CHALMERS

Multivariate signal approaches for object detection using microwave sensing

Master's Thesis in Engineering Mathematics and Computational Science

Claes Andersson

Department of Signals & Systems
Signal Processing & Biomedical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2013
Master's Thesis EX021/2013

Abstract

Microwave sensing can be implemented to detect objects in material flowing through a pipe by detecting changes in the signal. In this work simulated data modelling this scenario is used to examine the possibilities to separate changes caused by air bubbles from changes caused by plastic or metal objects.

The data used is obtained from numerically solving Maxwell's equations by FEM for an intersection of a cylindrical cavity at which four waveguides are placed. The solutions consists of the values of the scatter parameters at 400 frequencies in the microwave range. Data, modelling homogeneous background material with and without objects of the mentioned types, is analyzed and used to evaluate five types of classifiers: k nearest neighbors (KNN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM) and libSVM – a support vector machine algorithm for multilabel classification. It is also investigated which parts of the frequency span and what scatter matrix elements are most useful for classification. Furthermore the effects of adding measurement noise, modelled by white noise, is examined. It is found that in the case of no noise the lowest part of the frequency span is most useful for detecting objects, but also that the effects of adding noise is greatest at these frequencies. Concerning the classifiers, the LDA, QDA and libSVM classifiers all perform well in the absence of noise. While adding noise greatly reduces the accuracy of the discriminant analysis classifiers, the libSVM classifier shows promising results at SNR levels above 30 dB. At SNR levels from 20 dB and below non of the classifiers yields an accuracy that would meet the requirements of att useful detector in practice.

Finally, data modelling inhomogeneous background material is examined. In the initial analysis some none linear correlations was found, which led to evaluation of the effects of transforming the data to remove linear or linear and quadratic correlations. The evaluation method is a simple signal detection algorithm, and the results point to some improvement, although very small.

Acknowledgments

I would like to thank my supervisors, Tomas McKelvey and Thomas Rylander, for their help in the work with this thesis. I would also like to thank Johan Winges for generating and explaining the data used in this work, and for ideas and inputs on the work in general.

Claes Andersson, Göteborg, August 12, 2013

Contents

1 Introduction **1**
- 1.1 Aim . . . 2
- 1.2 Thesis outline . . . 2

2 Theory **3**
- 2.1 Classifiers . . . 3
  - 2.1.1 KNN . . . 3
  - 2.1.2 Discriminant analysis . . . 3
  - 2.1.3 Support vector machines . . . 5

3 Method **7**
- 3.1 The data . . . 8
  - 3.1.1 Modelling homogeneous background and objects . . . 8
  - 3.1.2 Modelling stochastically varying background and objects . . . 9
- 3.2 Definitions . . . 9
  - 3.2.1 The notion of signal . . . 10
  - 3.2.2 Adding noise to the simulated data . . . 10
- 3.3 Classifiers and feature selection . . . 11
  - 3.3.1 Implementation of classifiers . . . 11
- 3.4 Manual feature selection . . . 12
  - 3.4.1 Choosing scattering matrix elements . . . 12
  - 3.4.2 Using only a frequency span of fixed length . . . 12
  - 3.4.3 Using equally spaced frequencies . . . 12
  - 3.4.4 Using a part of the frequency span of varying length . . . 12

4 Results **14**
- 4.1 Data analysis . . . 14
  - 4.1.1 Data with fixed background parameters . . . 14
  - 4.1.2 Data with stochastically varying background parameters . . . 16
- 4.2 Evaluation of classifiers . . . 25

4.2.1 Parameter tuning for SVMs 25
4.2.2 Moving a frequency span of fixed width 26
4.2.3 Using equally spaced frequencies 29

5 Conclusions **32**

Bibliography **34**

A Additional figures **35**

B Resonance frequencies of a cylindrical cavity **37**

1

Introduction

Radar techniques can be used for detecting objects in the environment. In addition to the more conventional use to detect objects in the far field, i.e. at distances much greater than the wavelength of the radio waves used, it can also be used to detect objects in the near field, that is at a distance smaller than or equal to the wavelength considered. One use of this is to detect changes in material flowing through a pipe, by placing antennas used for measurement at some point along it. Such a measuring device is sensitive to changes in the permittivity and conductivity in the material flowing through the pipe, and can therefore be used to detect objects that have electrical properties which differs from the background, i.e. the material that normally flows through the pipe.

Constructing a detector that only uses changes in the signal for detection has an obvious drawback. Any object different from the background would, to some degree, result in a change in the signal. But in many cases, for optimal performance, not all types of objects should result in a detection. A typical scenario is that air bubbles are commonly appearing but not problematic, and as such should pass undetected. This work focuses on finding structures that distinguishes the signal changes caused by air bubbles from those caused by plastic and metal objects, and methods that could capture such differences. The tools evaluated for this are multivariate methods for classification. Such techniques allows for making distinctions between measurements resulting from the different cases when a foreign body is present and when only background material is in the pipe, rather than just looking for a significant change in the signal. But for a given problem there is rarely an obvious choice of method, and for every method there are parameters to tune to achieve good results. This has to be investigated by testing classifiers using acquired data.

If the background material is inhomogeneous, this is also a cause of changes in the signal. A simple way to take this into consideration in detection is to use means and variances to normalize measurement data. It might be the case, however, that there

are more structure in these changes than captured by such a simple method. In this work such structures have been found, and an attempt to use data transformations to eliminate them is evaluated in terms of the accuracy of a simple detector.

1.1 Aim

The goal of this work is to find ways to improve the accuracy in detecting objects in the setting at hand. The aims can be formulated as

1. Find the parts of the signal is most useful for separating samples containing background, or background and air from samples containing background and a plastic or metal object, at different levels of simulated measurement noise.

2. Examine the signal changes caused by background variability and possible ways to handle such changes to improve detection performance.

3. Find the most suitable classifier, from a set of classifiers chosen to compare, for distinguishing samples containing background or background and air from samples containing background and a plastic or metal object at different levels of simulated measurement noise.

1.2 Thesis outline

The structure of the thesis is mainly based on the distinction between data analysis and evaluation of classification methods. In the theory section brief descriptions of the different classifiers are found, and references to more detailed information on them. The methods section describes the data used in the project, some important, case specific, definitions of regularly used terms, and the evaluation methods for comparing classifiers. In the results the first part describes characteristics in the two types of data, while the second part contains the results from evaluating different classifiers. The discussion on the results is presented in connection to the respective findings. Finally, the results are summarized in a concluding section.

2

Theory

Throughout this section, as in the rest of the report, K will denote the total number of classes, N the number of observations and p the number of features.

2.1 Classifiers

2.1.1 KNN

The k-nearest-neighbor algorithm is the simplest method implemented. As the name suggests it uses the majority vote of the classes of the k data points closest to a sample for classification. However simple, there are parameters for the algorithm to be chosen, such as the number of neighbors considered, distance metric, distance weight etc. The distance metric can be altered to give different weights to distances in different components. The distance weight is used to give difference weights to neighbors in the majority vote step, depending on their distance from the sample. In this work, however, only the Euclidean distance with equal weights to all votes have been used.

2.1.2 Discriminant analysis

The discriminant analysis classifiers are parametric classification algorithms based on modelling class densities as multivariate Gaussian distributions. The probability density function of class k in $\mathbf{R}^p$ is given by

$$f_k(x) = \frac{1}{\sqrt{(2\pi)^p |\Sigma_k|}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}, \tag{2.1}$$

where μ_k and Σ_k are the class specific mean and covariance matrix respectively. Given that the prior of class k, that is the probability that a random sample belongs to class k, is π_k we can get the posterior probability that a sample $X = x$ belongs to class

k by applying Bayes theorem. Since G is discrete this posterior probability is given by

$$P(G = k|X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}. \tag{2.2}$$

The learning of a classifier of this type consists of estimating the parameters in the distribution, and the classification is done by maximizing this posterior probability over class labels, as

$$\hat{G} = \arg\max_k P(G = k|X = x) \tag{2.3}$$

where $\hat{G}$ is the predicted class label. $\hat{G}$ is piecewise constant on $\mathbf{R}^p$, and the boundaries between sets where it takes different values are of the form

$$\{x : P(G = k|X = x) = P(G = l|X = x), k,l \in \{1,...K\}, k \neq l\}. \tag{2.4}$$

To understand the shape of these sets it is useful to consider the log ratio of the pdf for two classes, since the priors only contributes with a constant:

$$\begin{aligned}\log \frac{f_k(x)}{f_l(x)} =& -\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\log\left((2\pi)^p|\Sigma_k|\right) + \\ & \frac{1}{2}(x-\mu_l)^T\Sigma_l^{-1}(x-\mu_l) + \frac{1}{2}\log\left((2\pi)^p|\Sigma_l|\right)\end{aligned} \tag{2.5}$$

When the covariance matrices are distinct for each class, this is quadratic in x. This is exactly the procedure in quadratic discriminant analysis (QDA), which also explains the name. The assumption when using linear discriminant analysis is that all classes have the same covariance matrix, $\forall k : \Sigma_k = \Sigma$, and a pooled estimate is used for all classes. The only parameter differing between the classes is then the mean. In this case equation (2.5) simplifies to

$$\log \frac{f_k(x)}{f_l(x)} = -\frac{1}{2}(\mu_k + \mu_l)^T\Sigma^{-1}(\mu_k - \mu_l) + x^T\Sigma^{-1}(\mu_k - \mu_l) \tag{2.6}$$

which, as expected, is linear in x.

An important property of these classifiers is the number of parameters estimated. For QDA they are

$$N_p = (K-1) + Kp + \frac{p(p+1)}{2}K \tag{2.7}$$

where the terms are from the priors, means and covariance matrices (which are symmetric), respectively. For LDA only one covariance matrix is estimated, and thus there is no factor K in the last term. The number of parameters to estimate does, however, increase quadratically with the number of features. This means that if the number of parameters estimated are more than the number of observations used for estimation, the covariance matrix will be singular. In this work this problem has been resolved by using the pseudo inverse of the covariance matrices. Another option would be to use regularization, i.e. to add a prior.

For a more detailed description of the discriminant analysis classifiers, see for example [1].

2.1.3 Support vector machines

The support vector machine (SVM) algorithm is based on the idea of finding an optimal separating hyperplane between two classes. In its original shape it can only be used for binary classification. The theory behind it is somewhat more involved than for the classifiers previously described, specifically the use of kernel functions to extend the feature space. Therefore only a brief introduction of it will be given here. A more detailed, but still brief, explanation can be found in [2]. An extensive description is presented in [1].

Consider the case where we have data, x_i with $i = 1,...N$, with labels $y_i \in \{\pm 1\}$. If the data from different classes can be separated by a hyperplane it is called linearly separable. A hyperplane can be described by

$$\{x : f(x) = x^T\beta + \beta_0 = 0\} \tag{2.8}$$

and a hyperplane separating the classes satisfies $y_i f(x_i) > 0$ for all i. It can be shown that $f(x)$ gives the signed distance from the hyperplane to x, and hence a new sample,x, is then classified by letting $\hat{y} = \text{sign}(f(x))$. Moreover we can find a hyperplane that maximizes the margin to all points by solving

$$\begin{aligned} \max_{\beta,\beta_0,||\beta||=1} \quad & M \\ \text{subject to} \quad & y_i(x_i^T\beta + \beta_0) \geq M,\ i = 1,\ldots,N. \end{aligned} \tag{2.9}$$

This can be rewritten, by replacing M by $\frac{1}{||\beta||}$ and dropping the norm constraint, as

$$\begin{aligned} \min_{\beta,\beta_0} \quad & ||\beta||^2 \\ \text{subject to} \quad & y_i(x_i^T\beta + \beta_0) \geq 1,\ i = 1,\ldots,N. \end{aligned} \tag{2.10}$$

It might of course be the case that our data is not linearly separable. In this case we introduce nonnegative slacks, ξ_i $i = 1,\ldots N$, and in this setting we look to solve

$$\begin{aligned} \min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & \xi_i \geq 0, y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i,\ i = 1,\ldots,N. \end{aligned} \tag{2.11}$$

The optimization problems (2.9)-(2.11) are all convex problems, guaranteeing that any optimum found is in fact a global optimum.

From the formulation above it is clear that the parameter C controls the penalization of the misclassified cases in the training data. It is referred to a s the *cost parameter* or *box constraint*. A larger value of C results in fewer misclassifications of the training data, while a smaller value of C gives a larger margin for the correctly classified data.

So far only a linear boundary for classification have been considered. The way to introduce a nonlinear boundary is to use a kernel function to transform the feature space.

In short this means that the input vectors, x, are mapped into a higher (possibly infinite) dimensional Hilbert space by the use of a function, $\phi(x)$. An optimal hyperplane is then sought in the new feature space, which translates to a non linear decision boundary in the original space. Studying the minimization problem in detail, it turns out that the function $\phi(\cdot)$ itself is not needed, but only the associated inner product as

$$K(x_i,x_j) = \langle \phi(x_i), \phi(x_j) \rangle \tag{2.12}$$

where $K(\cdot,\cdot)$ is the kernel function. In this work a radial basis kernel function was used when evaluating the SVM classifier. This function has the form of

$$K(x,z) = e^{-\gamma ||x-z||} \tag{2.13}$$

where the γ for the binary SVM is expressed as $\frac{1}{2\sigma^2}$. To fully explain the meaning of the kernel function and its parameter a more thorough description is needed, and the interested reader is referred to [1]. The effects on the results for different values of C in (2.11) is described above. As for the γ parameter it is clear that its value affects how the distance between two points will impact the value of the kernel function. For a large value of γ, the value of $K(x,z)$ will decrease faster than for a smaller value. Studying the optimization problem in more detail, it turns out that only some of the training samples are relevant for classification. These are called *support vectors*, and in the classification of a new sample only the values of the kernel function for the new sample and the support vectors are used. For a large value of γ the impact of the support vectors close to the new sample will be much larger than that of the more distant support vectors and decreasing the value of γ will increase the impact of support vectors far away from the new sample.

Finally the SVM is in its original shape, as mentioned, a binary classifier. There are a number of techniques to choose from to allow for more than two classes, of varying popularity and complexity. The two most popular to use seem to be the one-versus-all (OVA) and one-versus-one methods (OVO). Both of these are based on training a number of binary SVM classifiers, which are then used for multi class classification.

For K classes in total, the one-versus-all method trains one SVM for each class. In training these K SVMs, for $k = 1, \dots K$ the training set is divided into two groups, where one group is the samples from class k and the other group are all other samples. For a new sample a prediction is made using each one of these K classifiers, and for each k the predicted label is then "class k" or "not class k". The prediction for the multi class SVM is then that k for which the binary SVM predicted "class k" by the largest margin.

Using the one-versus-one method one binary SVM is trained for each pair of labels, in total $K(K-1)$. Thus, for each of these data from only two classes is used. The results from using all of these binary SVMs are then used in a voting process, and the majority vote decides the predicted label of the multi class SVM.

Other methods exist as well, and there are techniques that do not use a number of binary SVMs. A comparison of some of these methods are presented in [3].

3

Method

In this work two sets of data were used. Both sets were generated by numerically solving Maxwell's equations for the geometry of the problem. The first part of the method section will in detail describe the geometry, the different data sets and the models behind them. In the second section some terminology is introduced and described, and the third section will deal with the methods for testing and evaluating classifiers.

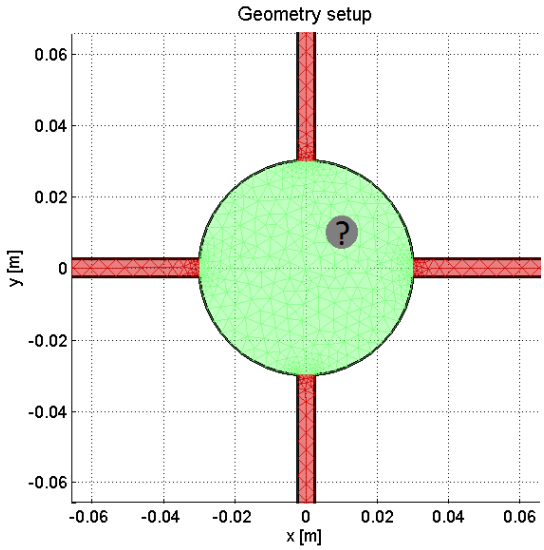


Figure 3.1: Visualization of the geometry of the model. The light green region is the background, waveguides are colored red and the grey region represents an object.

3.1 The data

The data was generated by modelling part of a hypothetical setup. The part modelled consists of an intersection of a pipe through which background material passes. At this intersection four waveguides are placed, equally spaced, around the pipe. The geometry described is illustrated in Figure 3.1. The geometry is divided into different regions: the interior of the pipe (background), the pipe itself and the waveguides. The instances modelling the presence of objects of other material than background also is represented by a second region in the interior of the pipe. The different regions are described by a set of electromagnetic parameters. These are the permittivity and the conductivity. The difference between the two data sets lies in the respective scenarios modelled. For the first data set the parameters of the background material were constant in the whole interior region, modelling completely homogeneous background material. For the second set the parameters varied stochastically, modelling inhomogeneous material. The stochastic behaviour is further described by two more parameters, explained in the sections below.

The two data sets have a common form. It consists of scatter matrix elements for a set of frequencies considered. Since there are four waveguides the full scatter matrix is 4×4, but since it is symmetric only the lower triangular part of it is needed. Thus there are 10 elements for each frequency. Element (i,j) in the scatter matrix is the quotient of the output signal from port i and the input signal at port j. These are complex valued. The number of frequencies considered was 400, equally spaced between 0.1 and 3 GHz.

3.1.1 Modelling homogeneous background and objects

This data is obtained by numerically solving Maxwell's equations using the finite element method (FEM), and the results models real world measurements using a network analyzer. The region modelling the background material is described by its relative permittivity, $\epsilon_{bg,r}$ and conductivity, σ_r. Both these parameters are constant over the interior region, modelling perfectly homogeneous material. Data was generated for all combinations of five levels of $\epsilon_{bg,r}$ (40, 50, 60, 70 and 80) and for levels of conductivity corresponding to $\delta/d = 1/2,1/4,1/8$ and $1/16$, where δ is the skin depth at 2.5 Ghz and d is the diameter of the pipe. The skin depth is a measure of the loss in current density when the field propagates through the material. The current density decreases exponentially with the distance from the surface, a relation that can be formulated as

$$J(l) = J_S e^{-\frac{l}{\delta}} \tag{3.1}$$

where J is the current density at distance l from the surface and J_S the current density at the surface. The values of $\epsilon_{bg,r}$ and σ_{bg} is a complete description of the scenario modelled in the case where no object or air bubble is present.

The air bubbles and objects are modelled by smaller circular regions in the intersection with different parameter values than the background. Air bubbles are represented by $\epsilon_r = 1$, three types of plastics by $\epsilon_r = 2$, 3 and 4, respectively, and metal as a perfect electrical conductor, i.e. a region with infinite conductivity. Simulations are carried

out for objects with radii spanning from 0.5 mm to 6 mm in 23 steps, equally spaced. The different positions modelled are about 3000 (somewhat fewer for larger objects), distributed over the interior region by altering radius and angle of in fixed steps. At most one object is included in every simulation and only one realization is produced for every class-position-size-backgruond combination, since the result would be identical if more were generated.

3.1.2 Modelling stochastically varying background and objects

In this data set the background parameters $\epsilon_{r,bg}$ and σ_{bg} are not constant, but varies stochastically over the region. The objects are modelled in the same way as for the data with homogeneous background. The values of the background parameters are set by generating a sample of a autocorrelated normal distribution. The means are 60 for $\epsilon_{r,bg}$ and σ_{bg} such that $\frac{\delta}{d} = 1$, for all cases. The root mean square (RMS) determines the mean amplitude in deviations from the mean. This value is varied in terms of the mean, and simulations are carried out for RMS = 1,2 and 4 % of the mean value. The autocorrelation between two points in the interior region is given by

$$e^{-\frac{2d^2}{\xi^2}}, \tag{3.2}$$

where d is the distance between points considered and ξ the correlation length. Thus the value of the correlation length determines how quickly the value of the parameters varies in space – larger correlation lengths results in larger areas with similar values, since they are more correlated.

The actual generation of the backgrounds with varying parameter values is done by first generating a sample of an uncorrelated, 2D, Gaussian distribution. The desired autocorrelation is then achieved by computing the convolution of this with this sample and the correlation function (3.2), using the Fourier transform.

For every scenario including at most one object, 5000 simulations were made with randomly generated background material as described above.

3.2 Definitions

Some definitions are introduced to simplify the description of different cases in what follows. Firstly the following short definitions are made:

- An *empty measurement* is a simulated measurement carried out on only background material.
- A *foreign body* is a region in the simulations representing air, plastics or metal.
- An *object* is a region representing plastics or metal, i.e. a foreign body that is not air.

3.2.1 The notion of signal

The data from all classes does of course consist of (simulated) signals. It is, however, useful to make another definition of what a signal is in this case. In terms of whether there is only background present or there is background and a foreign object, the normal case is only background. Therefore the data when only background is measured can be considered as no signal, and the results of something more than background signal. We write this as

$$S = S_{body} + \Delta S \tag{3.3}$$

where S is the total signal, S_{body} is the part of the signal resulting from a foreign body interacting with the field and ΔS is the noise in the signal. The noise part can be further divided into two parts as $\Delta S = \Delta S_{bg} + \Delta S_{meas}$. Here the ΔS_{bg} is the noise generated by variability in the background and ΔS_{meas} is the noise from measurement equipment.

3.2.2 Adding noise to the simulated data

The simulation of measurement noise in this work is based on the assumption that the power of the noise is equal for all channels, by using the mean signal power over all channels to define the signal to noise ratio (SNR). Another choice would be add noise with different variance for different channels, to get a constant SNR when considered channel by channel. Both choices are equally reasonable, and models different measurement equipment, but might lead to different effects of the noise added.

For a given signal, $x = [x_1,x_2,...,x_p]$, the power of the signal is given by

$$P_s = ||x||^2 = \sum_{i=1}^{p} |x_i|^2, \tag{3.4}$$

which can be used to estimate the mean signal power from a given data set. The noise is modelled by generating complex random Gaussian vectors, of length p, and adding such a vector to each sample in the data used. The power of the noise is then given by the expectation of the norm squared of a noise vector. Denoting a noise vector $\chi = [\chi_1,\chi_2,...\chi_p]$, every element of this has the form

$$\chi_i = X_{i,1} + iX_{i,2} \tag{3.5}$$

where both $X_{i,1}$ and $X_{i,2}$ follows a normal distribution with zero mean and variance σ_n. All components of χ are independent, as are the real and imaginary parts of the individual elements. Thus the expectation of the norm, and the noise power, is given by

$$P_n = E\left[||\chi||^2\right] = E\left[\sum_{i=1}^{p} |\chi_i|^2\right] = \sum_{i=1}^{p} \left(E\left[|X_{i,1}|^2\right] + E\left[|X_{i,2}|^2\right]\right) = 2p\sigma_n^2 \tag{3.6}$$

To obtain a given SNR when generating noise, the σ_n must be set so that $P_s/P_n = \text{SNR}$. This gives that $\sigma_n = \sqrt{\frac{P_s}{2p\text{SNR}}}$ is the proper choice.

3.3 Classifiers and feature selection

A large part of the work is to evaluate the performance of different classifiers. In connection to this the results of using different subsets of features were also investigated. In general the work flow has been to first evaluate the classifiers using simulated data free from noise. This gives information on how well different classifiers perform in an ideal setting. This information has then been used to investigate the effect of noise in the signals for the most promising candidate classifiers chosen from the test round without noise. The procedures for manual feature selection are explained in the section that follows.

The measure to compare classifiers was in most cases the correct classification rates in terms of object - no object. This is motivated by the task at hand. It is of no greater interest so distinguish between objects of different classes, or air and empty measurements. The objective is to optimize performance with respect to detecting objects. For classifiers allowing for multiple classification this was done by using all classes as labels in training and testing, and sorting the predictions and true labels into binary classes after classification. For the classifiers allowing for only two classes (e.g. standard SVM) the measurements were divided into object/no object classes before training and testing.

In the analysis of the data with non homogeneous background different data transformations were tested in terms of their impact on signal detection, using a simple detection algorithm based on featurewise deviations from the normal case (when only background is measured). Here the area under the receiver operating characteristic (ROC) curve was used.

3.3.1 Implementation of classifiers

The majority of the evaluation of classifiers was carried out using Matlab, where a variety of built in methods for classification is available. They do not, however, handle complex data. Therefore the data was split up into real and imaginary parts, thus doubling the number of features. The amount of data used for training and testing varied somewhat, depending mostly on computational times for different methods. In general cross-validation of some sort was used to evaluate the results.

The classifiers for which built in functions in Matlab was used was KNN (ClassificationKNN class was used), LDA and QDA (ClassificationDiscriminant) and SVM (svmtrain for training and svmclassify for classification).

The implementation used for the multi class SVM was the libSVM package [4], available online. This software uses a one-versus-one approach to perform classification into more than two classes. This is motivated by the fact that the results compared to using one-versus-all are comparable, but the one-versus-one algorithm is usually faster. It should be mentioned that OVA has often proven to be slightly better than OVO (see for example [5]), but with much longer computational times.

3.4 Manual feature selection

Given the properties of the data in its original form (8000 features, when separated into real and complex parts) it is in most cases necessary to perform some kind of feature selection manually before using the data. This simply because the computational times for different automated feature selection methods are too long if initially using all features. Therefore three types of manual feature selection is often used as an initial step. These three procedures are explained and motivated below.

3.4.1 Choosing scattering matrix elements

The different matrix elements corresponds to different kinds of mechanisms in the measuring process. A common way to describe the elements are as *reflection* and *transmission* elements, which is partly self explanatory. The reflection elements are the elements on the diagonal of the scattering matrix. These are the ratio of the output from one port and the input from the same port. The output signal is hence the part of the signal that has entered the cavity and then reaches the same port via reflection. A transmission elements is the ratio of the output from one port divided by the input from another, i.e. element S_{ij} is the ratio of output at port i and input at port j. Hence the signal is transmitted from one port to another (possibly partly due to reflection processes).

3.4.2 Using only a frequency span of fixed length

The data from different parts of the frequency span might contain more or less useful information and one way to evaluate this is to fix the length of the frequency span and move a span of this length along the full span, evaluating the accuracy of the classifier in every point.

3.4.3 Using equally spaced frequencies

In the data studied a high correlation between neighboring frequencies is expected. Therefore a decrease in the resolution in terms of frequencies might be a means to improve results. A simple way to investigate this is to use only a subset of all available frequencies (or a subset of all frequencies in an interval of choice). The most apparent way to do this is to use equally spaced frequencies from the full set of frequencies, and vary the number of frequencies used. Again this is evaluated by studying the accuracy of a given classifier for different numbers of frequencies used.

3.4.4 Using a part of the frequency span of varying length

The number of features used will, for mathematical reasons described in the theory section, have an impact on how well a classifier performs. Therefore it might be of interest to use only a part of the frequency span and varying the length of this part. As an example if it is found that the lowermost part of the frequency span is most useful

the limiting highest frequency to use can be evaluated roughly by using a span starting from the lowest available frequency and varying the highest used frequency.

4

Results

The results part will mainly be divided into two sections. The first part will be deal with the results on analyzing the data, both the data with constant background parameters and the data with stochastically varying parameters. The second part will treat the different classification methods and the feature selection involved in these.

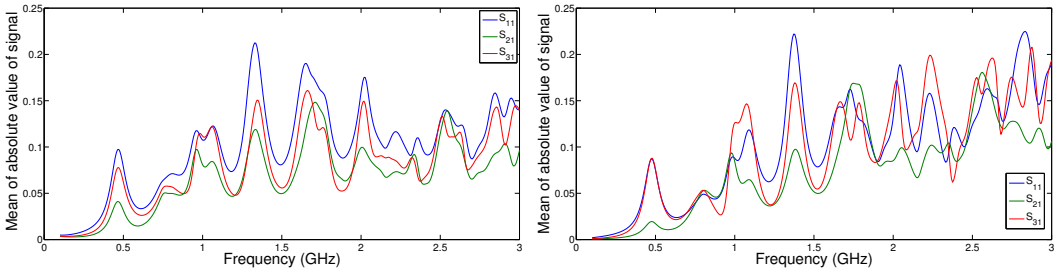
4.1 Data analysis

The analysis of the data aims to find characteristics in the data that might reveal and motivate methods for manual feature selection and/or transformations that could help in the classification of measurements. The methods are for the most part straightforward, comparing means and variances for different classes and graphically examining the distribution of data points.

4.1.1 Data with fixed background parameters

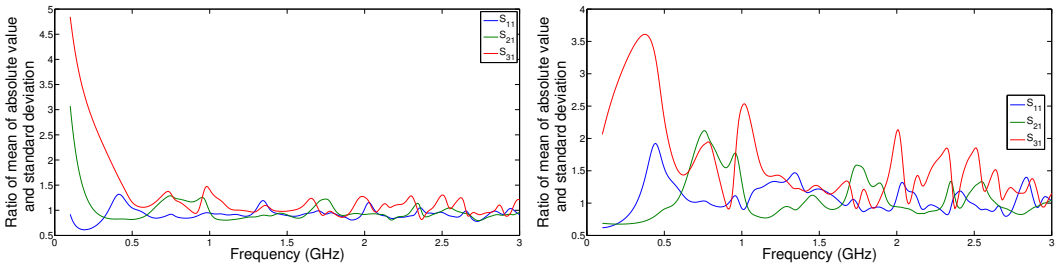
To begin the study data for fixed background parameters, set to $\epsilon_{bg,r} = 40$ and σ_{bg} corresponding to $\frac{\delta_{bg}}{d} = \frac{1}{2}$, is examined. From each class 1000 instances are loaded, all of which with object radius of 4 mm. The positions of the objects are chosen at random. The features studied are all frequencies from S_{11}, S_{21} and S_{31}. The reason for this is that these correspond to reflection, diagonal transmission and diametrical transmission respectively. When the positions are distributed in a circular symmetrical fashion parameters estimated by aggregating results from all samples will be equal for two elements within one of these three groups, and hence only one of them needs to be studied.

For the aim in this work the interest in studying the data is to investigate how different classes are separated by different features. By the definition of signal in this context, the absolute value of a feature for a given sample represents the separation



(a) Metal **(b)** Plastic $\epsilon_r = 2$

Figure 4.1: Mean of the absolute value of the signal for objects of size 4 mm. For metal objects the signal is strongest for the reflection element for all frequencies, while the S_{31} element gives a stronger signal for some frequencies for plastic objects.

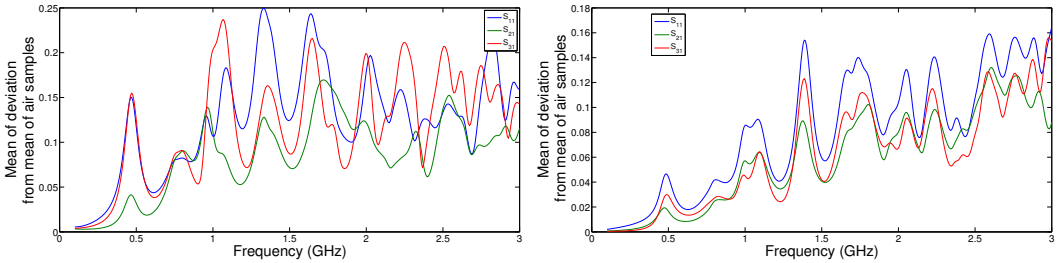


(a) Metal **(b)** Plastic $\epsilon_r = 2$

Figure 4.2: Ratio of the mean of the absolute value and the standard deviation for each feature for objects of size 4 mm. For metal objects the lowest part of the frequency span and the transmission elements are clearly giving the largest separation-standard deviation ratio from empty measurements. For the plastic class the picture is not as clear. The S_{31} element gives the largest values, again at the lowest frequencies. But for the other two elements frequency dependence it not as simple as before, and their values compared to each other varies more along the frequency span.

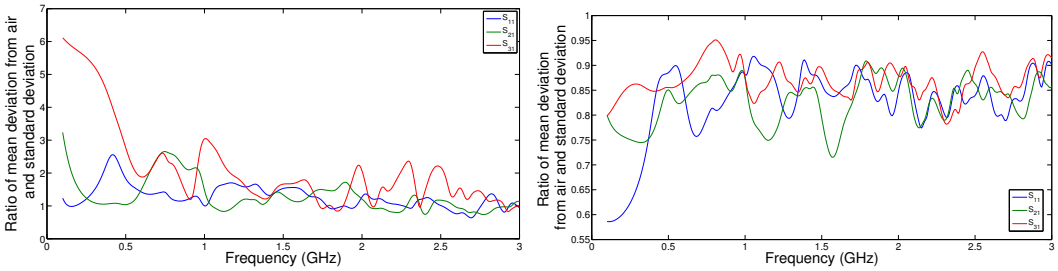
from an empty measurement at that feature. Thus an indication of how well a feature separate measurements with an object from empty measurements is given by the mean of the absolute value. Figure 4.1 shows plots of this. From these plots alone it seems that the reflection element (S_{11}) separates the objects the best. Also the absolute value shows an increasing trend with frequency. But the absolute value should be compared to the variability in the signal. Therefore plots of the ratio of the mean absolute value and the standard deviation of each feature, are presented in Figure 4.2.

Finally it is of interest not only to separate the object classes from the empty measurements but also from the instances containing air bubbles. Therefore a similar analysis is made now comparing, for each feature, the distance for a given realization containing an object to the corresponding mean of the air measurements. Furthermore the standard deviation used in this case is (again for each feature) the pooled standard deviation of the air measurements and the object class studied. The results are presented



(a) Metal **(b)** Plastic $\epsilon_r = 2$

Figure 4.3: Mean of the deviation (in absolute value) from the mean of the air samples for two object classes. The increase with frequency is similar to the earlier results.



(a) Metal **(b)** Plastic $\epsilon_r = 2$

Figure 4.4: Most notable is the fact that all values for the plastic class are below 1, that is the separation from the air samples is for all features smaller than the pooled sample standard deviation. For metal the behaviour is somewhat similar to when considering the distance to the mean of the empty measurements.

in Figures 4.3 and 4.4. Similarly to when comparing to empty measurements separation seems to increase with frequency. When considering the ratio of this deviation with the corresponding pooled sample standard sample deviation, the behaviour is very different between the two classes compared here. For metal the lowest part of the frequency span for S_{31} yields the highest ratio, with around 6 as the highest value. In general there seems to be quite a few features giving a separation of at least 2 times the standard deviation. For the plastics with $\epsilon_r = 2$ none of the values are above 1, which indicates that the classes might be difficult to separate. Compared to earlier results the ratios are also quite stable over the frequency span.

4.1.2 Data with stochastically varying background parameters

In practical terms the second dataset is very different from the first one. In the first set, for a given object (with a given size in a given position) and a given set of background parameters there was only one realization. In the second dataset, since the background parameters are varying, multiple simulations were made for each given object and given background parameters, but with different realizations of the background material. The

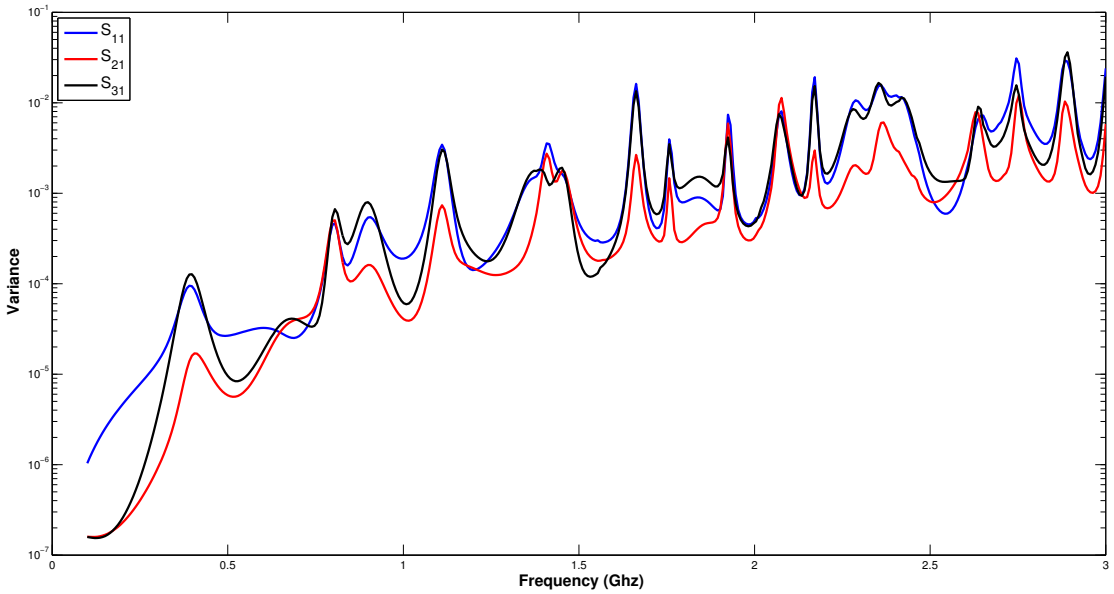


Figure 4.5: The behaviour of the variance is very similar for the three scatter parameters, with peaks appearing at the same frequencies. These arise from resonance modes in the circular cavity.

initial part of the analysis is carried out on data from background measurements.

Characteristics of background measurements

Firstly the variance for different frequencies was studied. As before we consider the scatter matrix elements S_{11}, S_{21} and S_{31}, as seen in Figure 4.5. Two features are immediately evident from this plot. The first one is that the variance generally increases with frequency. This is explained by the fact that the effect on variations in the background media will depend on the correlation length-to-wavelength relation. The second feature are the clear peaks that appear. The frequencies at which these are found seem to match quite well with the frequencies of the TE modes of the circular cavity, i.e. the resonant frequencies of the electric field (see Appendix B). What is more interesting for the application is what form the variations take for a single feature.

It turns out that the data for a single feature from several realizations in many cases takes a type of "banana"-shape in the complex plane. As an example, Figure 4.6 shows the data from two of the peak frequencies in the variance plot for a correlation length of 1.9 mm and RMS of 2%, from the S_{31} element. This shape appears with different curvatures and spread along and orthogonally to the curvature. At this point the question was posed whether the data could be transformed in a way that would ideally eliminate the correlation, both linear and non linear, between real and imaginary parts and standardize the variance. The reason for this is that some of the simpler algorithms for classification (e.g. discriminant analysis) are based on an assumption

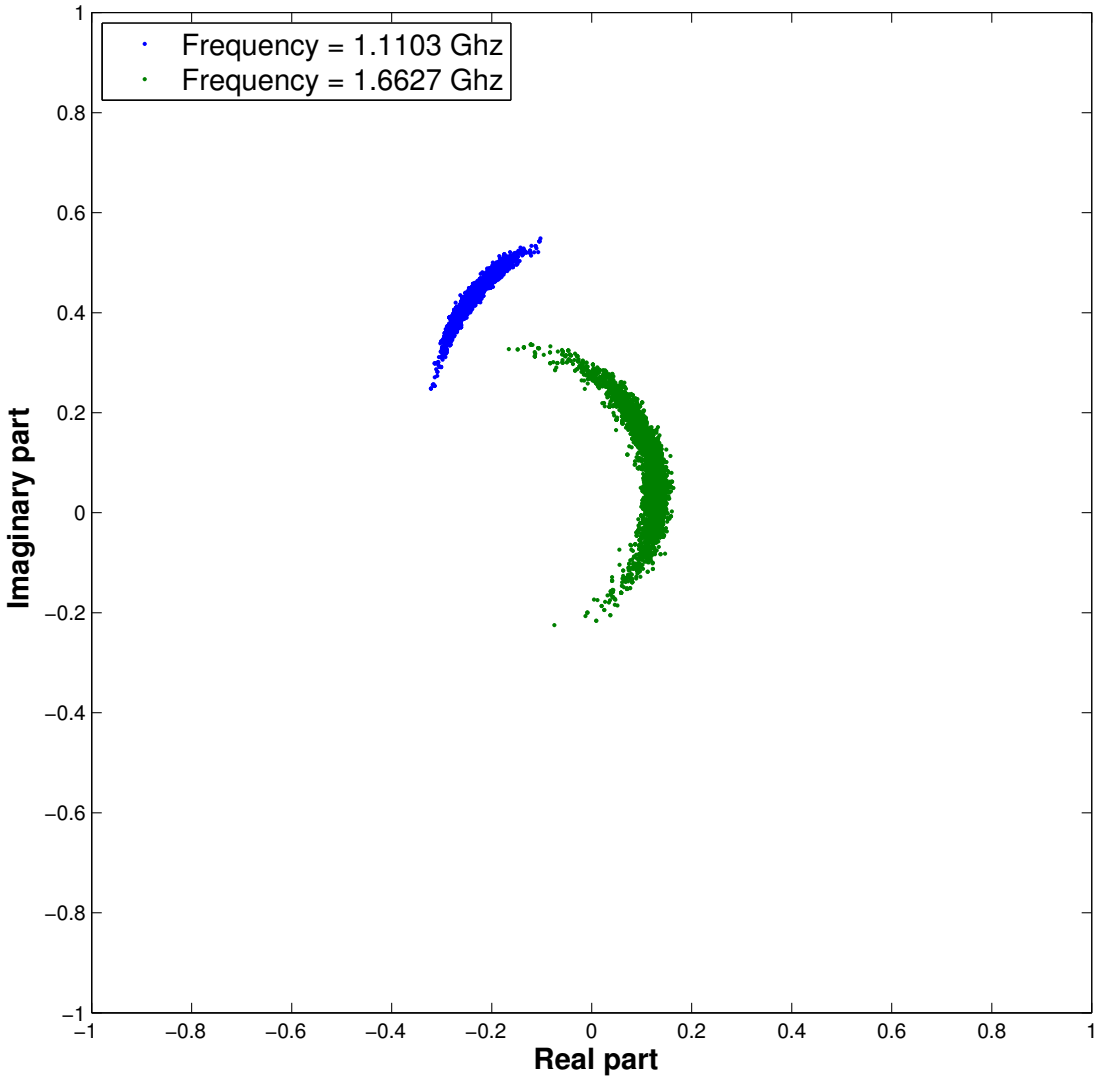


Figure 4.6: Data from two peak frequencies for the S_{31} element. The shape appears over the whole frequency span, and not only at variance peaks. The curvature and spread along the curve, and orthogonally to it, varies with frequency, scatter element, correlation length and RMS value.

that Gaussian distributions approximate the data well. Moreover, the perhaps simplest way to distinguish between the signal/no signal cases in the setting at hand would be to compare the featurewise deviation from the case when only background material is measured. It is likely that such a procedure would yield better results for data in which the parameters are uncorrelated and circular symmetrically distributed. The next section describes the transforms tested and the methods for evaluating them.

Transformations of data and evaluation of these

Three transforms will be evaluated, and the choices are perhaps most easily explained from the method to evaluate them. In the general setting in this work two distinctions of measurements are made in terms of signal/no signal, and object/no object. The transformations will be evaluated in terms of the first of these distinctions. That means that the interest is whether a transformation can improve the results in distinguishing between measurements made on only background, or background and air or object. A simple detector uses a number of measurements known (or assumed) to measure background material, to build statistics for each feature of the data. These statistics can then be used to set thresholds for each one of the features, used to determine when to classify a measurement as "signal".

A way to simplify this, is to use the statistics computed from the background measurements and use this to normalize the data for all features, so that a single threshold can be used for all features in testing. The idea is that if the data from the background measurements is distributed circularly symmetric with unit standard deviations in both directions (real and imaginary parts), a large enough deviation in one of the features is indication that something else than background material is passing the waveguides. The transformations can be seen as three such normalizations, taking different types of correlations into consideration.

For describing the transformations some notation will be convenient. Measurement k is denoted $x^k = x_r^k + ix_i^k$, and the corresponding normalized measurement is denoted $z^k = T(x^k)$, where T is the transformation used. Every measurement is a vector of length p, where p is the number of features. Letting $x^{k,l}$ denote entry l the value of this entry can itself be represented as a vector as

$$x^{k,l} = \begin{bmatrix} x_r^{k,l} \\ x_i^{k,l} \end{bmatrix} \tag{4.1}$$

The normalizations considered are meant to take into account the following types of correlations:

1. Different variances in real and imaginary parts, no correlation
2. Different variances in real and imaginary parts, linear correlation
3. Different variances in real and imaginary parts, linear and quadratic correlation

For the first two of these the normalization of a measurement is straightforward. The procedures are to estimate the diagonal and full covariance matrix and normalizing by this respectively. Formally this is expressed as

$$z^{k,l} = \begin{bmatrix} z_r^{k,l} \\ z_i^{k,l} \end{bmatrix} = \Sigma^{-1/2}(x^{k,l} - \bar{x}_{bg}^l) \tag{4.2}$$

Type	Radius (mm)	Position (mm)
Air	6	(13.8; 7.4)
	0.67	(3; 0.78)
$\epsilon_r = 2$	6	(13.8; 7.4)
	0.67	(3; 0.78)
PEC	6	(13.8; 7.4)

Table 4.1: The positions are expressed in a coordinate system where (0,0) is the center of the cavity.

where $\bar{x}_{bg}^l$ is the mean of feature l from the background measurements, Σ is the estimated covariance matrix, which in the first case is diagonal with the variances of the real part, $(\sigma_r^l)^2$, and imaginary part, $(\sigma_i^l)^2$ on the diagonal. For the second type the procedure is identical, with the exception that the full covariance matrix is now estimated. This implies that for the normalized background data real and imaginary parts will be uncorrelated with expected value zero and unit variance for all features. Again, the parameters are estimated using only background measurements, since scenario is that only this type of measurements are available for training, and the goal is the separate measurements of background material from measurements of background material and a foreign body of some type.

The third transformation involves another step, based on the assumption that the nonlinear correlation in the variables could be approximated by a second order polynomial. The first step is to normalize the data as in the second transformation. Two second order polynomials are then fitted to the data. One where the imaginary part is considered a function of the real part, and vice versa. The coefficients of the quadratic term from the fits are then compared, to examine along which axis the data is curved. From the data used it is found that the radius of curvature is almost exclusively along the real axis, that is the largest coefficient for the quadratic term is obtained when fitting a polynomial to the real parts as a function of the imaginary parts. In normalizing the data the value of the polynomial taken to describe the curvature is then subtracted from each point.

With the three transformations explained, the results of the evaluation will now be presented. There are three parameters affecting the variance in the data. These are the correlation length and RMS value of the background parameter, and the variance in normal distributed variables added to simulate measurement noise. The variance of the latter is expressed as a signal to noise ratio. The evaluation is carried out for a number of measurements noise levels with the parameters describing the variations in the background material are kept fixed. Moreover an AUC value is obtained for each object separately, which allows the study of both mean AUC values over all objects and the same value object by object, for all objects in Table 4.1.

Testing was carried out at all combinations of the two lower levels of correlations

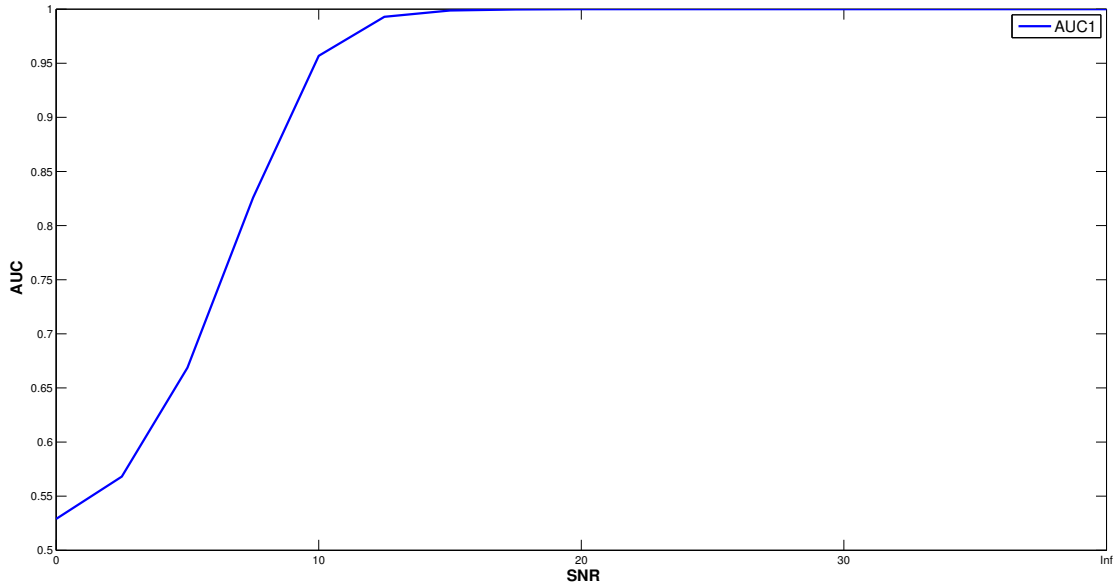


Figure 4.7: AUC values using the simplest transform for an object with $\epsilon_r = 2$, of radius 0.67 mm, for a correlation length of 1 mm RMS value of 1 %. At SNR levels above 20 dB detection is perfect.

length and RMS value, and at the highest RMS value and second highest correlation length. The SNR levels considered ranged from 0 to 30 dB in steps of 2.5 dB, and infinite SNR (no noise). For all combinations of background parameters the detection of the 6 mm objects of all classed yielded AUC values of 1, with one exception. This was for SNR of 0 dB and the highest RMS, and second highest correlation length. The values were then between 0.9992 - 0.9993 for the air bubble and 0.9987-0.9989 for the plastic object, with the simplest transformation scoring the highest in both cases. The smaller objects studied rendered some more interesting results.

The differences between the transformations used are in general small. To present them in a form which makes the differences clear, the difference in AUC value using the simplest transformation and the to others are considered, respectively. The first scenario considered is the setting where signal detection is easiest, that is for the lowest values of correlation length and RMS.

The AUC values using the simplest transformation at different SNR levels, for the smaller plastic object in Table 4.1, is seen in Figure 4.7. The differences from this for the other two transforms are seen in Figure 4.8. In this setting the simplest transformation gave the best detection rate at high SNR levels, while the other two worked better when there was more noise added. It should be noted that the AUC values at these noise levels are little above 0.5, which indicates that the detection is poor. The behaviour of the AUC values for the simplest transformation when instead considering an air bubble is very similar.

When increasing the RMS value and/or the correlation length, detection of the larger

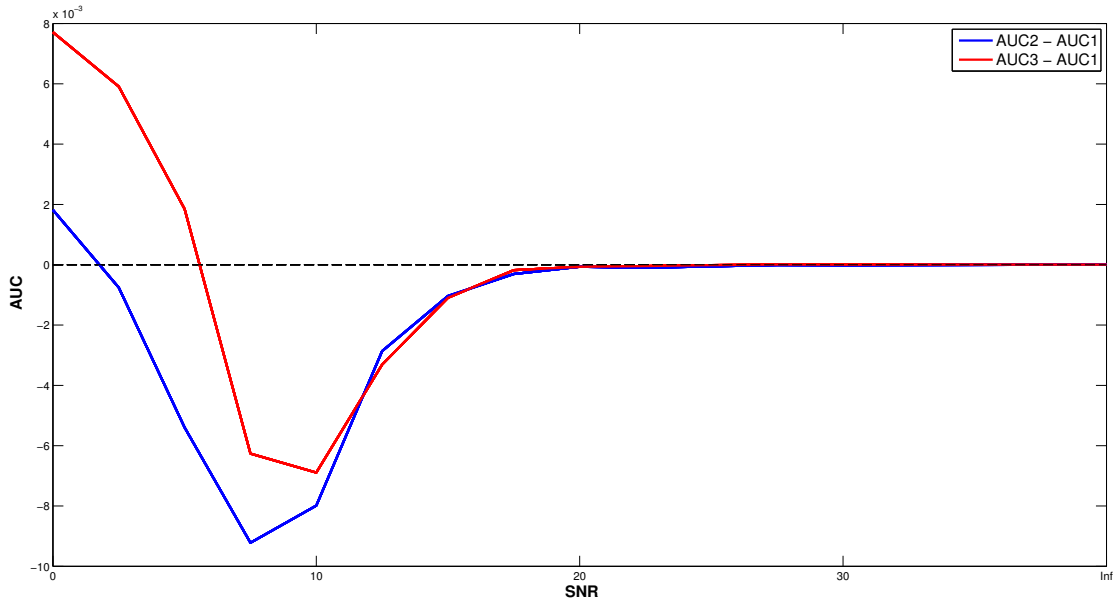


Figure 4.8: Differences in AUC values using the first transformation, or one of the other two, respectively, at correlation length of 1 mm and RMS of 1 %. The values are for a plastic object of radius 0.67 mm. The simplest transformation seems to be the better choice at lower noise levels. This plot is for the lowest amount of background variability considered.

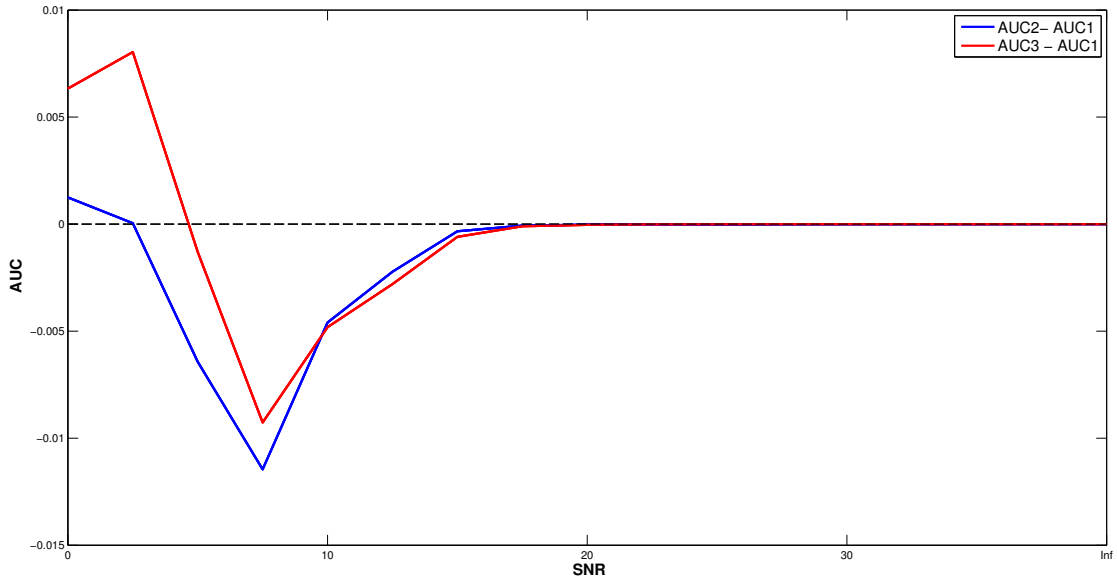


Figure 4.9: Differences in AUC values using the first transformation, or one of the other two, respectively, for an air bubble of radius 0.67 mm. Here the correlation length was 1 mm and RMS value 1 %.

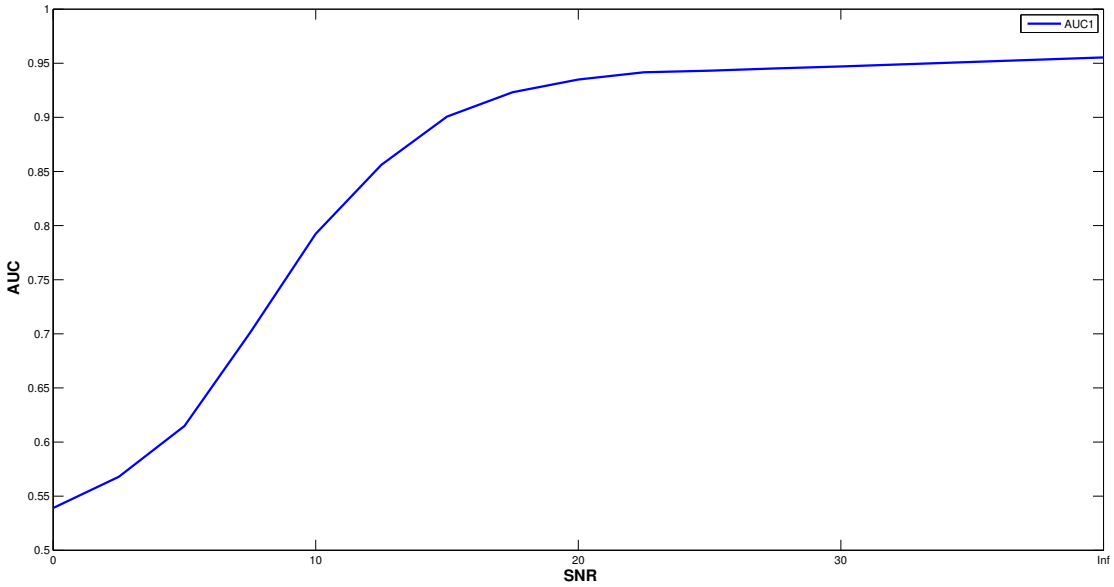


Figure 4.10: AUC values using the simplest transform for an object with $\epsilon_r = 2$, of radius 0.67 mm, for a correlation length of 1.9 mm RMS value of 1 %. The AUC does not reach 1 as at lower variability in the background material.

objects considered is still perfect, but changes the results for the smaller plastic object and air bubbles. In Figure 4.10 the results for the simplest transform at a correlation length of 1.9 mm and RMS value of 1 % is shown. Clearly detection is now more difficult, for example the AUC value in the noise free setting is now clearly below 1. Again comparing the differences between the transformations, as in Figures 4.11 and 4.12, we see that the results using the second and third transformation is now better at some noise levels, and worse at others. Most notably the results are better at higher SNR values. This might be because the curvature in the data clouds is more prominent at these noise levels, since the noise added is circularly symmetric.

If instead considering a RMS value of 2 % and correlation length of 1 mm the results are very similar, as seen in Figure A.1, A.2 and A.3. When considering higher values for both the correlation length and RMS the results for the smaller objects are generally poor, while the AUC values for the larger objects remain at 1 for all but the lowest SNR value, where it is still above 0.999 for all transformations. For a correlation length of 3.9 mm and RMS of 4 % the AUC values for all transformations are between 0.5 and 0.53 at all noise levels. This indicates that the classification of a sample as object/no object is more or less a coin toss. However, at this amount of background variability the results using the third transformation is uniformly better than using the second transformation, which is in turn better than using the first transform at all noise levels as seen in Figure 4.13. Thus it seems that the detection is slightly improved with the second and third transformation, but still very poor. It should be noted that the object size tested for must be considered very small, so the difficulties of detection is not surprising.

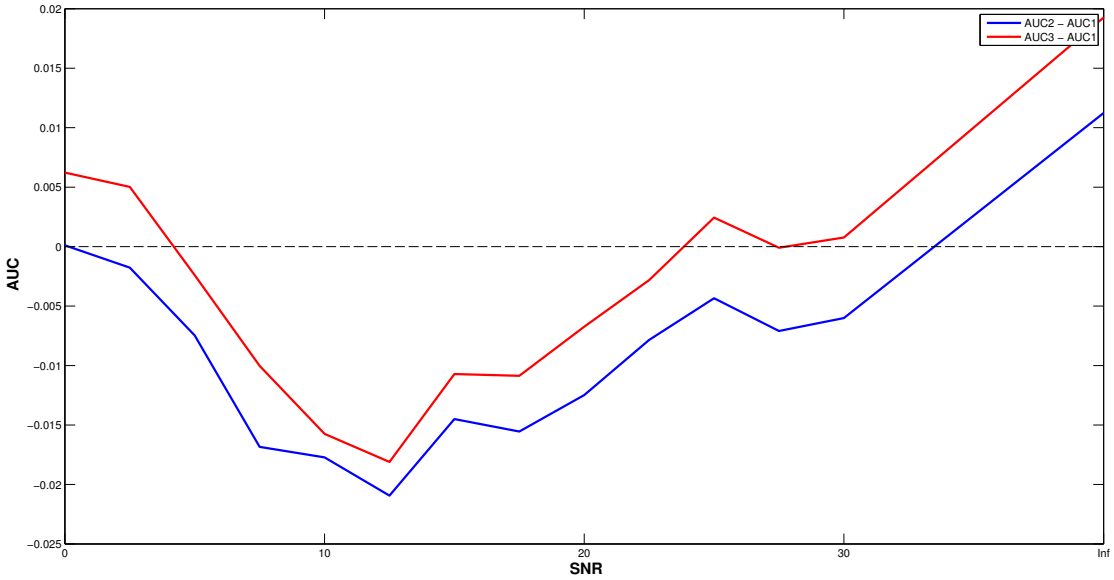


Figure 4.11: Differences in AUC values using the first transformation, or one of the other two, respectively. This plot is for a correlation length of 1.9 mm and RMS value of 1 %. The values are for a plastic object of radius 0.67 mm. The simplest transformation seems to be the better choice at lower SNR values. This plot is for the lowest amount of background variability considered.

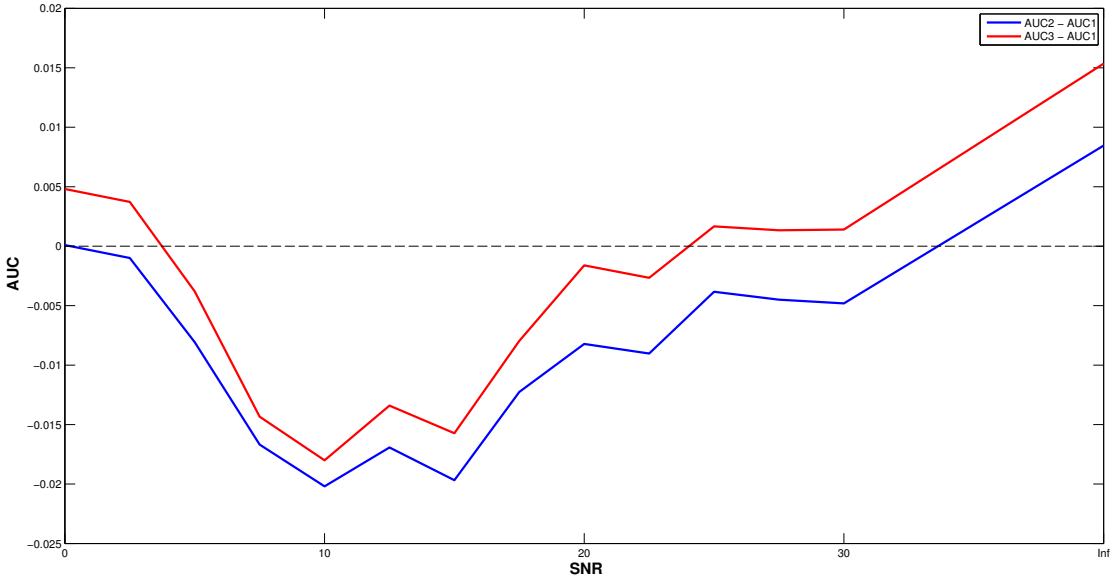


Figure 4.12: Differences in AUC values using different transformation, for an air bubble of radius 0.67 mm. The correlation length was 1.9 mm and RMS value 1 %.

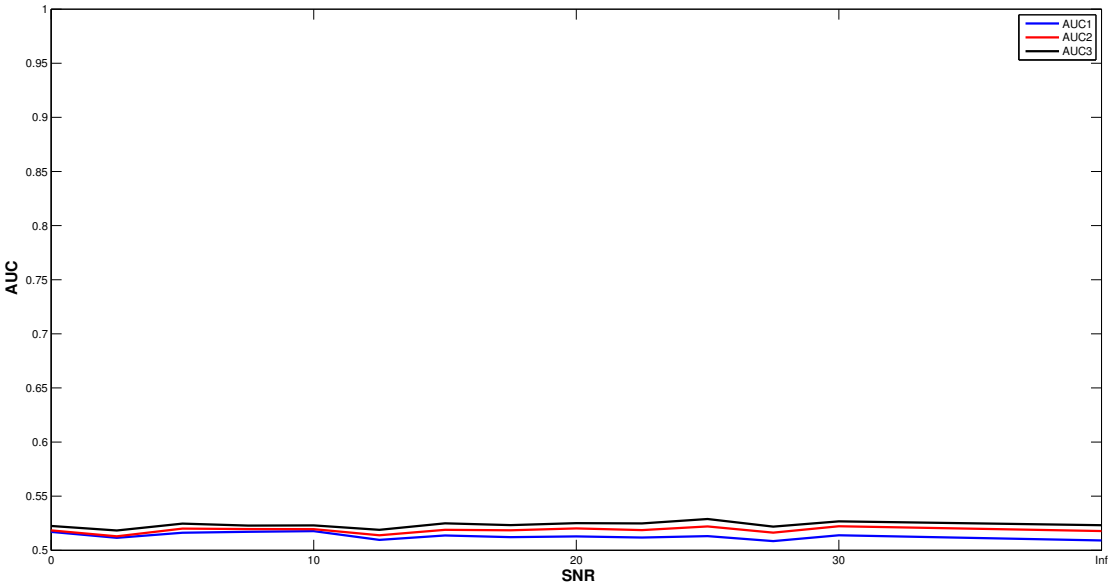


Figure 4.13: AUC values for the three different transformations at correlation length of 3.9 mm and RMS value of 4 %. The result is poor for all transformations, but the ordering of the transformations is the same at all noise levels, with the third transformation giving the best results and the first transformation the worst.

4.2 Evaluation of classifiers

This section describes the results found when testing different methods for classification, which also includes parameter tuning and feature selection. The aim has been to present the results in such a way that they make it easy to follow the effects both of choosing different classification methods and what features to use. The classifiers compared using manual feature selection are: QDA, LDA, KNN, SVM and libSVM. For the description of these refer to Chapter 2.

The performance measure used is the accuracy in terms of object/no object classes, and for the SVM classifier this is also how samples used for training and testing are labelled. The data, however, is originally divided into six different classes as described in section 3.1. Each of these six classes is a subclass to either the object or the no object class. For all classifiers but the SVM, training and testing is carried out using all six labels, but the accuracy is then computed only in terms of the object/no object distinction.

4.2.1 Parameter tuning for SVMs

For both the binary SVM and the libSVM that handles multiple classes there are parameters that need to be tuned. In both cases the cost parameter, C in (2.11), has to be tuned. Moreover a radial basis function (rbf) kernel was used for both types. This choice introduces a second parameter to tune. For the binary SVM the kernel function

	Binary
First gridsearch	$C = 10^{-2}$,$10^{-1.2}$,...$10^{9.2}$,10^{10}
	$\sigma = 10^{-5}$,10^{-4},...,10^{4},10^{5}
First min	$MCR = 0.19$ at $C = 10^{4.4}$, $\sigma = 10^{2}$
Second gridsearch	$C = 10^{3.5}$,$10^{3.6}$,...$10^{5.5}$
	$\sigma = 10$,$10^{1.1}$...10^{3}
Second min	$MCR = 0.1462$ at $C = 10^{3.6}$, $\sigma = 10^{1.8}$
fminsearch min	$C = 10^{2.83}, \sigma = 10^{1.60}$
	libSVM
First gridsearch	$C = 10^{-5}$,10^{-4}...,10^{14},10^{15}
	$\gamma = 10^{-5}$,$10^{-4.5}$,...,$10^{4.5}$,10^{5}
First min	$MCR = 0.1462$ at $C = 10^{6}$, $\gamma = 10^{-3}$
Second gridsearch	$C = 10^{5}$,$10^{5.1}$...10^{8}
	$\gamma = 10^{-3.5}$,$10^{-2.5}$,$10^{-1.5}$
Second min	$MCR = 0.09$ at $C = 10^{5.9}$, $\gamma = 10^{-2.5}$

Table 4.2: The results of the parameter tuning for the SVMs. They are generally not optimal when altering what features to use, but will still perform better than default values. The tuning is meant to make the comparison between the SVMs and other classifiers fair.

is expressed as

$$K(x,y) = e^{-\frac{\|x-y\|}{2\sigma^2}} \tag{4.3}$$

while for the multi class SVM the $\frac{1}{2\sigma^2}$ is instead expressed by a parameter γ, as

$$K(x,y) = e^{-\gamma\|x-y\|}. \tag{4.4}$$

For the two types two consecutive gridsearches was performed to find good parameter values. For the binary SVM these parameters was further refined by using MATLABs fminsearch on the misclassification rate (MCR). The data used was 1000 samples from each of the six classes, and in every point of the grid a five fold cross validation was used to obtain a mean MCR. The background parameters was $\epsilon_{bg,r} = 40$ and σ_{bg} corresponding to $\frac{\delta}{d} = \frac{1}{2}$.The results of the gridsearch and minimization are presented in Table 4.2.

4.2.2 Moving a frequency span of fixed width

In this round of testing the classifiers were tested using a frequency span of 60 frequencies altering the position of this subset of frequencies. For every position of the frequency span, the classifiers were tested using three different sets of scatter parameters: transmission only, reflection only and all parameters. The SNR level was also altered in 7

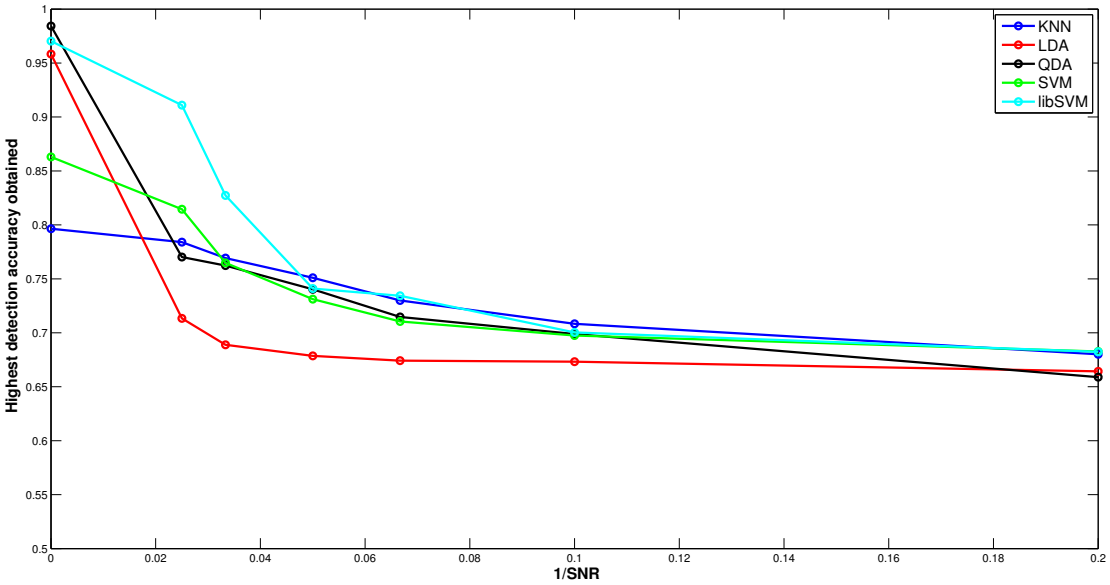


Figure 4.14: Highest obtained detection accuracy at different noise levels for the different classifiers. The value on the x-axis is 1/SNR and the leftmost point is for no noise. It is very clear that the impact of the noise on the performance differs between the classifiers, and that the best classifier using noise free data is not necessarily the best even at high but finite SNR levels. The KNN-data is for using one neighbor.

steps. The background parameters were chosen as $\epsilon_{bg,r} = 40$ and σ_{bg} corresponding to $\frac{\delta}{d} = \frac{1}{2}$. Position and size of foreign bodies was randomized. For KNN, LDA, QDA and the binary SVM the testing was done by using 1000 training samples and 1000 testing samples from each of the six subclasses. The classifiers were then evaluated at all element choices, subspan positions and noise levels using this data. This was then repeated five times reloading data in each iteration. For each of the foreign body classes data for 23 sizes in 2000-3000 positions was available (depending on the size), resulting in about 47500 samples from each class. In each iteration the subsets used for training and testing were randomly selected from all of the available data. For the libSVM classifier the computational times using that much data was too long, and hence a set of 500 measurements from each subclass was used in a 5 fold cross validation in each evaluation.

The results from this testing round was firstly meant to get indications on what classifiers to test further, but also on which parts of the frequency span and scatter matrix elements are most informative at different noise levels. To compare the classifiers at different noise levels the best achieved accuracy at each noise level is compared, as in Figure 4.14. The KNN-classifier was tested using 1,2,5 and 10 neighbors and the highest obtained accuracy was at all noise levels obtained using 1 neighbor. The only general conclusion on the results comparing classifiers is that LDA seems the performs the worst in all cases where noise is added to the data. This despite a high detection accuracy

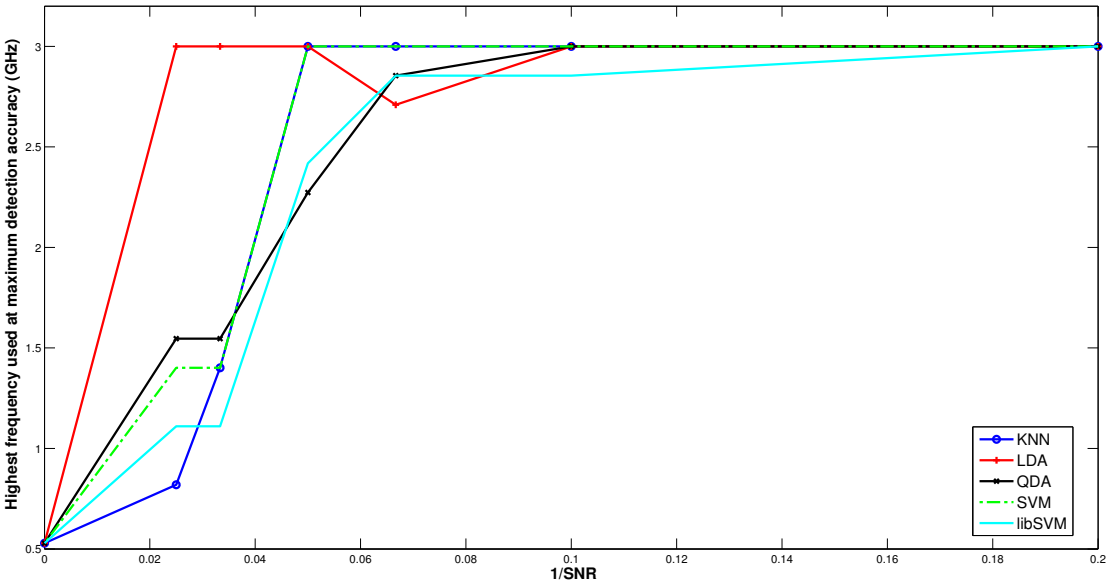


Figure 4.15: In general the frequency span yielding the highest accuracy moves to higher frequencies as the noise increases. Note that the graphs are all at the lowest possible frequency for the upper limit of the span, about 0.4 Ghz, when there is no noise. At the lowest SNR level considered, 5 dB, the highest possible position of the frequency span yields the best accuracy for all classifiers. The span was moved in steps of 20 frequencies, that is about 0.4 Ghz.

(about 0.95) for noise free data. The behaviour is similar for the QDA classifier, although the decrease is not as severe. The KNN gives the lowest accuracy at no noise, but the performance is rather robust to noise and at SNR = 20 it yields the best results. At this and lower SNR levels the performance of the SVM, libSVM, QDA and KNN are very similar. It is also clear that for the lower SNR values none of the classifiers perform very well. Since an equal number of samples is used from each of the six classes, two thirds of all the samples are measurements containing objects and one third containing only background material or background material and an air bubble. Thus classifying every sample as object would yield an accuracy of 2/3. The accuracies for SNR values below 20 dB are little above this value.

The other interest in this part was to investigate which frequencies and scatter elements renders the best accuracy. Figure 4.15 shows the upper limit of the frequency span giving the highest detection accuracy for all classifiers and noise levels. As the noise increases, so does the frequencies which are most useful for classification. This reflects the results in the Data Analysis section: at low frequencies the distances between the means of the object/no object classes are small, but so is the variance. This makes these frequencies very useful when there is no noise present. When noise with the same variance is added over the whole frequency span, however, the information at the low frequencies is more easily lost than at high frequencies.

Table 4.3: The matrix elements used for best accuracy. No pattern is evident. The reason for the many uses of reflection elements could very well be because of the information contained in that part of the data, but it could also be that they are fewer (4) than the transmission elements (6). The large number of parameters estimated for QDA, which grows quadratically with the number of features, is a possible explanation.

SNR	KNN	LDA	QDA	SVM	libSVM
∞	Transmission	All	All	All	All
40	All	All	Reflection	All	All
30	All	All	Reflection	All	All
20	Reflection	All	Reflection	All	All
15	Reflection	Reflection	Reflection	Transmission	All
10	Reflection	All	Reflection	Transmission	Transmission
5	All	Reflection	Reflection	Transmission	All

Finally the scatter matrix elements are studied. The results are not easily interpreted at this point. The reflection elements appear in total 11 times, compared to 5 for transmission. This could be indication that the reflection elements are more informative. On the other hand the KNN classifier is generally considered a poor choice when using high dimensional data, and the fact that QDA estimates a full covariance matrix for each class, could explain why it tends to benefit from using reflection elements only at this stage. The number of features when using reflection, transmission or all elements are 480, 720 and 1200 respectively, and the enhanced performance could be a consequence of the smaller set of features. It does not seem that any one set of scattering elements is generally preferable at this point.

4.2.3 Using equally spaced frequencies

In this section the results of using subsets of the frequency span consisting of equally spaced frequencies is presented. Again, the interest is both to compare classifiers and the effects of varying the subsets of frequencies and scattering matrix elements to use. The noise levels considered were the same as in the previous section. The number of frequencies considered were 10, 30, 60, 100, 150 and 200. Again the best obtained detection accuracy was for each noise level is compared between the classifiers, as seen in Figure 4.16.

In Figure 4.17 the number of frequencies used for the obtained accuracies are plotted. The general picture is that the SVMs perform well using many frequencies, while the DA classifiers benefit from reducing the dimension. This is especially clear when considering Table 4.4, since QDA uses reflection elements only except at the lowest SNR, while the libSVM uses all elements up to SNR = 10 dB, and transmission thereafter.

When comparing using 60 frequencies together or spread out, the results are better in the former case for the SVMs and KNN at all noise levels. The DA classifiers perform

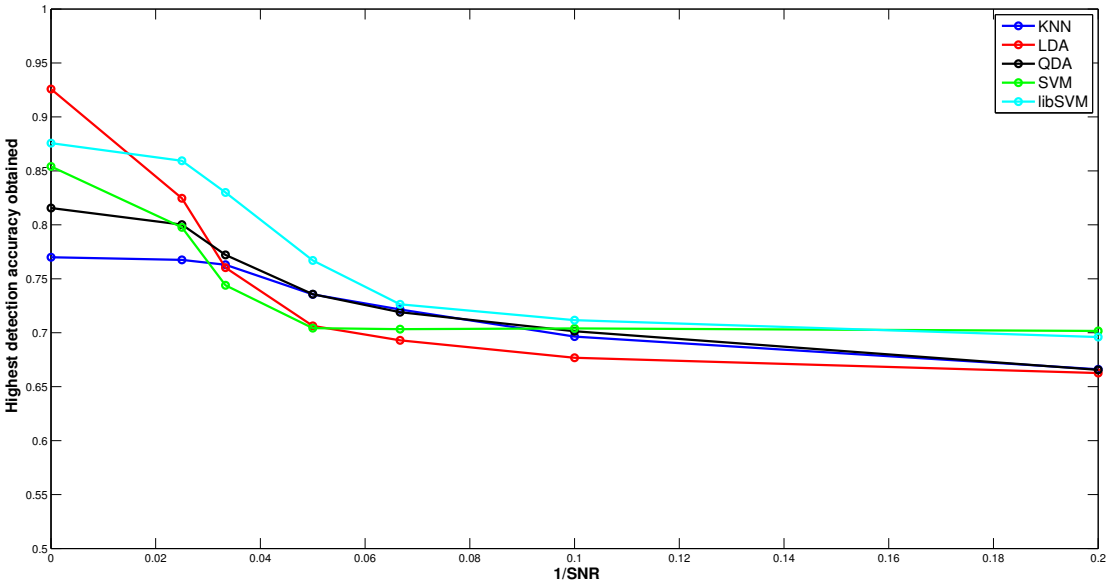


Figure 4.16: The best detection accuracies obtained using equally spaced frequencies over the whole available range. At the highest noise level the accuracies range from 0.67 to .70, which should be compared to classifying everything as object which would yield a detection accuracy of 2/3.

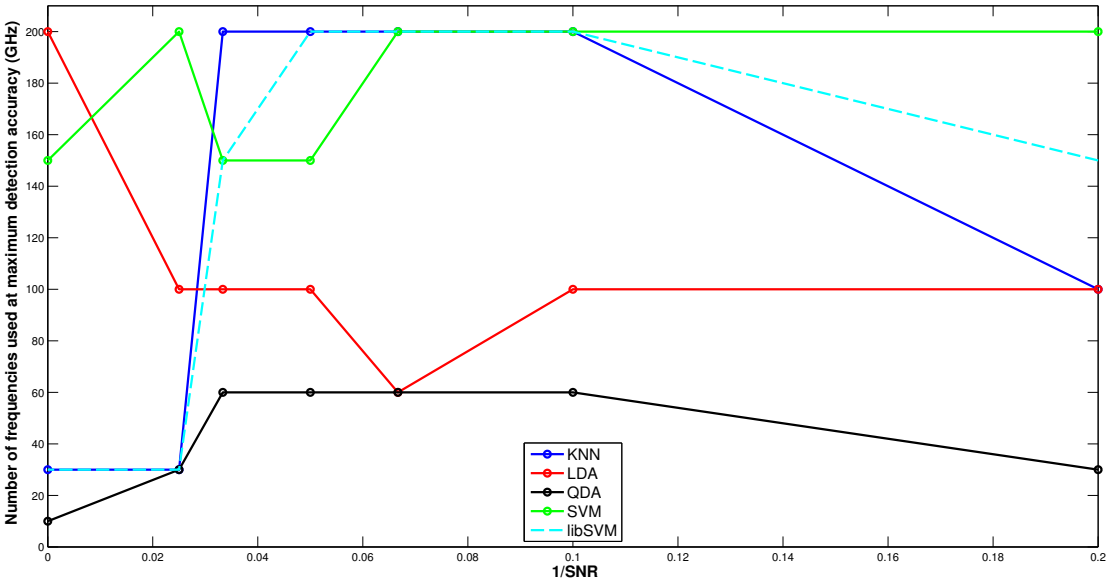


Figure 4.17: The benefits from dimension reduction when using QDA is quite clear. Both SVMs seems to gain from using many features, at least when noise is added.

Table 4.4: The matrix elements used for best accuracy. The result is very similar to that in Table 4.3.

SNR	KNN	LDA	QDA	SVM	libSVM
∞	Transmission	Transmission	All	Transmission	Transmission
40	All	All	Reflection	Reflection	Transmission
30	Reflection	All	Reflection	Reflection	All
20	All	All	Reflection	All	All
15	All	All	Reflection	All	All
10	Transmission	Reflection	Reflection	All	All
5	Transmission	Transmission	Reflection	All	All

better or at least as well using the latter option at the intermediate noise levels (SNR between 10 and 40 dB).

5

Conclusions

The data analysis shows that in the absence of noise the lowest parts of the frequency span gives the largest separation between signals from objects and background measurements, in relation to the corresponding standard deviations. The behaviour of this measure differs somewhat between object classes and scatter parameters, however. For metal objects the transmission elements at the very lowest part of the frequency span clearly yields the largest mean absolute value - standard deviation ratio for objects with a radius of 4 mm. For plastic objects with $\epsilon_r = 2$, the largest ratio for each scatter parameter is found below 1 GHz, even if not at the very lowest available frequency. In total this indicates that in the case of homogeneous background material and no measurement noise, the lowest part of the frequency span is most useful for detecting objects.

When using the same measure of separation, but considering the separation from signals from air bubbles, the picture is very similar for the metal objects, with even larger separation values obtained for the diametrical transmission elements. For the plastic objects however, the ratio varies around 0.85 over the whole frequency span, which only points to that distinction between these classes is bound to be difficult.

The plots of the separations also shows that information in the lowest frequencies is easily lost in the presence of noise. The absolute values of the separation at these frequencies are very small. It should be stressed again, however, that the assumption that the noise level is equal for all channels is not a proper model for all measuring devices.

The signal changes caused by variability in the background material was studied featurewise, and some non linear correlations were found between the real and imaginary parts. As for the use of different data transformations together with the simple classifier for this data, the differences in performance in terms of AUC values are in general small. Firstly, the results using the third transformation, taking linear and quadratic correlations into account, is uniformly better than using the second one, where only linear correlation is considered. Which of the first and the third transformation gives the best

detection, however, depends on the variability parameters of the background material and the amount of noise added. The third transformation consistently yields the best results at the lowest SNR value considered (0 dB). At higher SNR values, it also gives the best performance for the cases when at least one of correlation length and RMS is at the second lowest value considered (1.9 mm and 2 % respectively). The second part of this might be explained by the fact that the correlations are more prominent with little or no noise, while the circular symmetric noise tends to dominate at lower SNR values. The results at the lowest noise levels, however, is not in line with this explanation. There is another pattern appearing, when also considering the results for when the correlation length is 3.9 mm and the RMS 4 %, and that is that the third transformation performs the best when the AUC values are low (little above 0.5) for all transformation, though this has not been found helpful for finding an explanation for the behaviour.

Concerning the classifiers, the libSVM together with the discriminant analysis classifiers perform well when no noise is added. When adding noise, however, the libSVM is the only one showing some promising results. The fact that the kNN-classifier yields the best results at SNR levels below 30 dB should be interpreted as none of the classifiers are doing well at these noise levels, and the accuracy levels at these noise levels are well below what would be the requirements of a useful detector. The results also gives an indication that better accuracy is obtained using tightly spaced frequencies from a small part of the frequency span than widely spread frequencies from the whole span. The results from the data analysis are also reflected in the performance of the classifiers at different noise levels. When there is little or no noise, the lowest part of the frequency span clearly separates the object/no object classes and this is successfully used for classification. But at lower SNR values this information drowns in the noise added.

Bibliography

[1] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition, Springer Science+Business Media, 2009.

[2] M. Martínez-Ramón, C. Christodoulou, Support vector machines for antenna array processing and electromagnetics, Synthesis Lectures on Computational Electromagnetics 1 (1) (2005) 1–120.

[3] C.-W. Hsu, C.-J. Lin, A comparison of methods for multiclass support vector machines, Neural Networks, IEEE Transactions on 13 (2) (2002) 415–425.

[4] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27, software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[5] J. Milgram, M. Cheriet, R. Sabourin, et al., "one against one" or "one against all": Which one is better for handwriting recognition with svms?, in: Tenth International Workshop on Frontiers in Handwriting Recognition, 2006.

[6] R. Sorrentino, G. Bianchi, Microwave and RF Engineering, Wiley, Hoboken, NJ, USA, 2010.

A

Additional figures

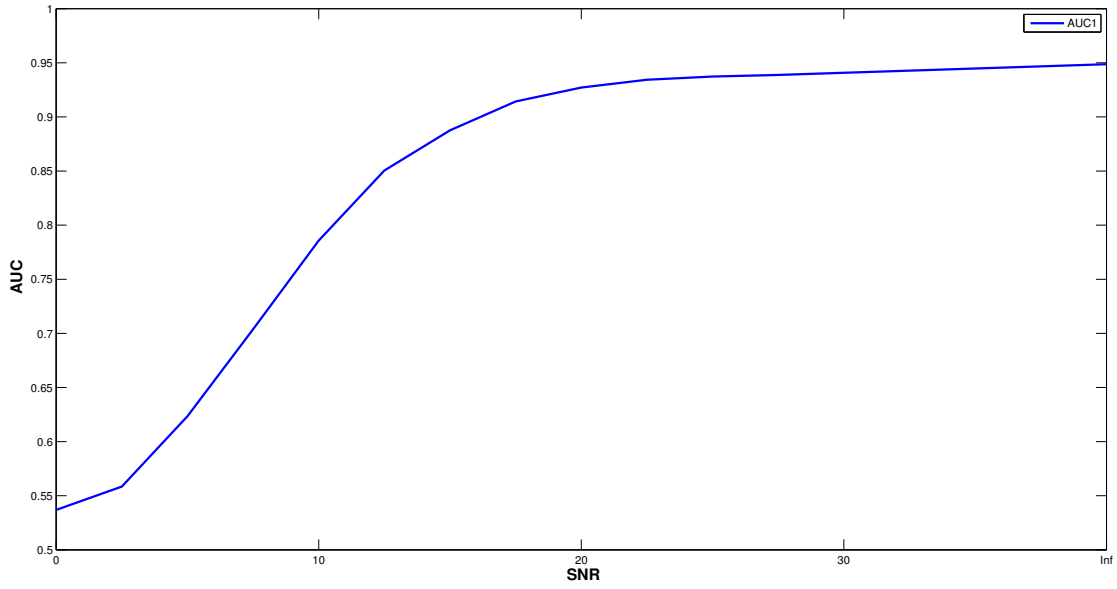


Figure A.1: AUC values using the simplest transformation for an object with $\epsilon_r = 2$, of radius 0.67 mm, for a correlation length of 1 mm RMS value of 2 %.

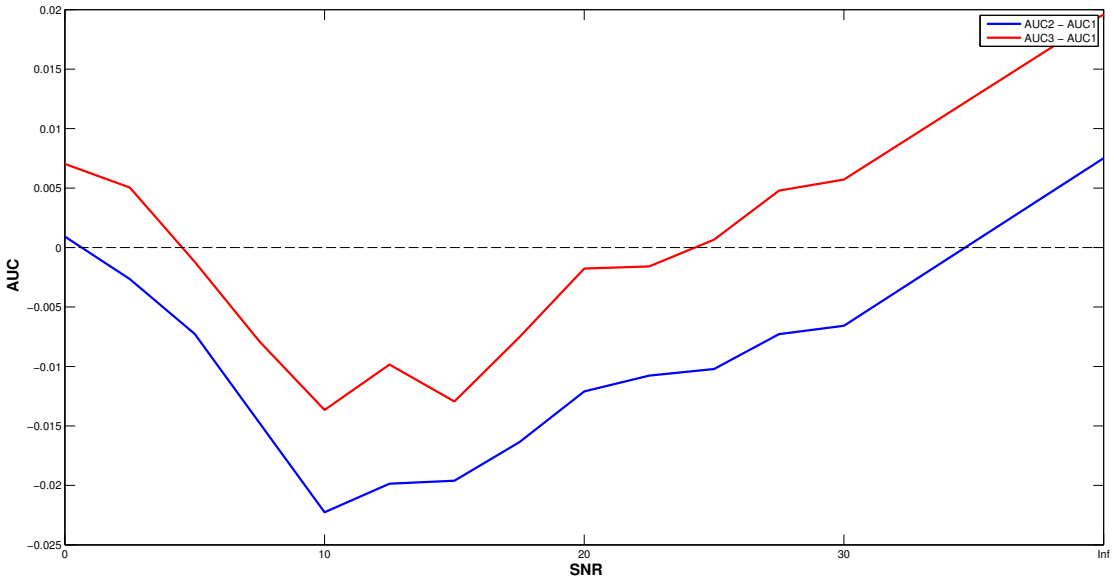


Figure A.2: Differences in AUC values using different transformation, for an plastic object of radius 0.67 mm. The correlation length is 1 mm and RMS value 2 %.

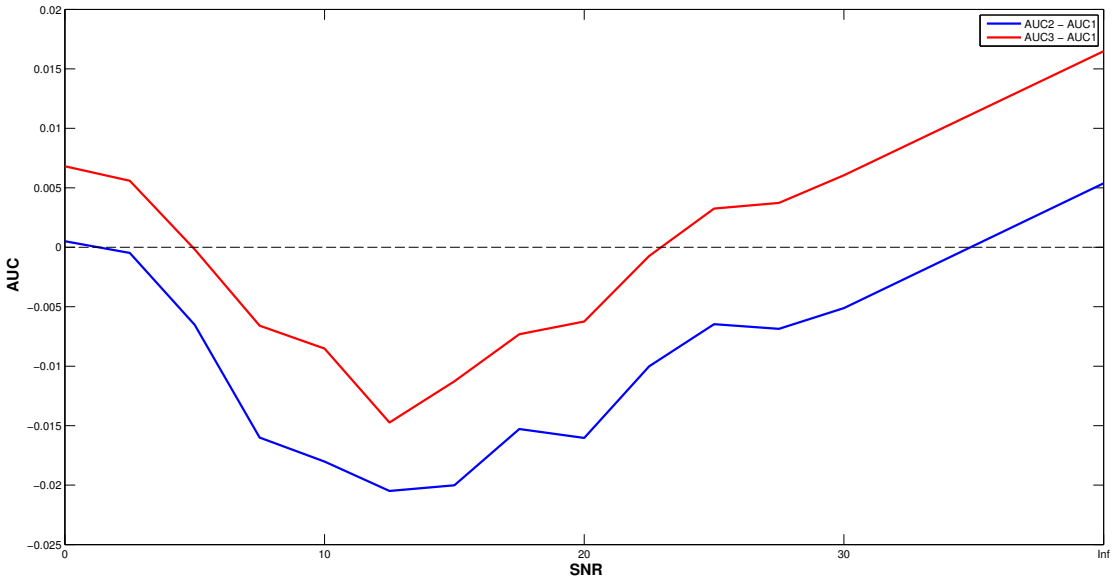


Figure A.3: Differences in AUC values using different transformation, for an air bubble of radius 0.67 mm. The correlation length is 1 mm and RMS value 2 %.

B

Resonance frequencies of a cylindrical cavity

An electromagnetic signal that propagates in a cylindrical cavity will at some frequencies give rise to resonances. The resonance frequencies are different for the electric and magnetic parts of the signal. Such resonances is the probable explanation of the peaks in standard deviation in the signal for the model with varying background material used in this work. In this section the formulas for calculating these frequencies are presented along with a comparison of the positions of the peaks in the standard deviation in the simulated signal response.

For the data used the only resonance frequencies for the electric signal are relevant, and the field patterns at these frequencies are called the *TE modes* of the cavity. The frequencies at which these modes occur are given by the following formula, as found in [6]:

$$f_{nmq} = \frac{c}{2\pi\sqrt{\epsilon_r}}\sqrt{\left(\frac{p'_{nm}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2} \tag{B.1}$$

where c is the speed of light in vacuum, ϵ_r the relative permittivity of the material in the pipe, p'_{nm} the m-th zero of the n-th Bessel function and a the radius of the cavity and d the length of the cavity. The cavity modelled can be considered to be of infinite length, and hence the second term under the square root is zero for all values of the index q. This then simplifies to

$$f_{nm} = \frac{c}{2\pi\sqrt{\epsilon_r}}\left(\frac{p'_{nm}}{a}\right) \tag{B.2}$$

Using the mean value of ϵ_r (60 F/m) and the value of a (3 cm), the frequencies for the TE modes can be computed and compared to the peaks in the standard deviation of the signal for the cavity modelled. In Table B.1 the computed resonance frequencies and

Resonance frequencies (GHz)	Standard deviation peaks (GHz)
0.3783	0.3907
0.6275	0.6015
0.7873	0.8050
0.8632	0.9068
1.0926	1.1103
1.0954	
1.3182	
1.3779	1.4083
1.4415	1.5536
1.6469	1.6627
1.7539	1.7571
	1.8371
1.9072	1.9243
2.0484	
2.0903	2.0769
2.1615	2.1714
	2.2877
2.3312	2.3604
2.4052	2.4040
	2.4185
2.6057	2.6511
2.7061	
2.7376	2.7456
2.8739	2.8837
2.9969	

Table B.1: The resonance frequencies and the standard deviation peaks approximately matched. The matches are not perfect, and in some cases a resonance frequency does not have a corresponding standard deviation peak or vice versa. The correspondence is, despite this, clear.

obtained standard deviation peaks are listed sorted by frequency. The matches are far from perfect, but the pattern is still clear. The figure B.1 is gives a graphical comparison.

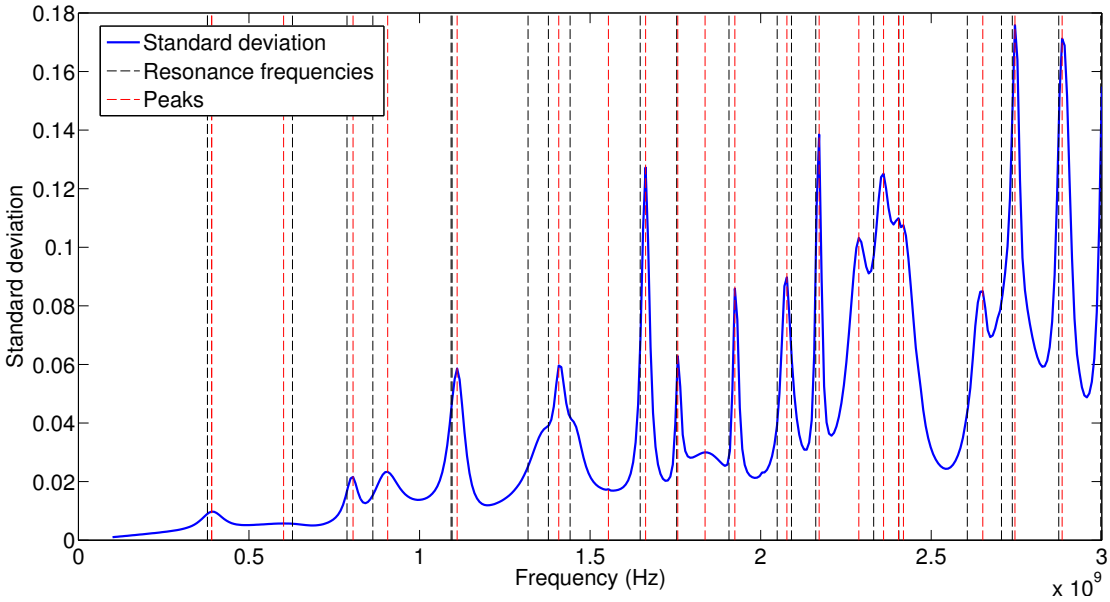


Figure B.1: Plot of the standard deviation with resonance frequencies and peaks indicated by vertical lines.