E}[\epsilon] \\
&= \left(X^T X\right)^{-1} X^T X\beta + DX\beta \\
&= (I_P + DX)\,\beta
\end{aligned}$$

Since β is unobserved, $\tilde{\beta}$ is unbiased if and only if $DX = 0$.

$$\begin{aligned}
\operatorname{Var}(\tilde{\beta}) &= \operatorname{Var}(Cy) \\
&= C\operatorname{Var}(y)C^T \\
&= \sigma^2 CC^T \\
&= \sigma^2\left(\left(X^T X\right)^{-1} X^T + D\right)\left(\left(X^T X\right)^{-1} X^T + D\right)^T \\
&= \sigma^2\Bigg(\left(X^T X\right)^{-1} X^T X\left(X^T X\right)^{-1} + \\
&\qquad DX\left(X^T X\right)^{-1} + \left(X^T X\right)^{-1} X^T D^T + DD^T\Bigg) \\
&= \sigma^2\left(\left(X^T X\right)^{-1} + \left(X^T X\right)^{-1}(DX)^T + DD^T\right) \\
&= \sigma^2\left(\left(X^T X\right)^{-1} + DD^T\right) \\
&= \operatorname{Var}(\hat{\beta}) + \sigma^2 DD^T
\end{aligned}$$

since DD^T is positive semidefinite,

$$\begin{aligned}
\operatorname{Var}(\rho^T X\tilde{\beta}) &= \rho^T X\operatorname{Var}(\tilde{\beta})X^T\rho \\
&= \rho^T X\left(\operatorname{Var}(\hat{\beta}) + \sigma^2 DD^T\right)X^T\rho \\
&\geq \rho^T X\operatorname{Var}(\hat{\beta})X^T\rho = \operatorname{Var}(\rho^T X\hat{\beta})
\end{aligned}$$

□

R. Christensen presents an even stronger result in *Plane Answers to Complex Questions (2011)*, theorem 10.4.5, which will be included in part here without proof [12].

Theorem 2. *Consider the linear model* (4) *with* $\mathrm{Cov}(\epsilon) = \sigma^2 V$ *for some matrix* $V \in \mathbb{R}^{N\times N}$. *Let* $C(A)$ *denote the column space of a matrix* A, *then the OLS estimate is a BLUE of the linear model if and only if* $C(VX) \subset C(X)$.

A common way to evaluate regression models is the coefficient of determination R^2. First define the sum of squared residuals, SSR, and the total sum of squares, SST:

$$\mathrm{SSR} := \sum_{i=1}^{N} (y_i - \hat{y}_i)^2,$$

$$\mathrm{SST} := \sum_{i=1}^{N} (y_i - \mu_y)^2,$$

where μ_y is the mean of the vector y and $\hat{y}$ is the regression estimate of y. The coefficient of determination is then defined as follows:

$$R^2 := 1 - \frac{\mathrm{SSR}}{\mathrm{SST}} \tag{6}$$

A R^2 value of 1 represents a perfect fit while a value of 0 means that the regression did not fit the data better than just picking the mean. The coefficient of determination can take negative values if the sum of squared residuals is larger than the total sum of squares.

Another common way to evaluate regression models is the root mean squared error, RMSE. It is defined as follows:

$$\mathrm{RMSE} := \sqrt{\frac{\mathrm{SSR}}{N}}$$

Multivariate linear regression

This section presents an estimator for linear models with multiple dependent variables. The presentation largely follows that of R. Christensen in *Linear Models for Multivariate, Time Series, and Spatial Data (1991)* [13].

Let a process have a multivariate response $y \in \mathbb{R}^J$ and let each dependent variable follow a linear relationship with a vector of independent variables $x \in \mathbb{R}^P$ that is shared between the dependent variables.

$$y_j = x \cdot \beta_j + \epsilon_j \quad j = 1, ..., J \tag{7}$$

where $\beta_j \in \mathbb{R}^P$ is a vector of unknown but fixed parameters and $\epsilon_j \sim \mathcal{N}(0, \sigma_j^2)$ is random noise. If N observations are taken of such a process then the relationship in equation (7) can be expressed in matrix form as

$$Y = XB + e, \tag{8}$$

where $Y \in \mathbb{R}^{N\times J}$, $X \in \mathbb{R}^{N\times P}$, $B \in \mathbb{R}^{P\times J}$, $e \in \mathbb{R}^{N\times J}$. Each column of Y contains values of a dependent variable and each column of X contains values of an independent variable.

The columns of Y can be calculated independently of one another as

$$Y_{\cdot j} = XB_{\cdot j} + e_{\cdot j}$$

Hence (8) is equivalent to

$$\begin{bmatrix} Y_{\cdot 1} \\ Y_{\cdot 2} \\ \vdots \\ Y_{\cdot J} \end{bmatrix} = \begin{bmatrix} X & 0 & \dots & 0 \\ 0 & X & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & X \end{bmatrix} \begin{bmatrix} B_{\cdot 1} \\ B_{\cdot 2} \\ \vdots \\ B_{\cdot J} \end{bmatrix} + \begin{bmatrix} e_{\cdot 1} \\ e_{\cdot 2} \\ \vdots \\ e_{\cdot J} \end{bmatrix} \tag{9}$$

Assume that the errors are uncorrelated, and have equal variance, between the observations,

$$\mathrm{Cov}(e_{ih}, e_{i'h'}) = \begin{cases} \sigma_{hh'} & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases}$$

It follows that $\mathrm{Cov}(e_{\cdot h}, e_{\cdot h'}) = \sigma_{hh'} I_N$. Equation (9) is univariate and the error vector has mean zero and covariance matrix:

$$\begin{bmatrix} \sigma_{11} I_N & \sigma_{12} I_N & \dots & \sigma_{1J} I_N \\ \sigma_{12} I_N & \sigma_{22} I_N & \dots & \sigma_{2J} I_N \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1J} I_N & \sigma_{2J} I_N & \dots & \sigma_{JJ} I_N \end{bmatrix} \tag{10}$$

Definition 4. The vectorization transformation, $\mathrm{Vec}(\cdot)$, transforms a matrix into a column vector. Let A be a $m \times n$ matrix then $\mathrm{Vec}(A)$ is a $mn \times 1$ column vector,

$$\mathrm{Vec}(A) = [A_{1,1}, ..., A_{m,1}, A_{1,2}, ..., A_{m,2}, ..., A_{1,n}, ..., A_{m,n}]^T.$$

Definition 5. Let A be a $m \times n$ matrix and B be a $p \times q$ matrix. Then the *Kronecker product* $A \otimes B$ is the $pm \times qn$ block matrix:

$$\begin{bmatrix} A_{1,1}B & \dots & A_{1,n}B \\ \vdots & \ddots & \vdots \\ A_{m,1}B & \dots & A_{m,n}B \end{bmatrix}$$

The following theorem presents selected properties of the Kronecker product. It is included without proof and the interested reader is referred to e.g. *Topics in Matrix Analysis* by Horn and Johnson (1991) [14].

Theorem 3 (Properties of $\otimes$)**.**

- *Mixed-product property:*

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

- *Inverse of a Kronecker product:*
  $(A \otimes B)$ *is invertible if and only if* A *and* B *are invertible and,*

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

- *Transpose of a Kronecker product:*

$$(A \otimes B)^T = A^T \otimes B^T.$$

Using the vectorization transformation and the Kronecker product, rewrite equation (9) as

$$\operatorname{Vec}(Y) = (I_J \otimes X)\operatorname{Vec}(B) + \operatorname{Vec}(e), \tag{11}$$

and the covariance matrix as $\Sigma \otimes I_N$, where $\Sigma_{ij} = \sigma_{ij}$ for $i, j = 1, 2, ..., J$. By theorem 2,

$$\operatorname{Vec}(\hat{B}) = \Big((I_J \otimes X)^T (I_J \otimes X)\Big)^{-1} (I_J \otimes X)\operatorname{Vec}(Y), \tag{12}$$

is the best linear unbiased estimate of $\operatorname{Vec}(B)$ if and only if

$$C((\Sigma \otimes I_N)(I_J \otimes X)) \subset C(I_J \otimes X). \tag{13}$$

By the mixed-product property:

$$\begin{aligned}
(\Sigma \otimes I_N)(I_J \otimes X) =& (\Sigma I_J) \otimes (I_N X) \\
=& \Sigma \otimes X \\
=& \begin{bmatrix} \sigma_{11}X & \dots & \sigma_{1J}X \\ \vdots & & \vdots \\ \sigma_{J1}X & \dots & \sigma_{JJ}X \end{bmatrix} \\
=& \begin{bmatrix} X & 0 & \dots & 0 \\ 0 & X & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & X \end{bmatrix} \begin{bmatrix} \sigma_{11}I_N & \sigma_{12}I_N & \dots & \sigma_{1J}I_N \\ \sigma_{12}I_N & \sigma_{22}I_N & \dots & \sigma_{2J}I_N \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1J}I_N & \sigma_{2J}I_N & \dots & \sigma_{JJ}I_N \end{bmatrix} \\
=& (I_J \otimes X)(\Sigma \otimes I_N)
\end{aligned}$$

For any two matrices R and S of conforming size it holds that $C(RS) \subset C(R)$. Hence equation (13) holds and equation (12) is the BLUE. Using the properties of theorem 3, equation (12) can be reduced back to the original dimensions.

$$\begin{aligned}
\Big((I_J \otimes X)^T (I_J \otimes X)\Big)^{-1} (I_J \otimes X)^T =& \Big((I_J \otimes X^T)(I_J \otimes X)\Big)^{-1} (I_J \otimes X^T) \\
=& (I_J \otimes X^T X)^{-1} (I_J \otimes X^T) \\
=& (I_J \otimes (X^T X)^{-1})(I_J \otimes X^T) \\
=& (I_J \otimes (X^T X)^{-1} X^T)
\end{aligned}$$

Equation (12) is thus equivalent to

$$\text{Vec}(\hat{B}) = \begin{bmatrix} (X^TX)^{-1}X^T & 0 & \dots & 0 \\ 0 & (X^TX)^{-1}X^T & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & (X^TX)^{-1}X^T \end{bmatrix} \begin{bmatrix} Y_{\cdot 1} \\ Y_{\cdot 2} \\ \vdots \\ Y_{\cdot J} \end{bmatrix}$$

which in turn is equivalent to

$$\hat{B} = (X^TX)^{-1}X^TY. \tag{14}$$

2.2 Bayesian analysis

Bayesian and frequentist statistics are the two dominant paradigms of statistics. The key difference between the two is the view of what probability is. While the frequentist reserves the notion of probability for random processes, the Bayesian may use probability to represent uncertainty. The Bayesian can thus assign probability to processes that are not viewed as random but whose outcome is uncertain due to limited knowledge. This allows the analyst to include prior information in the analysis, whether this information is from another experiment, subjective belief or a combination of the two. The prior information is formulated into a prior distribution and is combined with the model using Bayes' theorem. Bayes' theorem and a prior distribution can also be used as a form of regularization.

Let A and B be events and $P(B) \neq 0$. Bayes' theorem then states that,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

When working with continuous random variables this is expressed as

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y)f_X(x)}{f_Y(y)}, \tag{15}$$

where $f_{X|Y=y}(x)$ is referred to as the posterior, $f_{Y|X=x}(y)$ as the likelihood, $f_X(x)$ as the prior and $f_Y(y)$ as the marginal likelihood. The likelihood is derived from the model and the prior is selected by the analyst. The marginal likelihood can be expressed, using the law of total probability, as

$$f_Y(y) = \int f_{Y|X=\xi}(y)f_X(\xi)d\xi. \tag{16}$$

Note that for given data y the marginal likelihood is constant and $f_Y(y)$ can be seen as the normalizing constant that ensures that the posterior integrates to one. Hence the geometry of the posterior distribution is completely determined by the product of the likelihood and the prior. In light of this, it is common not to include the denominator when writing Bayesian models. Instead the posterior is expressed as proportional to the product of the likelihood and the prior,

$$P(A|B) \propto P(B|A)P(A).$$

Since the entire distribution of the parameters is expressed in the posterior distribution, it is relatively easy to tailor an evaluation method to the problem at hand. However if the exact application and requirements of the model is unknown, a well known standard method of evaluation might be the best choice. One such method is equal tailed credible intervals. The α equal tailed credible interval is the interval that a parameter falls in with probability α such that the probability that the parameter is outside of the interval is the same, $\frac{1}{2}(1-\alpha)$, at the lower end and the upper end.

Markov chain Monte Carlo

The numerator in equation (15) is typically straightforward to calculate but the denominator given by equation (16) quickly becomes unfeasible to calculate as the number of parameters increase. One way to get around this is by using Markov chain Monte Carlo (MCMC) methods. A MCMC algorithm creates a Markov chain whose stationary distribution is the sought probability distribution. Under certain conditions this stationary distribution is unique and the Markov chain converges to the stationary distribution. It is then possible to sample from the sought probability distribution by sampling from the converged Markov chain.

A rigorous discussion on the convergence of the Markov chains used in this thesis is out of scope of the thesis. Instead an informal introduction to MCMC methods will be presented. The foundation of this introduction is the detailed balance and its connection to stationary distributions. The following two definitions, theorem and proof are inspired by *Monte Carlo Statistical Methods* by Robert and Casella (2004) [15]. Let k be the transition kernel which describes the dynamics of the Markov process $(X_t, t = 1, 2, \dots)$ via

$$P(X_{t+1} \in B \mid X_t = x) = \int_B k(x, dy), \quad \text{for all } t.$$

Here k is a map such that for each x in a space S $k(x, \cdot)$ is a measure on S in its second argument.

Definition 6. Let S be the state space for a Markov chain defined by its transition kernel k. A probability distribution π, defined on S, is *stationary* if

$$\int_B \pi(dx) = \int_S \int_B k(x, dy)\pi(dx),$$

for all $B \subset S$.

Definition 7. A Markov chain with a transition kernel k satisfies the detailed balance condition if there exists a function f satisfying

$$\int_A \int_B k(x, dy) f(x) dx = \int_B \int_A k(y, dx) f(y) dy,$$

for any subsets A and B in the state space S.

Theorem 4. *Suppose a Markov chain with a transition kernel k satisfies the detailed balance condition with π a probability density function defined on the state space S,*

$$\int_A \int_B k(x, dy)\pi(dx) = \int_B \int_A k(y, dx)\pi(dy), \tag{17}$$

for any subsets A and B in S. Then the density π is a stationary distribution of the Markov chain.

Proof. For any measurable set $B \subset S$,

$$\begin{aligned}\int_S \int_B k(x, dy)\pi(dx) &= \int_B \int_S k(y, dx)\pi(dy) \\ &= \int_B \pi(dy)\end{aligned}$$

□

Many MCMC algorithms utilizes theorem 4 to sample from distributions that are difficult to sample from directly. In some algorithms, such as the Metropolis-Hastings algorithm, the transition is split into two steps, a proposal step g and an acceptance step a [16],

$$P_g(X \in B|x) = \int_B g(x, dy).$$

A transition from state x works out as follows: A new state y is drawn from the proposal distribution $g(x, \cdot)$. This state is then accepted or rejected randomly determined by the acceptance probability $a(x, y)$. If the state is rejected the chain stays at x, if it is accepted it moves to y. This corresponds to the following transition kernel [17]:

$$k(x, dy) = g(x, dy)a(x, y) + \delta_x(dy) \int_S (1 - a(x, u))g(x, du), \tag{18}$$

where δ_x is the Dirac measure,

$$\delta_x(A) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

The choice of proposal and acceptance distributions should be such that the transition kernel, equation (18), fulfils the conditions of theorem 4. To this end, define

$$I(x) := \int_S (1 - a(x, u))g(x, du)),$$

and insert equation (18) into equation (17):

$$\begin{aligned}&\int_A \int_B g(x, dy)a(x, y)\pi(dx) + \int_A \int_B \delta_x(dy)I(x)\pi(dx) = \\ &\int_B \int_A g(y, dx)a(y, x)\pi(dy) + \int_B \int_A \delta_y(dx)I(y)\pi(dy)\end{aligned} \tag{19}$$

By the definition of the Dirac measure, the second terms of the left and right hand side are 0 unless x and y, respectively, are in $A \cap B$. Therefore,

$$\int_A \int_B \delta_x(dy) I(x) \pi(dx) = \int_{A \cap B} I(x) \pi(dx),$$

$$\int_B \int_A \delta_y(dx) I(y) \pi(dy) = \int_{A \cap B} I(y) \pi(dy).$$

Hence the second terms of both sides of equation (19) cancel out and the transition kernel, equation (18), satisfies the conditions of theorem 4 together with the posterior distribution π if and only if,

$$\int_A \int_B g(x, dy) a(x, y) \pi(dx) = \int_B \int_A g(y, dx) a(y, x) \pi(dy),$$

for any subsets A and B in S.

Assume there is a function f such that $\pi(dx) = Cf(x)dx$, for some constant C, and choose the following acceptance probability:

$$a(x, y) = \min\left(1, \frac{f(y)}{f(x)}\right) \tag{20}$$

Further assume that the proposal kernel g is symmetric so that:

$$g(x, dy) h(x, y) dx = g(y, dx) h(x, y) dy, \tag{21}$$

for all functions h. If $f(y) \leq f(x)$ (the other case being similar),

$$\begin{aligned} g(x, dy) a(x, y) \pi(dx) &= g(x, dy) \frac{f(y)}{f(x)} C f(x) dx \\ &= g(y, dx) \underbrace{1}_{a(y,x)} C f(y) dy \\ &= g(x, dy) a(y, x) \pi(dy), \end{aligned}$$

using equation (21) in the second step, therefore:

$$\int_A \int_B g(x, dy) a(x, y) \pi(dx) = \int_B \int_A g(y, dx) a(y, x) \pi(dy)$$

Thus any Markov chain with symmetric proposal distribution g and acceptance distribution given by equation (20) will have the sought distribution π as its stationary distribution, given that there exist a function f satisfying $\pi(dx) = Cf(x)dx$. When the sought distribution is the posterior distribution, the product between the likelihood and the priors are taken as f.

In this thesis, Stan is used for MCMC sampling. More specifically the default sampler of Stan, the *No-U-Turn-Sampler* or NUTS is used. NUTS is a Hamiltonian Monte Carlo (HMC) algorithm and in principle does not need to follow the

structure presented above. However the Stan implementation of HMC uses the acceptance step, equation (20), to counteract errors from numerical integration [18]. Furthermore the proposal process is symmetric and the algorithm thus takes the structure presented above with an especially clever proposal function [19]. A detailed discussion on HMC is not included but can be found in e.g. R. M. Neal 2011 [19].

3 Model

In the univariate method one tries to calibrate a function for each element p, that is usually affine, so that

$$\|\tilde{x}_p - f(\tilde{y}, \beta)\|^2,$$

is minimized. Here $\tilde{x}_p \in \mathbb{R}^N$ is a vector with the relative concentrations of element p and $\tilde{y} \in \mathbb{R}^N$ is a vector with pre-processed intensities of a peak corresponding with element p. Each entry in the vectors $\tilde{x}_p$ and $\tilde{y}$ thus contains a relative concentration or a pre-processed intensity, respectively, of one of the N samples. With an affine function this minimization is accomplished by simple linear regression. Let

$$f(\tilde{y}, \beta_0, \beta_1) := \beta_0 + \tilde{y}\beta_1,$$

using simple linear regression one estimates the $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squared errors between $\tilde{x}_p$ and $f(\tilde{y}, \hat{\beta}_0, \hat{\beta}_1)$. This regression problem corresponds to the model

$$\tilde{x}_p = \beta_0 + \beta_1\tilde{y} + \varepsilon. \tag{22}$$

If one would like to extend this method to take advantage of more information from the spectra then a natural choice would be multiple linear regression. If one would like to use the entire spectrum, i.e. all of the J pixels, then

$$f(\tilde{Y}, \boldsymbol{\beta}) := \tilde{Y}\boldsymbol{\beta}, \tag{23}$$

where $\tilde{Y} \in \mathbb{R}^{N\times J}$ contains the pre-procesed captured spectra and $\boldsymbol{\beta} \in \mathbb{R}^J$ contains the parameters estimated in the regression.

A remark on notation: Hereafter the notation $y(\lambda_j)$ will sometimes be used for different variants of y (different accents, sub- or superscripts). $y(\lambda_j)$ is the signal at pixel j and λ_j is the wavelength corresponding to that pixel. $y(\lambda_j)$ is scalar. The natural association to continuous functions is embraced and the index j will sometimes be dropped to heighten this association. However no attempt is made at modelling how the pixels register the continuum of wavelengths. Instead the modelling is done on the pixel level. There is nothing stopping us however, other than reality, from increasing the number of pixels J to be arbitrarily large and thereby approach the continuum.

The multiple linear regression problem, with the function declared in equation (23), can then be seen as the problem of finding weights β_j so that the weighted sum

$\sum_{j=1}^{J} \tilde{y}(\lambda_j)\beta_j$ maps to the concentrations of the samples as well as possible. This regression corresponds to the model

$$\tilde{x}_p = \sum_{j=1}^{J} \tilde{y}(\lambda_j)\beta_j + \varepsilon \tag{24}$$

While both the univariate and the multivariate models discussed above can be used to map LIBS spectra to sample concentrations; their corresponding models, equations (22) and (24), do not reflect current understanding of LIBS. The linear relationship between concentration and intensity has support from models building on an understanding of the plasma[4] [7][20]. However this form of the model suggests a view of the concentration as random and there is no obvious interpretation of β. A more intuitive univariate model would be

$$\tilde{y}(\lambda) = b_0 + \tilde{x}_p b_1 + \varepsilon. \tag{25}$$

Here $\tilde{x}_p$ is not random, instead the measured intensity $\tilde{y}(\lambda)$ is random as a result of ε, representing noise.

Now consider a non-normalized spectrum at a wavelength corresponding with element p, $y(\lambda_p)$. All other parameters fixed it is expected that $y(\lambda_p)$ follows an affine relationship with the concentration of element p, similar to equation (25). However as discussed in the introduction, large variations between different measurements are expected. Assume that these variations are multiplicative and constant over the the spectrum, the spectrum can then be modelled as.

$$y(\lambda) = h(b_0 + x_p b_1) + \epsilon. \tag{26}$$

where h is the spectrum constant factor, called hit factor, accounting for the measurement to measurement variations. b_0 can now be interpreted as the background signal and b_1 as the expected signal if the concentration was 1 and there was no background signal.

Using this form of the linear relationship also offers a natural way to deal with interference from peaks of other elements. Since emission from different sources with the same wavelengths are added to the signal in the spectrometer it is motivated to view the signal at a pixel as a superposition of signals from different elements. Motivated by the interpretation of b_0 as the background signal and b_1 as the expected signal of the pure element ($x_p = 1$), the notation is adjusted and the following model is suggested

$$y(\lambda) = h y^{bg}(\lambda) + h \sum_{p=1}^{P} y_p^{\mathcal{P}}(\lambda) x_p + \epsilon, \tag{27}$$

where P is the total number of elements, $y^{bg}(\lambda)$ is the background signal at λ and $y_p^{\mathcal{P}}(\lambda)$ is the expected signal of pure element p at λ. Note that unlike the univariate

[4] given that non-linear effects such as self-absorption are negligible

model we do not include second order terms from (2) and leave that out for future work.

The model, equation (27), is however not a linear model when h is allowed to vary. To get around this it is normalized in the same way as in the univariate method, i.e. the spectra are normalized to the reference peak and the concentrations to the concentration of the reference element. Assume the reference peak (pixel) has negligible contribution from elements other than the reference element, negligible background contribution and negligible error,

$$y(\lambda_{ref}) \approx h y^{\mathcal{P}}_{ref}(\lambda_{ref}) x_{ref}. \tag{28}$$

The normalized spectra thus follows,

$$\frac{y(\lambda)}{y(\lambda_{ref})} \approx \frac{y^{bg}(\lambda)}{y^{\mathcal{P}}_{ref}(\lambda_{ref})}\frac{1}{x_{ref}} + \sum_{p=1}^{P} \frac{y^{\mathcal{P}}_p(\lambda)}{y^{\mathcal{P}}_{ref}(\lambda_{ref})}\frac{x_p}{x_{ref}} + \frac{\epsilon}{h y^{\mathcal{P}}_{ref}(\lambda_{ref}) x_{ref}}.$$

Let,

$$\begin{aligned} \tilde{y}(\lambda) := \frac{y(\lambda)}{y(\lambda_{ref})}, \quad \tilde{y}^{bg}(\lambda) := \frac{y^{bg}(\lambda)}{y^{\mathcal{P}}_{ref}(\lambda_{ref})}, \\ \tilde{y}^{\mathcal{P}}_p(\lambda) := \frac{y^{\mathcal{P}}_p(\lambda)}{y^{\mathcal{P}}_{ref}(\lambda_{ref})}, \quad \tilde{x} := \frac{x_p}{x_{ref}}, \quad \tilde{\epsilon} := \frac{\epsilon}{h y^{\mathcal{P}}_{ref}(\lambda_{ref}) x_{ref}}. \end{aligned} \tag{29}$$

The normalized model can then be expressed as,

$$\tilde{y}(\lambda) = \tilde{y}^{bg}(\lambda) x^{-1}_{ref} + \sum_{p=1}^{P} \tilde{y}^{\mathcal{P}}_p(\lambda)\tilde{x}_p + \tilde{\epsilon}. \tag{30}$$

If N samples with concentrations of P elements are taken with a spectrometer capturing J pixels, let $\tilde{X} \in \mathbb{R}^{N\times P}$ contains the relative concentrations of the samples and again let $\tilde{Y} \in \mathbb{R}^{N\times J}$ contain the pre-processed captured spectra. In a similar fashion let $\tilde{Y}^{\mathcal{P}} \in \mathbb{R}^{P\times J}$ contain the normalized pure spectra in its rows, $\tilde{y}^{bg} \in \mathbb{R}^{J}$ contain the background signals at the pixels and let $\tilde{\boldsymbol{\epsilon}}$ contain the errors. The model (30) can be expressed in matrix form as

$$\tilde{Y} = DB + \tilde{\boldsymbol{\epsilon}}, \tag{31}$$

where row n of D is given by,

$$\left[X^{-1}_{n,ref}, \tilde{X}_{n,1},\ \tilde{X}_{n,2}, ..., \tilde{X}_{n,P}\right],$$

and column j of B is given by,

$$\left[\tilde{y}^{bg}_j,\ \tilde{Y}^{\mathcal{P}}_{1,j},\ \tilde{Y}^{\mathcal{P}}_{2,j}, ..., \tilde{Y}^{\mathcal{P}}_{P,j}\right]^T.$$

Given a calibration set $\{X^{cal}, Y^{cal}\}$, the normalized pure spectra $\tilde{Y}^{\mathcal{P}}$ can be estimated using linear regression. For simplicity we assume that the normalized errors are both uncorrelated and have equal variance between samples. Substituting X for D in equation (14), the BLUE for (31) is

$$\hat{B} = (D^T D)^{-1} D^T Y^{cal}. \tag{32}$$

Noting that

$$\hat{B}_{\cdot,j} = (D^T D)^{-1} D^T Y^{cal}_{\cdot j}, \tag{33}$$

i.e. the estimated signals of the normalized pure spectra and background signal at a pixel from the multivariate regression is the same as those estimated from a multiple regression of the data at that pixel. Hence one shouldn't hope for any better results, at a given pixel, than from multiple linear regression and since a typical LIBS calibration set contains a couple of dozen samples and around a dozen elements one should expect overfitting. In addition to a poor samples-to-variables ratio, it is quite likely that there is some correlation in the concentration data as samples typically come from, or approximates samples from, some industrial process.

Now assume that $\hat{\tilde{y}}^{bg}$ and $\hat{\tilde{Y}}^{\mathcal{P}}$ have been estimated and that we want to determine the concentration x given a normalized LIBS spectrum $\tilde{y}^{obs}$. In the case with a single variable, equation (25), the relative concentration can be estimated by solving the equation presented by the model when ignoring the error term, i.e. dividing the measured signal with the expected signal of the pure element after subtracting the background contribution. However it is unlikely that an exact solution is available in the multivariate case so a least-squares approach is adopted instead,

$$\underset{\hat{d}\in\mathbb{R}^{P+1}}{\text{minimize}} \, \|\tilde{y}^{obs} - B^T \hat{d}\|^2, \tag{34}$$

where

$$\hat{d} = \left[\hat{x}_{ref}^{-1}, \hat{\tilde{x}}_1, \hat{\tilde{x}}_2, ..., \hat{\tilde{x}}_P\right]^T.$$

The concentrations are then calculated by inverting the first regression coefficient ($\hat{x}_{ref}^{-1}$) and multiplying the relative concentrations with this value.

The suggested full spectra extension of the univariate calibration method is thus a multivariate calibration method where the normalized pure spectra $\tilde{Y}^{\mathcal{P}}_{1,\cdot}$, $\tilde{Y}^{\mathcal{P}}_{2,\cdot}, ..., \tilde{Y}^{\mathcal{P}}_{P,\cdot}$ and the background spectrum $\tilde{y}^{bg}$ are first estimated from the calibration set whereafter predictions are made by solving the model, with the error term ignored, for the relative concentrations $\tilde{x}$. Since an exact solution is unlikely in the multivariate case, the estimated concentrations are solved for in a least-squares sense.

3.1 Bayesian approach

As mentioned, overfitting is expected at any given pixel for a typical LIBS calibration set. Ideally the issues from this overfitting would cancel out over the large number of pixels so that the final concentration estimate would be accurate and robust.

However when trying the method on the available datasets this was not found to be the case.

To alleviate the problem with overfitting a Bayesian version of the model was constructed. The idea being that with additional information, relayed with the priors, the overfitting could be reduced enough that over a large number of pixels, accurate and robust estimates could be obtained. Priors as well as probability distributions for the errors ϵ are needed to create a Bayesian model. The errors are assumed to be normally distributed and independent over both samples and pixels so that for pixel j of a pre-processed spectrum $\tilde{y}$,

$$\tilde{y}_j \sim \mathcal{N}(\tilde{y}_j^{bg} x_{ref}^{-1} + \tilde{Y}_{\cdot j}^{\mathcal{P}} \cdot \tilde{x}, \sigma_j^2). \tag{35}$$

This equation formulates a forward model for the observable pre-processed spectrum in terms of the unknown normalized background spectrum $\tilde{y}_j^{bg}$, the unknown normalized pure spectra $\tilde{Y}_{\cdot j}^{\mathcal{P}}$ and the known or unknown x (both x_{ref}^{-1} and $\tilde{x}$ are determined from the concentration x). In the case of the calibration x is known and in the case of the prediction $\tilde{y}_j^{bg}$ and $\tilde{Y}_{\cdot j}^{\mathcal{P}}$ are the results from the calibration step while x is the parameter to estimate.

We first consider a two-step procedure where the $\tilde{y}_j^{bg}$ and $\tilde{Y}_{\cdot j}^{\mathcal{P}}$ are estimated in the calibration step and thereafter the x is estimated, using the estimated $\tilde{y}_j^{bg}$ and $\tilde{Y}_{\cdot j}^{\mathcal{P}}$ from the calibration step, in the prediction step.

Here one has the choice of plugging in a point estimator for equation (35) or one write the joint posterior of the model. In the calibration step when x is known, the joint posterior of everything unknown given everything observed, based on equation (35), is expressed as:

$$P(\tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma | \tilde{Y}^{cal}) \propto P(\tilde{Y}^{cal} | \tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma) P(\tilde{y}^{bg}) P(\tilde{Y}^{\mathcal{P}}) P(\Sigma). \tag{36}$$

Similarly the prediction step is expressed as:

$$P(x, \tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma | \tilde{y}^{obs}, \tilde{Y}^{cal}) \propto P(\tilde{y}^{obs} | x, \tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma) P(x) \times P(\tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma | \tilde{Y}^{cal}). \tag{37}$$

Σ is the covariance matrix and with the assumptions of equation (35) it takes the following form:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & \sigma_P^2 \end{bmatrix}$$

The Bayesian model can be seen as a two-step process, first the parameters are estimated on the calibration set and then the prediction is performed on new data

using the posterior of the calibration step as one of its priors. However the two steps can also be combined into a one-step process and this one-step process is easier to implement in Stan. As a one-step process the model is written as:

$$P(x, \tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma|\tilde{y}^{obs}, \tilde{Y}^{cal}) \propto P(\tilde{y}^{obs}|x, \tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma)P(x)\times \\ P(\tilde{Y}^{cal}|\tilde{y}^{bg}, \tilde{Y}^{\mathcal{P}}, \Sigma)P(\tilde{y}^{bg})P(\tilde{Y}^{\mathcal{P}})P(\Sigma). \quad (38)$$

An example will help illustrating the equivalence between the two-step process, equations (36) and (37), and the one-step process with its rather daunting expression, equation (38).

Example Consider some forward model for y in terms of x and θ. Given two sets of observations, the first set with observations of both y and x and the second with observations of only y, we want to estimate the unknowns θ as well as the x giving rise to the second set of observations of y. To clarify, we have a forward model with probability distribution,

$$f(y|x, \theta),$$

two sets of observations, (x_1, y_1) and y_2; from this we want to determine the unknowns x_2 and θ. Let h be the distribution for the unknown variables conditional on the known variables,

$$h(x_2, \theta|x_1, y_1, y_2),$$

and let g be the joint distribution for the observations y_1 and y_2,

$$g(y_1, y_2|x_1, x_2, \theta).$$

Using Bayes' theorem on the distribution, h, for the unknown variables conditional on the known variables,

$$h(x_2, \theta|x_1, y_1, y_2) = \frac{g(y_1, y_2|x_1, x_2, \theta)\pi_\theta(\theta)\pi_x(x_2)}{c_1}, \quad (39)$$

where c_1 is the normalizing constant, π_θ and π_x are priors for θ and x. The observations y_1 and y_2 are independent, conditional on θ, of one another as well as the x of the other observation set, i.e.

$$g(y_1, y_2|x_1, x_2, \theta) = f(y_1|x_1, \theta)f(y_2|x_2, \theta).$$

Substituting into equation (39) and rearranging,

$$\frac{g(y_1, y_2|x_1, x_2, \theta)\pi_\theta(\theta)\pi_x(x_2)}{c_1} = \frac{f(y_2|x_2, \theta)\pi_x(x_2)f(y_1|x_1, \theta)\pi_\theta(\theta)}{c_1},$$

we identify the one-step process, equation (38). To get the corresponding two-step process, note that the corresponding calibration step is

$$q(\theta|x_1, y_1) = \frac{f(y_1|x_1, \theta)\pi_\theta(\theta)}{c_2}, \tag{40}$$

where c_2 is the normalizing constant. Putting it all together,

$$\begin{aligned} h(x_2, \theta|x_1, y_1, y_2) &= \frac{g(y_1, y_2|x_1, x_2, \theta)\pi_\theta(\theta)\pi_x(x_2)}{c_1} \\ &= \frac{f(y_2|x_2, \theta)\pi_x(x_2)f(y_1|x_1, \theta)\pi_\theta(\theta)}{c_1} \\ &= \frac{f(y_2|x_2, \theta)\pi_x(x_2)\left(f(y_1|x_1, \theta)\pi_\theta(\theta)c_2^{-1}\right)}{c_1 c_2^{-1}} \\ &= \frac{f(y_2|x_2, \theta)\pi_x(x_2)q(\theta|x_1, y_1)}{c_1 c_2^{-1}}. \end{aligned}$$

We identify this as the prediction step, equation (37), and together with the calibration step, equation (40), the two-step process.

□

Noting that the two-step process, equations (36) and (37), and the one-step process, equation (38), are equivalent but that the two-step process is easier to understand while the one-step process is easier to implement in Stan; the model is explained in terms of the two-step process and implemented in Stan as the one-step process.

In the prediction step the priors consists of the posterior from the calibration step, equation (36), and a Dirichlet distribution for the concentrations x. In the calibration step, the priors for $\tilde{y}^{bg}$, $\tilde{Y}_j^{\mathcal{P}}$ and σ_j consists of Gaussian-, exponential- and Cauchy distributions, respectively. The Cauchy distribution for the variances is clipped at 0 to prevent undefined negative variances.

For the parameters of the priors of the pure spectra, equation (3) was taken as the foundation together with data from the NIST atomic spectra database [10]. Values for g_r, E_r and A_{rq} were taken from the NIST database for all emission lines of all elements present in the samples. Together with chosen values for F and T, the integrated line intensities could be calculated using equation (3). The integrated line intensities were then converted to pixel intensities. To do this a peak was constructed from each integrated line intensity by multiplying it with Gaussian

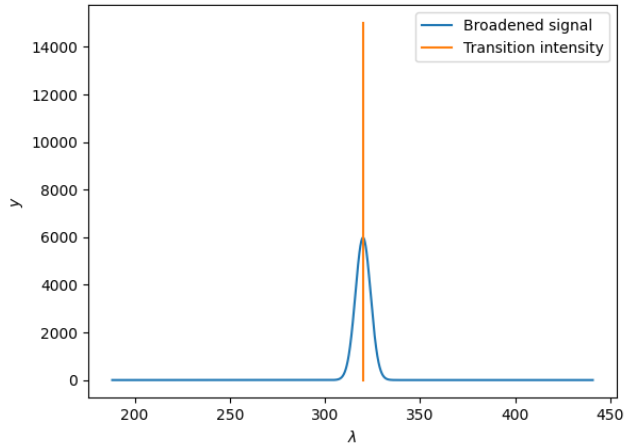

Figure 4: Multiplying the integrated line intensity, or the transition intensity, with a Gaussian distribution approximates the broadening of the signal that naturally occurs.

distributions, see figure 4. The Gaussians were centered at the wavelength of their corresponding emission line. All peaks belonging to a given element were then added together to create the prior pure spectra $y_p^{\mathcal{P}_0}(\lambda)$ of that element. For simplicity all Gaussians had the same variance B referred to as the broadening factor.

The prior pure spectra were then normalized by dividing their intensities with the intensity of the reference element (Al) at the reference pixel. This normalization step cancels out the F factor in equation (3) so that only T and B remains to be determined. Plug-in estimates were used for both of them. The normalized prior pure spectrum of Si is shown in figure 5.

The normalized prior pure spectra contains the information that will be used to attempt to overcome the overfitting issue. The exponential priors for the pure spectra are the means to relay this information to the model. The rate of the exponential distributions was set proportional to the inverse of the prior pure spectra. This choice highly constrains the pure spectra that are not expected to have any signal at a given pixel while leaving the pure spectra that are expected to have signal at the pixel relatively free, see figure 6. It also forces the pure spectra to be positive, which is in line with the interpretation of the parameters $Y^{\mathcal{P}}$ as the expected spectra of the pure elements.

Using the inverse of the prior pure spectra inevitably causes problems in practice when a prior pure spectra is close to 0 at a pixel. For example, at the pixel corresponding to $\lambda = 404.58$, the prior pure spectrum of Zn is rounded to 0 by the computer, see table 1, and its inverse is therefore undefined. To get around this the prior pure spectra is simply clipped at a small value, e.g. 10^{-9}, so that any value smaller than 10^{-9} is replaced by 10^{-9}. This is in line with the general idea of the taken approach. The main objective of the priors is to reduce overfitting and this is

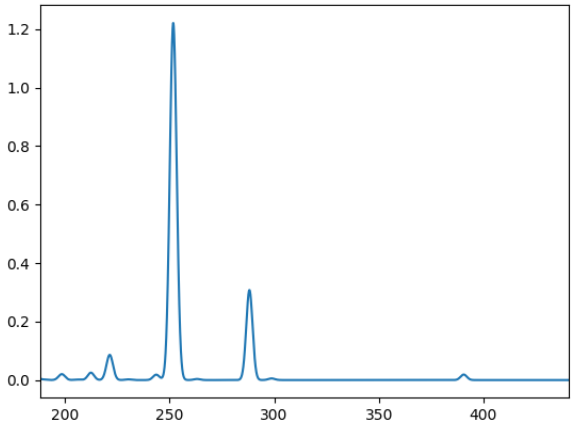

Figure 5: Normalized prior pure spectrum of Si calculated with $T = 4800$ and $B = 1.5$.

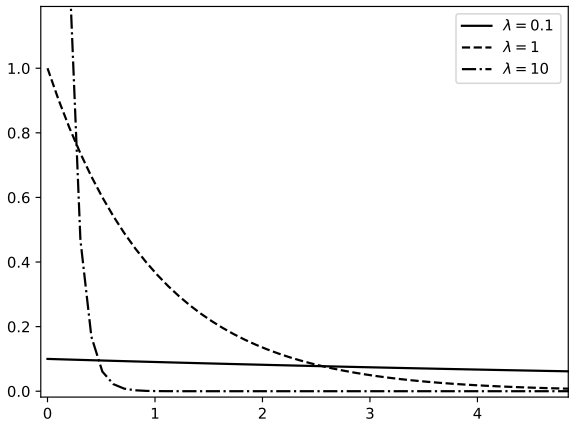

Figure 6: Exponential distributions for rates 0.1, 1 and 10. Larger rates have very small probability mass for large values and will thus constrain the pure spectra with small (large inverse) prior pure spectra from taking large values in the regression.

Table 1: Prior pure spectra at $\lambda = 404.58$.

Si	$8.85 \cdot 10^{-8}$
Fe	$6.79 \cdot 10^{-3}$
Mg	$5.15 \cdot 10^{-6}$
Ni	$2.13 \cdot 10^{-6}$
Zn	0
Cu	$8.07 \cdot 10^{-5}$
Mn	1.42
Cr	$5.81 \cdot 10^{-4}$

accomplished by constraining the pure spectra from taking unreasonable values.

For the calibration step, it remains to address the parameters of the priors for the background signal y^{bg} and for the variances σ_j. These are taken to be only weakly informative.

4 Result

To evaluate the models and gain insight into their strengths and limitations they are applied to data from aluminum samples. The data consists of around 40,000 spectra taken on 27 samples. Each spectrum consist of signals at 4094 pixels. The spectra are filtered, removing any spectra with too high (60,000) maximum signal or too low (5000) signal at the reference pixel $\lambda_{ref} = 308.23$nm.

After the weak and strong spectra have been removed, the spectra are normalized. For every spectrum the signal of each pixel is divided with the signal of the reference pixel. The concentrations are likewise normalized by dividing the concentrations of each sample with its aluminum concentration. Finally, the spectra are grouped by their samples and averaged so that there is one spectrum per sample. The data now consist of 27 normalized spectra, each paired with relative concentrations of 12 elements as well as the inverse of the concentration of the reference element Al.

To evaluate the multivariate linear regression model, leave-one-out cross validation is performed. For each sample, the remaining 26 samples are used as the calibration set. The samples have relative concentrations of 12 elements: Si, Fe, Mg, Ti, Ni, Zn, Cu, Mn, Pb, Sn, Cr and Al. However concentrations of Ti, Sn, and Pb are low and to reduce the degrees of freedom these are dropped from the model. The calibration set, (X^{cal}, Y^{cal}), is used to estimate the normalized background signal $\tilde{y}^{bg}$ and the normalized pure spectra $\tilde{Y}^{\mathcal{P}}$ using equation (32). The relative concentration of the left out sample is then estimated according to equation (34). The solution to this minimization problem is the well known ordinary least squares estimate.

When using normalized spectra and relative concentrations in calculations, the resulting estimates are relative concentrations. These are transformed into concentra-

Table 2: Coefficents of determination, R^2, and root mean squared errors, RMSE, for the univariate- and multivariate method. The univariate method is evaluated on the same data as used in the regression while the multivariate method is evaluated using leave-one-out cross-validation. Note that this is not an entirely fair comparison because we are comparing with a univariate model with second terms and hand-selected baseline correction.

	R^2 Univariate	R^2 MV	RMSE Univariate	RMSE MV
Si	0.960	-0.065	$5.48 \cdot 10^{-3}$	$4.37 \cdot 10^{-2}$
Fe	0.473	-0.840	$2.29 \cdot 10^{-3}$	$4.25 \cdot 10^{-3}$
Mg	0.988	0.943	$1.33 \cdot 10^{-3}$	$2.90 \cdot 10^{-3}$
Ni		-0.399		$3.60 \cdot 10^{-3}$
Zn	0.992	0.105	$1.23 \cdot 10^{-3}$	$1.93 \cdot 10^{-2}$
Cu	0.984	0.760	$1.47 \cdot 10^{-3}$	$6.46 \cdot 10^{-3}$
Mn	0.927	0.788	$1.09 \cdot 10^{-3}$	$1.92 \cdot 10^{-3}$
Cr		0.572		$6.93 \cdot 10^{-4}$
Al		-0.599		$5.51 \cdot 10^{-2}$

tions and the estimated concentrations are then used to calculate individual coefficients of determination and root mean squared errors for each element, these can be found in table 2 together with coefficients of determination from estimates by the univariate method.

The Bayesian model is evaluated by comparing approximated 90% equal tailed credible intervals for the concentrations with the actual concentrations, see figure 8. The credible intervals are computed from samples of the posterior distribution obtained through the MCMC method implemented in Stan. Since the evaluation method is visual and since the MCMC sampling takes far longer to compute, 7 random aluminum samples are used instead of using all aluminum samples as in leave-one-out cross validation. For each aluminum sample, the posterior is estimated using the data from the other 26 aluminum samples as calibration data. The same preprocessing as for the multivariate linear regression is used but Ti, Sn and Pb are included in the analysis. However, due to computational constraints, all pixels could not be included. Instead 52 selected pixels and their two neighbor-pixels were used.

Samples from the posterior, equation (38), was drawn using the default Stan settings but with the number of iterations increased to 3000. The parameters used for the priors can be found in table 3 and in figure 7. Stan did not print any warnings and no further convergence checks were made.

5 Discussion

This thesis presents a shift in perspective. Instead of attempting multivariate analysis of LIBS data through calibration of a multivariate mapping from the spectrum to the concentrations; the spectrum is modelled as a multivariate response to a

Table 3: Prior distributions used in evaluation of the Bayesian model. x is transformed into relative concentration $\tilde{x}$ according to equation (1).

y^{bg}	Gaussian distribution with mean 0 and standard deviation 0.5.
$Y^{\mathcal{P}}$	Exponential distribution with rate $0.1 \times \left(Y^{\mathcal{P}}\right)^{-1}$, $Y^{\mathcal{P}}$ clipped at 10^{-6}.
x	Dirichlet distribution with parameter $a_p = 1$ for all elements except Al, 99 for Al.
Σ	Cauchy distribution with mean 0.1 and scale 0.4, clipped at 0.

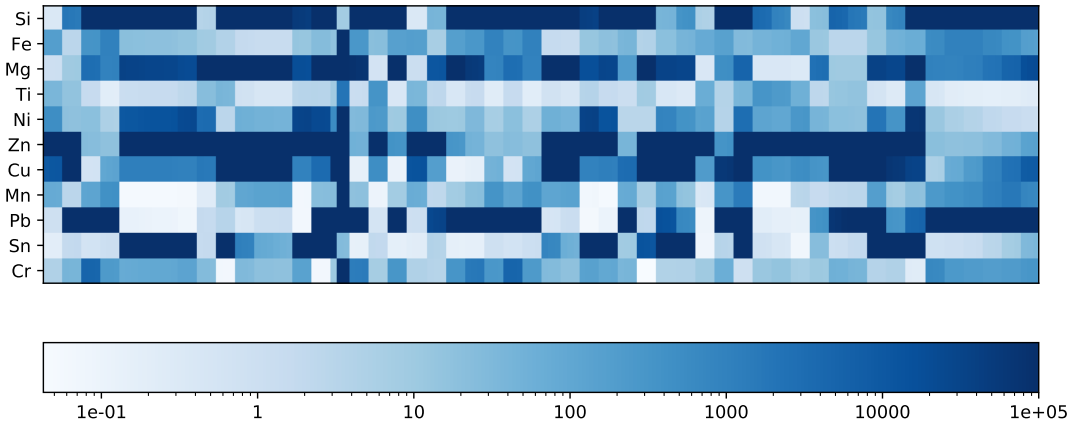

Figure 7: Rate parameters for the exponential distributions used as priors for the pure spectra. Each column shows the rate parameters used for one of the selected pixels. The rate parameters are one tenth of the inverse of the clipped prior pure spectra.

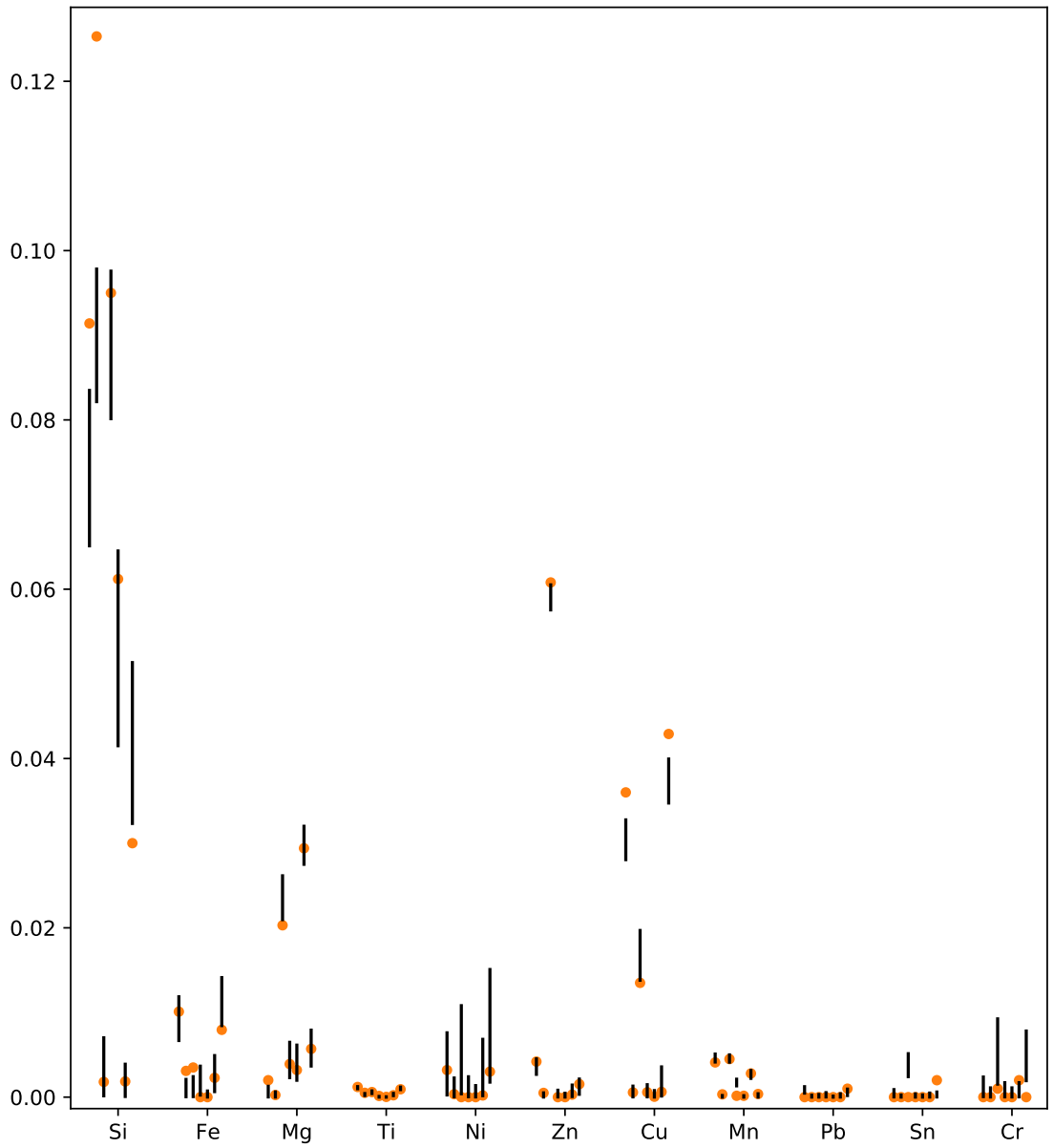

Figure 8: Equal tailed credible intervals for the concentrations predicted by the Bayesian model (lines) and actual concentrations (orange dots) for seven randomly selected samples. The samples were removed one by one when evaluating the Bayesian model, i.e. the predicted sample is not in the calibration set. Aluminum is not included for illustrative purposes (concentration is close to 0.9).

process and finding the concentrations is the inverse problem to this model. This perspective has two key advantages.

Firstly it is a model based approach. This makes results more interpretable, assumptions easier to scrutinize and further development easier. Take the normalization as an example. The reference peak method has been used both in this thesis and in the classical univariate LIBS analysis [4]. The model based approach taken here made it possible to derive this normalization method and in the process be very clear about the assumptions being made.

Indeed, the normalization method presents an example on how the model based approach can help scrutinizing the made assumptions. To claim that equation (32) is BLUE for the model, equation (31), the normalized errors were assumed to have equal variance between the samples. However looking at the definition for the normalized error, equation (29), this hardly seems reasonable. The errors of the non-normalized spectra would somehow need to correlate with the signal at the reference pixel to compensate for the scaling resulting from the normalization. A multivariate weighted linear regression estimator would likely be a better choice of estimator than the one used, equation (14), when analyzing the normalized spectra. Another approach could be to use a different transformation than normalization to deal with the varying measurement to measurement intensity. Using the same assumptions, equation (28), h can be expressed as,

$$h \approx \frac{y(\lambda_{ref})}{y^p_{ref}(\lambda_{ref})x_{ref}},$$

so the model, equation (27), can be transformed into,

$$\begin{aligned} y(\lambda) =& h y^{bg}(\lambda) + h \sum_{p=1}^{P} y_p^{\mathcal{P}}(\lambda) x_p + \epsilon \\ \approx& \frac{y^{bg}(\lambda) y(\lambda_{ref})}{y^p_{ref}(\lambda_{ref}) x_{ref}} + \sum_{p=1}^{P} \frac{y(\lambda_{ref})}{y^p_{ref}(\lambda_{ref})} \frac{x_p}{x_{ref}} y_p^{\mathcal{P}}(\lambda) + \epsilon \\ =& y(\lambda_{ref}) \tilde{y}^{bg}(\lambda) x^{-1}_{ref} + y(\lambda_{ref}) \sum_{p=1}^{P} \tilde{y}_p^{\mathcal{P}}(\lambda) \tilde{x}_p + \epsilon, \end{aligned}$$

making the assumption of equal variance between the samples more reasonable.

Another way that the approach facilitate further improvements is through sensitivity analysis. The model, equation (27), presents a highly simplified view of LIBS spectra and contain errors. Indeed comparing the multivariate model with the more detailed CF-LIBS model, equation (3), coupled with broadening; we note that variability between samples in either temperature T or amount of broadening would not be accounted for. The sensitivity to these effects could be investigated by simulating spectra from equation (3) and applying the method to these. This can of course be done with methods such as multiple linear regression too, but the increased interpretabilty stemming from the model based approach, clearly stated assumptions

and the clear interpretation of the parameters arguably makes it easier to identify error sources to investigate.

A careful sensitivity analysis of the model was deemed out of scope for this thesis but would likely be highly valuable in the pursuit of a good method for multivariate analysis of LIBS spectra. The effect of known non-linear effects such as variation in plasma temperature, broadening and self-absorption could be investigated by applying the method to computer generated spectra. The insights from this analysis could potentially be used to improve the model or to develop criteria for the use of the model. These criteria could then replace or complement the filtering method described in the introduction and hopefully improve performance.

The second key advantage of the taken approach is that it decouples degrees of freedom and number of pixels. Mapping the spectrum to concentration using methods like multiple linear regression means that the degrees of freedom are linked to the number of pixels. This leads to the counter intuitive behaviour that an increase in resolution of the spectrometer likely results in worse predictions. Indeed it is unlikely that the majority of the pixels could be used in the first place as the number of samples are likely counted in the dozens while the number of pixels are counted in thousands. The approach taken in this thesis eliminates this problem. Only the number of elements in the analysis counts towards the degrees of freedom. This makes the success of the analysis more feasible. More pixels also correspond to more data in the prediction step and thus hopefully better predictions.

Even with this improvement, the ratio of samples to degrees of freedom was poor. In an attempt to reduce this issue the Bayesian model was created. Both approaches show promise as potential tools for quantitative analysis of LIBS spectra with many or all pixels. However neither perform as well as the univariate method, see table 2 and figure 8, and further improvement is needed. Fortunately it is much easier to identify potential improvements, as argued above, compared to an approach based on e.g. PLS. The models could thus be systematically investigated and hopefully improved to the degree that they constitute an improvement to the univariate method.

A noteworthy idea is to exploit the sparsity of the spectra. Not every element emits on each frequency - on the contrary, only a few peeks are important, see figure 7. This makes the problem of finding the pure spectra a sparse regression problem. In regression respective the Bayesian paradigm can incorporate such sparsity constrains naturally through a penalty terms or the choice of prior. Note that in this situation very good prior information about the frequency response of the elements exist, for example the competing approach CF-LIBS is purely based on prior information and it is conceivable that combining prior information of CF-LIBS with a non-parametric Bayesian statistical approach. This seems to be a subject to future research.

Another related idea is to weight the pixels in the prediction step, equation (34). One could be fooled to identify this as a conventional linear regression problem and

proclaim that a weighted linear regression would be BLUE. However that would be a mistake as B is random. Which weights to use is thus not straightforward and left for future work; however some things to consider are signal strength at the pixel, uncertainty in the estimates of the pure spectra at the pixel and amount of interference between the elements at the pixel.

6 Conclusion

Neither of the two approaches outperform the univariate method and are in need of further improvement. The positive aspects are however that both approaches present a model based approach to multivariate quantitative LIBS analysis and with their clearly stated assumptions, they allow for systematic development. Perhaps most importantly the approaches shift the degrees of freedom from being dependent on the number of pixels to being dependent on the number of elements. This is a two orders of magnitude improvement. The thesis can thus be seen as a first step in a new direction for multivariate quantitative LIBS analysis. Hopefully with further steps, this leads to increased performance so that LIBS can be used for in-situ measurements; enabling decreased energy and material consumption, better quality and usable biproducts.

Bibliography

[1] R. Noll et al. “Laser-induced breakdown spectroscopy expands into industrial applications”. In: *Spectrochim. Acta B* 93 (2014), pp. 41–51.

[2] R. Noll et al. “LIBS analyses for industrial applications – an overview of developments from 2014 to 2018”. In: *J. Anal. At. Spectrom.* 33 (2018), pp. 945–956.

[3] S. Moncayo et al. “Classification of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS)”. In: *Talanta* 158 (2016), pp. 185–191. DOI: https://doi.org/10.1016/j.talanta.2016.05.059.

[4] Arne Bengtson. “Laser Induced Breakdown Spectroscopy compared with conventional plasma optical emission techniques for the analysis of metals – A review of applications and analytical performance”. In: *Spectrochimica Acta Part B* 134 (2017), pp. 123–132. DOI: https://doi.org/10.1016/j.sab.2017.05.006.

[5] V. Motto-Ros et al. “Critical aspects of data analysis for quantification in laser-induced breakdown spectroscopy”. In: *Spectrochimica Acta Part B: Atomic Spectroscopy* 140 (2018), pp. 54–64. DOI: https://doi.org/10.1016/j.sab.2017.12.004.

[6] Tomoko Takahashia and Blair Thornton. “Quantitative methods for compensation of matrix effects and self-absorption in Laser Induced Breakdown Spectroscopy signals of solids”. In: *Spectrochimica Acta Part B* 138 (2017), pp. 31–42. DOI: http://dx.doi.org/10.1016/j.sab.2017.09.010.

[7] A. CIUCCI et al. “New Procedure for Quantitative Elemental Analysis by Laser-Induced Plasma Spectroscopy”. In: *Applied Spectroscopy* 53.8 (1999), pp. 960–964.

[8] P.J. Kolmhofer et al. “Calibration-free analysis of steel slag by laser-induced breakdown spectroscopy with combined UV and VIS spectra”. In: *Spectrochimica Acta Part B* 106 (2015), pp. 67–74.

[9] E. Tognoni et al. “Calibration-Free Laser-Induced Breakdown Spectroscopy: State of the art”. In: *Spectrochimica Acta Part B* 65 (2010), pp. 1–14.

[10] A. Kramida et al. *NIST Atomic Spectra Database (version 5.7.1), [Online].* DOI: https://doi.org/10.18434/T4W30F. URL: https://physics.nist.gov/asd. (accessed: 2020-07-14).

[11] Wikipedia contributors. *Gauss–Markov theorem — Wikipedia, The Free Encyclopedia*. [Online; accessed 29-October-2020]. 2020. URL: https://en.wikipedia.org/w/index.php?title=Gauss%E2%80%93Markov_theorem&oldid=976302150.

[12] Ronald Christensen. *Plane Answers to Complex Questions*. Springer, 2011.

[13] Ronald Christensen. *Linear Models for Multivariate, Time Series, and Spatial Data*. Springer, 1991.

[14] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991. DOI: 10.1017/CBO9780511840371.

[15] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. 2nd edition. Springer, 2004.

[16] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57 (1970), pp. 97–109.

[17] Luke Tierney. "A note on Metropolis-Hastings kernels for general state spaces". In: *The annals of Applied Probability* 8 (1998), pp. 1–9.

[18] Stan Development Team. *Stan Reference Manual (Version 2.24)*. URL: https://mc-stan.org/docs/2_24/reference-manual/index.html. (accessed: 2020-09-11).

[19] Radford M. Neal. "Handbook of Markov Chain Monte Carlo (Chapter 5)". In: Chapman and Hall/CRC, 2011. Chap. 5.

[20] Reinhard Noll. "Fundamental Algorithms". In: Springer, 2012. Chap. 10.