



Zentrum für Technomathematik
Fachbereich 3 – Mathematik und Informatik

**Regression Models for Ultrasonic
Testing of Carbon Fiber Reinforced
Polymers**

**Carsten Brandt, Marco Hamann,
Johannes Leuschner**

Report 19–01

Berichte aus der Technomathematik

Report 19–01

September 2019

Contents

Preface	ii
1 Introduction	1
1.1 Carbon fiber reinforced polymers	1
1.2 Ultrasonic testing for CFRP	1
1.3 Task	4
2 Evaluation criteria	5
2.1 Normalized confusion matrix	5
2.2 Receiver operating characteristic curve	6
2.3 Correlation	7
3 Regression models	8
3.1 Time-linear Regression	9
3.2 FFT-linear Regression	9
3.3 Wavelet-based Linear Regression Models	13
4 Evaluation	14
4.1 Dataset	14
4.2 Fourier transform	14
4.3 Time shifts tested on time-linear regression	15
4.4 Performance test of basic FFT-linear regression including time shifts	15
4.5 Comparison of FFT-linear regression methods	18
4.6 Comparison of Wavelet-based Linear Regression Models	20
4.7 Time shifts for FFT-linear and Wavelet-based linear regression models	21
4.8 Validation on different data sets of a specimen	21
4.9 Restriction to parts of the intermediate echo gate	25
4.10 Regression coefficients	26
4.11 Regularization of the regression	29
5 Conclusion	32
Bibliography	33
A Results for lower reference threshold	34

Preface

This report shows results of regression models for the evaluation of ultrasonic testing of porosity in carbon fiber reinforced polymers, based on work in a modeling project within the master degree studies in Industrial Mathematics at the University of Bremen. The authors thank the Airbus Operations GmbH in Bremen (for providing samples and ultrasonic testing equipment for the investigations), Kai-Wah Chan, Matthias Rick, Jim van Kleef and Matthias Otten (for work in two former modeling projects) and especially Professor Dr. Dr. h.c. Peter Maaß as supervisor at the Center for Industrial Mathematics, faculty of Mathematics and Computer Science, University of Bremen.

Chapter 1

Introduction

Before describing the task of the work presented in this report in section 1.3, a brief overview about carbon fiber reinforced polymers and ultrasonic testing is given in section 1.1 and 1.2.

1.1 Carbon fiber reinforced polymers

Carbon fiber reinforced polymers (CFRP) are layered composite materials. The layers consist of carbon fibers, either aligned unidirectional or interwoven, and are embedded into a synthetic resin matrix. The

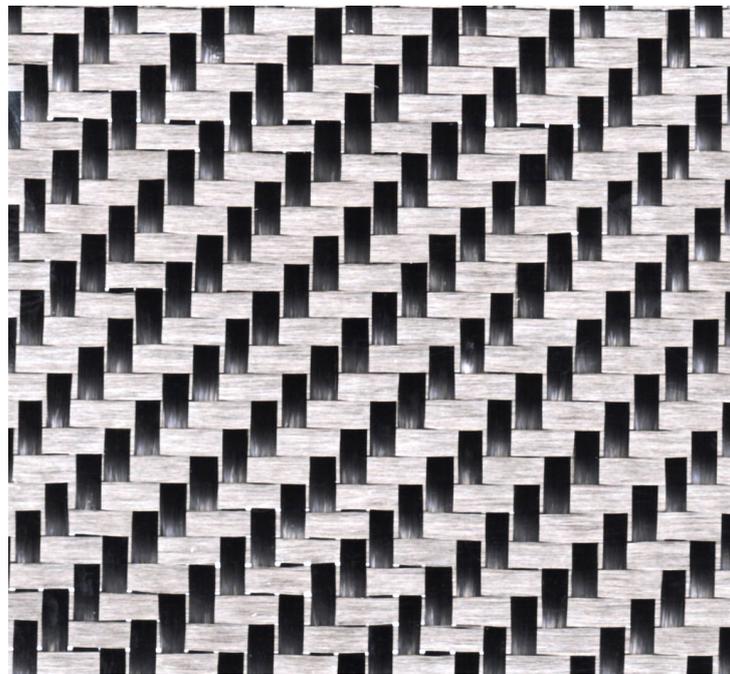


Figure 1.1: Carbon fiber fabric with interweaved structure. Image: [5]

main benefit of CFRP compared to conventional metallic materials is their low density and nevertheless high loading capacity, especially in fiber direction. More specifically the density of CFRP is by a factor near 5 lower than the one of steel, while the tensile strength in fiber direction is comparable (cf. [9, p. 2ff.]).

Because of these properties CFRP is in great demand, among others for the application in aircrafts. Decrease in weight by 1 kg results in fuel saving of up to 2000l during an airplane lifetime [10]. At Airbus the usage of composites like CFRP has increased heavily over the last decades (figure 1.2).

1.2 Ultrasonic testing for CFRP

The goal of *non-destructive testing* (NDT) is to assess the quality while preserving the functionality. In contrast, destructive testing techniques like tensile or hardness tests usually result in the test object being

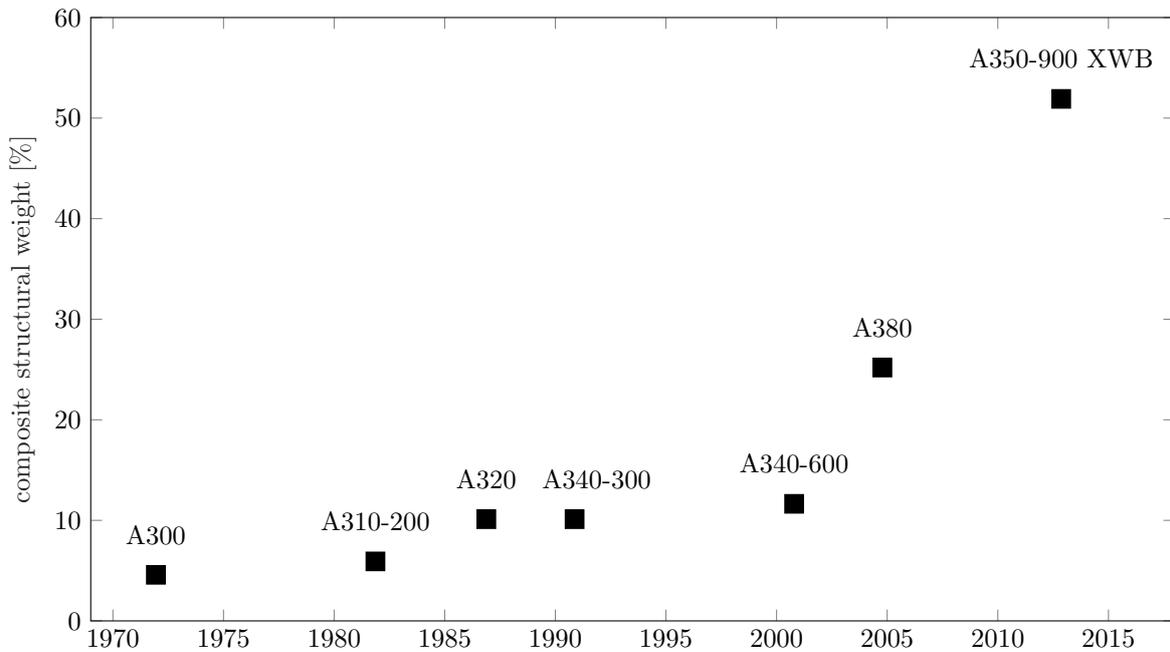


Figure 1.2: Increasing quantity of composites at Airbus aircrafts
Graphic based on [1]

destroyed or damaged. NDT allows not only for testing during the development process or in samples, but also for inspection of products and parts in use or ready for use.

One common NDT method is *ultrasonic testing*. It can be applied on CFRP by scanning over the surface in a grid pattern to detect porosity (figure 1.3).

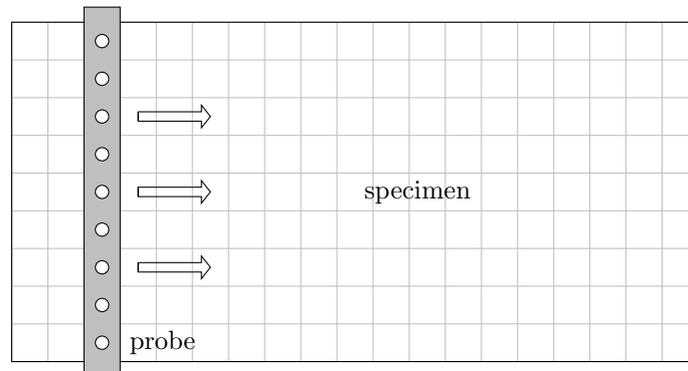


Figure 1.3: Scanning of a specimen. The phased array probe is moved horizontally over the surface.

The probe operates by phased array technique, which means that it contains an array of ultrasonic pulse emitting elements, which can be delayed individually. The same elements also receive the ultrasound echoes. By subsequent firing it is possible to sweep through the material in the direction of the array. Thus in the scanning process (figure 1.3) the vertical resolution is obtained from multiple measurements of a software-controlled linear scan. By moving the probe horizontally over the surface and repeating these measurements at multiple locations the horizontal resolution is achieved. The inspection is performed with a coupling medium (often water) between probe and sample.

Finally the result of an ultrasonic scan are the echo time series at each point of the grid. Each of those is called an A-scan and holds the measured amplitudes for a finite time interval starting from the time of the emission of the pulse (figure 1.4).

In CFRP an ultrasonic pulse gets reflected at the front and the back of each layer due to the high differences of acoustic impedance (cf. [8, p. 15f.]). This also applies to pores, but here the pulse also gets scattered due to their shape or rough surface.

Currently used classification technique

Usually two quite strong reflections occur when inspecting a CFRP part of simple geometry. The first reflection happens at the surface of the specimen (front-wall) directly at the probe, which emits the signal. The second reflection occurs at the back-wall, where the test material ends and the neighboring material follows. In figure 1.4 the peak of the back-wall echo (BWE) of an A-scan is marked by a green circle.

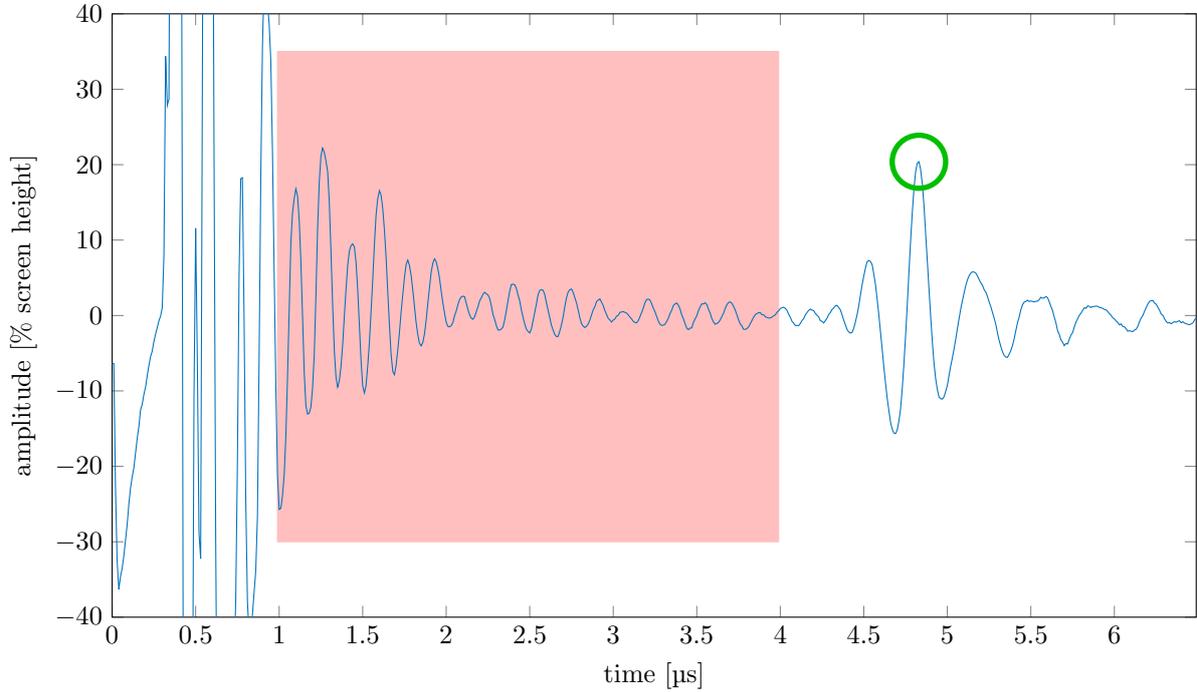


Figure 1.4: Example of an A-Scan. The red rectangle marks the intermediate echoes and the green circle marks the back-wall echo.

If the specimen is porous, the pulse gets reflected and scattered at the pores before it reaches the opposite surface (back-wall). Therefore the intensity is measurably smaller than it is for non-porous specimens. The currently used method to classify each point on the grid as porous or non-porous only depends on the amplitude of the BWE. At all points with a BWE lower than a reference threshold the specimen is assumed to be porous (provided no other defects like delaminations—material separations of larger size than pores—are present). All other points are considered to be non-porous. The reference threshold usually is determined as a fixed fraction (e.g. $-6 \text{ dB} \approx 1/2$ or $-12 \text{ dB} \approx 1/4$) of the mean of the BWE height in a non-porous region.

The BWE of a CFRP specimen can be visualized in a so called C-scan. One kind of C-scan shows the BWE amplitudes at each point on the grid using a color map (figure 1.5).

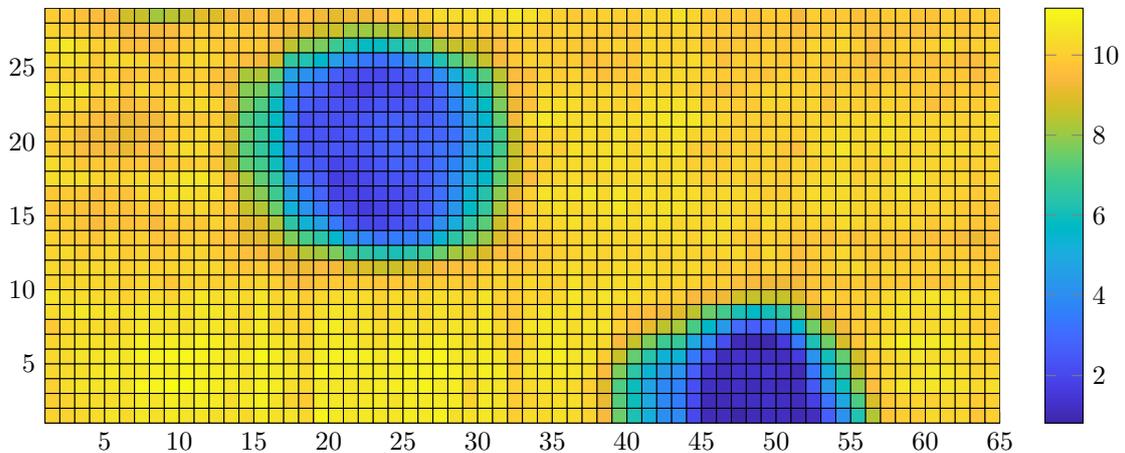


Figure 1.5: Example of a C-scan

1.3 Task

Obviously it is not possible to apply the currently used technique if there is no constant back-wall echo. This is the case for example if the back surface of the material is bonded to another material of similar density, if the back-wall is not parallel to the front wall, such that the pulse reflection does not return to the probe, or for sandwich materials. Such parts, for which it is currently not possible to measure the porosity content, have to be designed in a way that they would have sufficient strength even if being porous, which leads to a gain of weight (cf. [3]).

Hence an alternative classifier only using the signal before the back-wall echo is sought-after.

Since the front-wall echo does not contain valuable information, only the part of the signal between both walls, the intermediate echo, is relevant. In figure 1.4 the intermediate echoes of an A-scan are marked by a red rectangle.

The approach we follow here is to calculate a back-wall echo equivalent, on which again a threshold classifier can be applied. For the task of finding a back-wall echo equivalent we use regression methods. Afterwards a threshold is determined by optimizing for a specific quality measure of the classification, evaluated on the training data.

In this report we will present some regression models and statistically evaluate the classifiers.

Chapter 2

Evaluation criteria

In order to rate the quality of the classifiers we are going to present in chapter 3 for the application of porosity in CFRP, we have to use specimens with known ground truth. Therefore we use specimens providing back-wall echoes, by which the class can be determined using the conventional classification technique. This classification is assumed to represent the truth, to which we compare the classification made by the new classifiers, which only use the intermediate echoes of the same specimens.

Section 2.1 and 2.2 give background about means to assess classifiers; section 2.3 shows the background for correlation.

2.1 Normalized confusion matrix

A detailed way to present the classification accuracy is the normalized confusion matrix, showing the rates at which all combinations of true and predicted classes occur.

true positive rate:

$$\begin{aligned} \text{TPR} &= P(\text{predicted to be porous}|\text{in fact porous}) \\ &= \frac{P(\text{predicted to be porous and in fact porous})}{P(\text{in fact porous})} \end{aligned}$$

false negative rate:

$$\begin{aligned} \text{FNR} &= P(\text{predicted to be non-porous}|\text{in fact porous}) \\ &= \frac{P(\text{predicted to be non-porous and in fact porous})}{P(\text{in fact porous})} \end{aligned}$$

false positive rate:

$$\begin{aligned} \text{FPR} &= P(\text{predicted to be porous}|\text{in fact non-porous}) \\ &= \frac{P(\text{predicted to be porous and in fact non-porous})}{P(\text{in fact non-porous})} \end{aligned}$$

true negative rate:

$$\begin{aligned} \text{TNR} &= P(\text{predicted to be non-porous}|\text{in fact non-porous}) \\ &= \frac{P(\text{predicted to be non-porous and in fact non-porous})}{P(\text{in fact non-porous})} \end{aligned}$$

		True Class	
		<i>porous</i>	<i>non-porous</i>
Predicted Class	<i>porous</i>	TPR	FPR
	<i>non-porous</i>	FNR	TNR
Σ		100 %	100 %

Table 2.1: Normalized confusion matrix for porosity tests

The prediction quality is good if the entries on the main diagonal, TPR (also called sensitivity) and TNR (also called specificity), are close to 100%. The distribution of the rates in the normalized confusion matrix depends on the threshold used for the prediction. If for example the threshold is lower than every

regression value, all sample points are predicted to be non-porous, meaning $\text{TPR} = 0\%$, $\text{FPR} = 0\%$, $\text{FNR} = 100\%$ and $\text{TNR} = 100\%$. With increasing threshold TPR and FPR increase and FNR and TNR decrease. If the threshold finally is higher than every regression value, all sample points are predicted to be porous, meaning $\text{TPR} = 100\%$, $\text{FPR} = 100\%$, $\text{FNR} = 0\%$ and $\text{TNR} = 0\%$.

Our goal is to achieve $\text{FPR} < 10\%$ while maintaining $\text{TPR} > 90\%$. From the safety point of view, it is important to have a low FNR , because it measures how many porous sample points are not detected as such. By $\text{TPR} + \text{FNR} = 100\%$ this is equivalent to a high TPR .

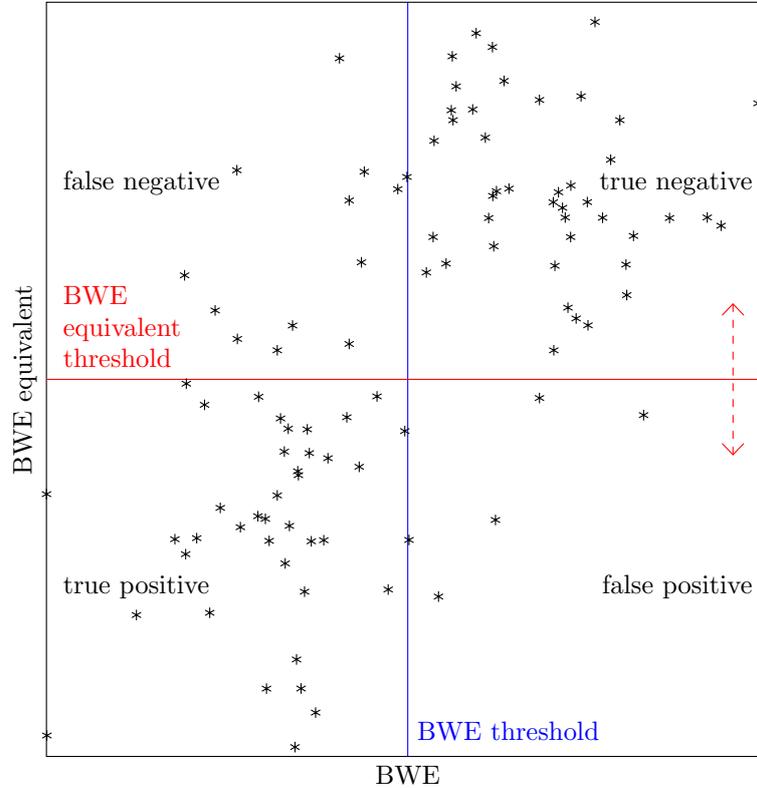


Figure 2.1: Visualization of confusion matrix for threshold-based classification

Balanced accuracy

One measure commonly used for classification evaluation is the accuracy, defined as the part of correctly classified points,

$$\begin{aligned} \text{acc} = & P(\text{predicted to be porous and in fact porous}) \\ & + P(\text{predicted to be non-porous and in fact non-porous}). \end{aligned}$$

While this intuitively reflects the quality of the class prediction, it does not take into account the size of the true classes. This can be very misleading if the classes have drastically different sizes, which is usually the case for the application of quality assurance. Since the class of the defect sample points is typically small, a quite good accuracy would be achieved by the simple classifier that predicts every sample point to be non-defect.

An alternative accuracy measure, which does take the class sizes into account, is the balanced accuracy, defined as the mean of the diagonal of the normalized confusion matrix,

$$\text{balAcc} = \frac{1}{2}(\text{TPR} + \text{TNR})$$

(cf. [6]).

2.2 Receiver operating characteristic curve

One way to display the classification quality for various thresholds is the receiver operating characteristic curve (ROC curve). It is given by the coordinates $(x, y) = (\text{FPR}, \text{TPR})$. Both rates increase monotonically with the threshold; the resulting curve is the graph of a monotonically increasing function

$c : [0; 1] \rightarrow [0; 1], \text{FPR} \mapsto \text{TPR}$. The choice of a threshold is equivalent to picking a point from the ROC-curve. One usually looks for a point in the top left corner.

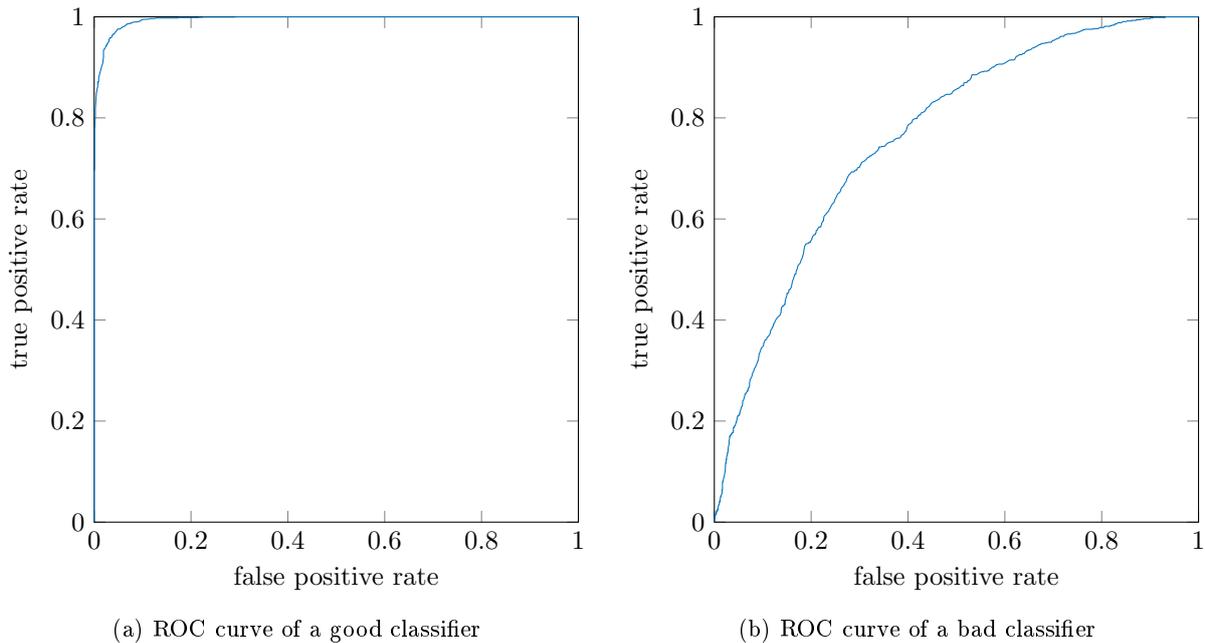


Figure 2.2: ROC curve examples

The *area under curve* (AUC) is a quality measure that is calculated based on the ROC curve by integrating,

$$\text{AUC} = \int_0^1 c(\text{FPR}) \, d\text{FPR}.$$

It significantly differs from the rates in the normalized confusion matrix and the balanced accuracy in that it summarizes their values for all possible thresholds. Desirably the TPR should reach high values when the FPR is still low. This results in the curve bending strongly to the top left and a high area under curve, like shown in figure 2.2a. The AUC in figure 2.2a is 99.5% and the AUC in figure 2.2b is 76.3%. If the curve is close to a straight line, the classification is not meaningful.

2.3 Correlation

Another evaluation approach is not to consider the true and predicted classes, but to directly compare the BWE p and its equivalent q . The Pearson correlation coefficient, defined by

$$\text{corr} = \frac{\text{cov}(p, q)}{\sigma_p \sigma_q},$$

measures the linear dependence of variables (with cov being the covariance and σ denoting the standard deviation). For a finite series of n measurements it can be calculated by

$$\text{corr} = \frac{\sum_{i=0}^{n-1} (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\left(\sum_{i=0}^{n-1} (p_i - \bar{p})^2\right) \left(\sum_{i=0}^{n-1} (q_i - \bar{q})^2\right)}}.$$

Here \bar{p} and \bar{q} denote the mean values of p and q .

The correlation coefficient can take values in the interval $[-1, 1]$. For the purpose of obtaining a good equivalent of the BWE, a large correlation coefficient value (close to 1) is desired.

Chapter 3

Regression models

The general linear regression approach is based on the assumption that the BWE p can be approximated by

$$p \approx F(E)x,$$

where

- $E = \begin{bmatrix} b_1^T \\ \vdots \\ b_n^T \end{bmatrix} \in \mathbb{R}^{n \times m_s}$ row-wise contains the intermediate echoes b_k , $k = 1, \dots, n$ of length m_s ,
- $F : \mathbb{R}^{n \times m_s} \rightarrow \mathbb{R}^{n \times m}$ is the feature map, and
- $x \in \mathbb{R}^m$ is a vector of coefficients.

We formulate the model as the minimization problem

$$x = \arg \min_x \|F(E)x - p\|_2. \quad (3.1)$$

This is a typical linear least squares problem for the coefficients x . Afterwards the BWE equivalent for new test data \tilde{E} can be calculated by

$$q = F(\tilde{E})x.$$

A simple way to include a trainable constant offset in this model is to add a 1-column to $F(E)$ and $F(\tilde{E})$. The offset is then given by an additional entry of x . We use this approach for all models.

Finally a threshold has to be chosen that is applied on the BWE equivalent for classification. If not stated otherwise, the classification threshold is determined by maximizing the balanced accuracy of the resulting classifier on the training data E .

Solution of the linear least squares problem The main task is to solve

$$x = \arg \min_x \|Ax - p\|_2 = \arg \min_x \frac{1}{2} \|Ax - p\|_2^2.$$

This problem is always solvable in contrast to the possibly overdetermined system of linear equations $Ax = p$. We assume x to be an arbitrary solution if it should be non-unique. The following methods for solving this linear least squares problem can also be found in [4, p. 890ff.].

First we note that the objective is convex. Hence we can minimize it by finding the x for which its derivative is zero,

$$\frac{\partial}{\partial x} \frac{1}{2} \|Ax - p\|_2^2 = A^T(Ax - p) = 0.$$

This leads to the normal equation

$$A^T A x = A^T p, \quad (3.2)$$

which is a linear system of equations and can be solved by well known methods.

This method is numerically disadvantageous, because the condition of the system matrix $A^T A$ depends in a quadratic manner on the condition of A . An alternative is to compute the QR decomposition $A = QR$ with an orthogonal matrix Q and an upper triangle matrix R . Then the linear system of equations $Rx = Q^T p$ can be solved by back substitution for x . In practice we use the backslash operator $x = A \backslash p$ provided by Matlab[®], which is also documented to solve the least-squares approximation if the system matrix A is not square.

In the next section 3.1 we introduce the 'time linear regression model' as the most classical approach to our application. It serves as state-of-the-art comparison with the following, more elaborated model. In section 3.2 we introduce our FFT-based linear regression model, and section 3.3 contains a description of the competing wavelet model as introduced in [7].

3.1 Time-linear Regression

The simplest approach is to assume a linear connection between the columns of the intermediate echo matrix E and the BWE vector p . This can be achieved by choosing the identity for the feature map F in (3.1),

$$F(E) = E,$$

leading to the *time-linear regression problem*

$$x = \arg \min_x \|E x - p\|_2.$$

The BWE equivalent for new test data \tilde{E} is then given by

$$q = \tilde{E} x.$$

A drawback of this approach is the following: The time-linear regression assigns fixed coefficients to all time points and thus to all depths in the material. In case of a time shift due to a varying distance of the probe to the surface the assignment changes, leading to the coefficients being mapped to other time points than they were trained for.

3.2 FFT-linear Regression

A new idea developed in this work is to apply linear regression on the amplitude spectrum, i.e. on the magnitudes of the Fourier transform of the time series. The amplitude spectrum is invariant regarding time shifts for periodic time series, which is an advantageous property for our application. Therefore we choose $F(E)$ to row-wise compute the absolute value of the discrete Fourier transform (FFT)¹ of E ,

$$(F(E))_{k,l} = \left| \sum_{j=0}^{m_s-1} E_{k,j} e^{-i2\pi \frac{j l}{m_s}} \right|, \quad k = 0, \dots, n-1, \quad l = 0, \dots, \left\lfloor \frac{m_s}{2} \right\rfloor. \quad (3.3)$$

Note that we cut off the FFT after $m = \lfloor \frac{m_s}{2} \rfloor + 1$ values, because the signal is real and thus the remaining absolute values of the FFT are symmetric to the first ones and do not provide further information. Now the problem reads

$$x = \arg \min_x \left\| \left(\sum_{l=0}^{m-1} \left| \sum_{j=0}^{m_s-1} E_{k,j} e^{-i2\pi \frac{j l}{m_s}} \right| x_l \right)_{k=0, \dots, n-1} - p \right\|_2$$

and the solution can be used to calculate the BWE equivalent by

$$q_k = \sum_{l=0}^{m-1} \left| \sum_{j=0}^{m_s-1} E_{k,j} e^{-i2\pi \frac{j l}{m_s}} \right| x_l.$$

¹The discrete Fourier transform is usually computed by the fast Fourier transform (FFT) algorithm.

Weighting in time domain

Next we consider some custom weightings of the signal in the time domain. With the choice of a row vector $w \in \mathbb{R}^{1 \times m_s}$ of weighting coefficients and an arbitrary function \tilde{F} , possibly one of those proposed for F above, we obtain another possibility for F ,

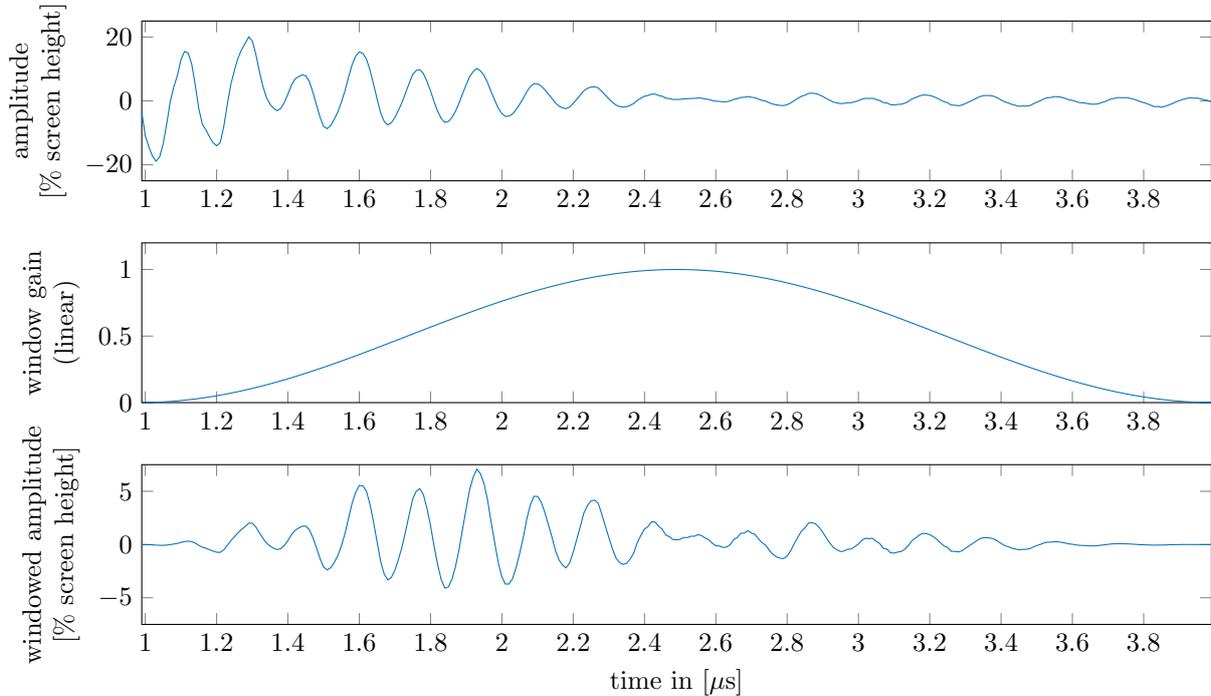
$$F(E) = \tilde{F}(E \circ W), \quad W = \begin{bmatrix} w \\ \vdots \\ w \end{bmatrix}.$$

Here “ \circ ” denotes point-by-point multiplication. While this approach does not make sense for the time-linear Regression, as the linear coefficients x would just be optimized to compensate for the change, one can expect an influence if F is more complex in the sense that it does not just compute a linear mapping $F(E) = EA$ with a matrix A . We introduce two weightings for the FFT-amplitude-function (4.1) now.

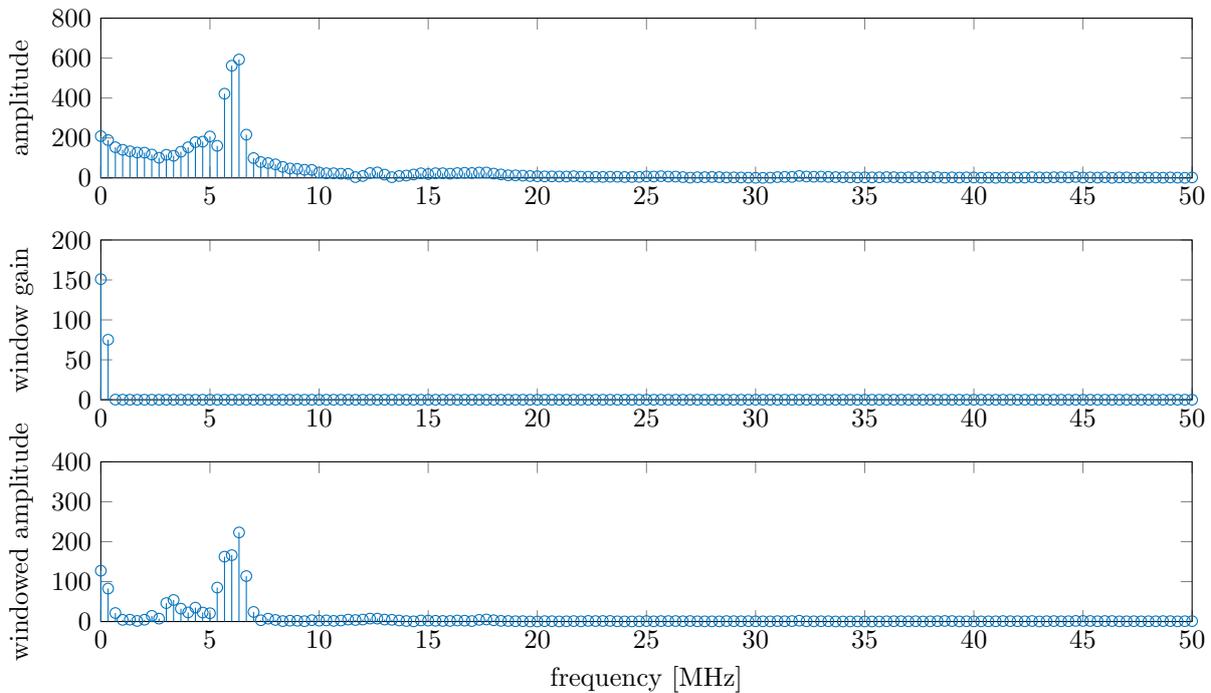
Windowing

When applying the FFT the finite time signal is implicitly continued infinitely periodically. In general, as in our application, we have to expect a discontinuity after each period. This leads to distortions in the frequency domain, known as the leakage effect. To overcome this problem one can replace the rectangular window function used when just taking a finite time signal with a window function, which also applies a weighting on the samples.

In particular we apply the Hann window (cf. [11, p. 239ff.]) on the intermediate echoes. It smoothly forces the values at the beginning and the end of the window to get near zero, achieving a signal that is well continuable periodically, because the boundary values in the window match (figure 3.1).



(a) Time signals

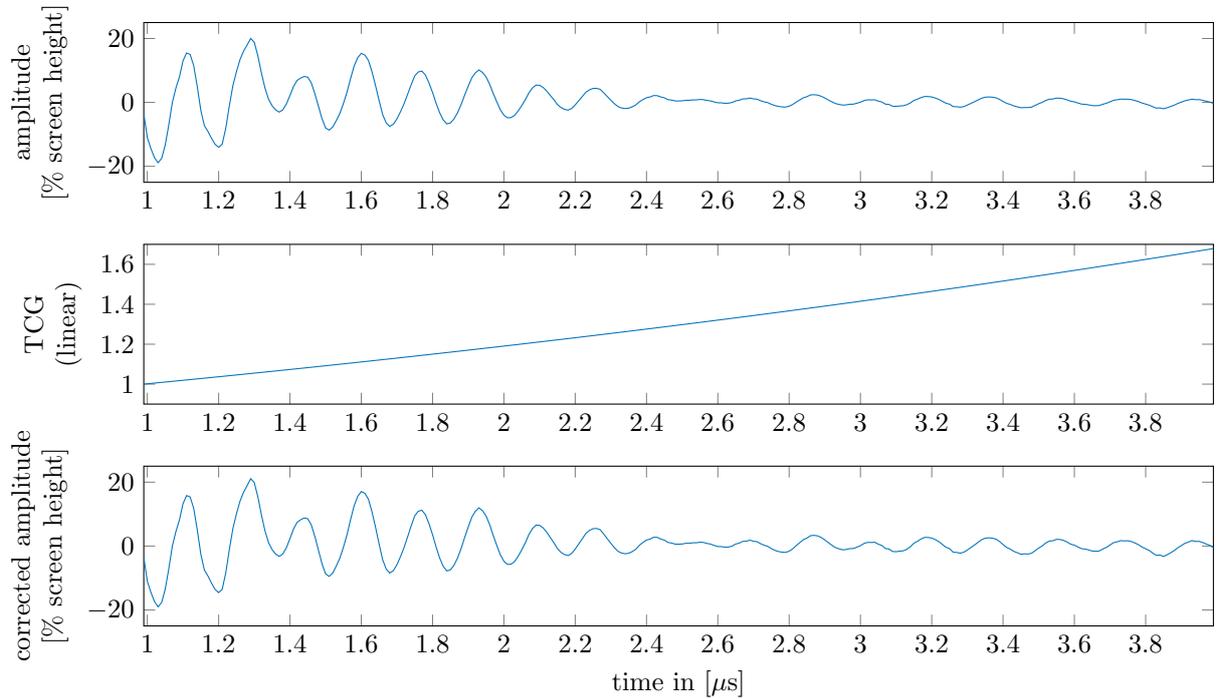


(b) Magnitudes of frequency spectra

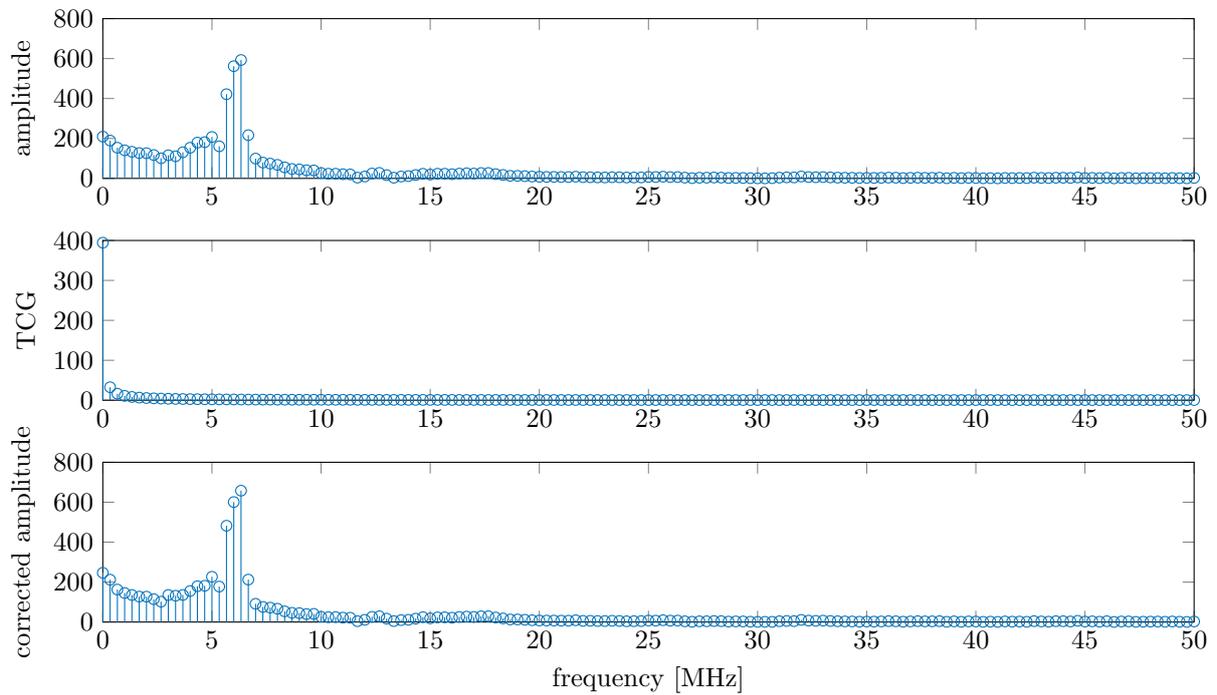
Figure 3.1: Windowing of an A-scan: in both sub-figures original at the top, Hann window in the middle and windowed at the bottom.

Time corrected gain

The material is known to damp the ultrasound signal by a fixed rate of approximately $1.5 \text{ dB}/\mu\text{s}$ (figure 3.2). Since this effect is independent of the damping and reflecting effect of pores, compensating for this effect is a useful preprocessing step for porosity detection, very common in ultrasonic testing of CFRP.



(a) Time signals



(b) Magnitudes of frequency spectra

Figure 3.2: TCG of an A-scan: in both sub-figures original signal at the top, TCG function in the middle and corrected signal at the bottom.

3.3 Wavelet-based Linear Regression Models

A different approach utilizing wavelet transformations for constructing a BWE equivalent is proposed by Ki-Bok Kim et al. in [7]. Similar to the exponential functions in Fourier transforms, wavelets are another basis of the function space L^2 , which have the property to be localized not only in the frequency domain but also in the time domain. In this sense wavelet transforms can be compared to short time Fourier transforms. The discrete wavelet transform (DWT) is based on the multiresolution analysis, which allows to split a signal up into finer and coarser scales. In [7] wavelets from the Daubechies wavelet family (figure 3.3) are chosen, the DWT was calculated up to level 5, yielding the two signals A_5 and D_5 . According

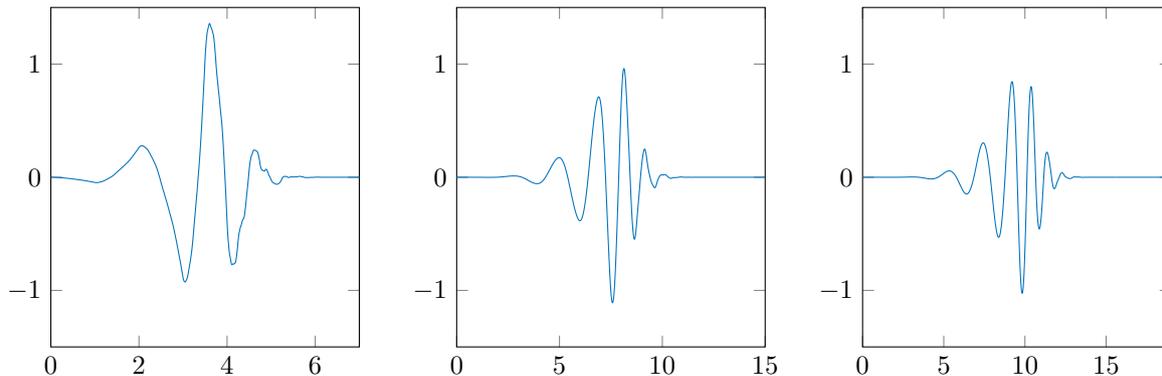


Figure 3.3: The wavelet functions $db4$, $db8$ and $db10$ (from left to right) of the Daubechies wavelet family.

to [7, p. 13, l. 32–36], “the correlations between ultrasonic parameters and the porosity content suggest that it is possible to predict the porosity content of the composite materials by measuring their peak amplitudes and magnitudes of peak frequency of the backscattered signal, A_5 and D_5 ”. Therefore they suggest the linear regression model

$$q_k = \underbrace{[1, \ln(A_{A_5,k}), \ln(M_{A_5,k}), \ln(A_{D_5,k}), \ln(M_{D_5,k})]}_{k\text{-th row of } F(E)} \cdot [x_1, \dots, x_5]^T \quad (3.4)$$

(cf. [7, eq. (9)]). Here $A_{X,k}$ denotes the peak amplitude of the signal $X \in \{A_5, D_5\}$ for A-scan k and M_X denotes the magnitude of peak frequency. In [7] the $db4$ wavelet was chosen, since the evaluated wavelets $db4$ up to $db10$ showed similar results.

We consider four wavelet-based models, the first one being (3.4) with the $db4$ wavelet. The other models utilize more levels,

$$q_k = [1, \ln(A_{A_1,k}), \ln(M_{A_1,k}), \ln(A_{D_1,k}), \ln(M_{D_1,k}), \dots, \ln(A_{A_7,k}), \ln(M_{A_7,k}), \ln(A_{D_7,k}), \ln(M_{D_7,k})] \cdot [x_1, \dots, x_{29}]^T, \quad (3.5)$$

where the $db4$, $db8$ and $db10$ wavelets are chosen for the second, third and fourth model, respectively. The consideration of all signals A_1, D_1 up to A_7, D_7 enables the linear regression to choose the relevant level(s) instead of picking level 5 a priori. Like in (3.4), $[1, \ln(A_{A_1,k}), \dots, \ln(M_{D_7,k})]$ constitutes the k -th row of $F(E)$ here.

Chapter 4

Evaluation

In this chapter we present classification results of the different time series analysis methods described in chapter 3 on ultrasonic measurements of CFRP samples without and with porosity. After describing the dataset in section 4.1 we motivate the investigations by showing that standard FFT results do not lead to sufficient classification results in section 4.2. The influence of time shifts on results with linear regression in the time domain in section 4.3 are followed by investigations to the main results of this work, FFT-linear regression in section 4.5 and 4.6, and to the wavelet approach in 4.6. Time shifts, validation on other data sets and the use of smaller time series are investigated in section 4.7, 4.8 and 4.9; section 4.10 and 4.11 look at the regression coefficients and regularization.

4.1 Dataset

We apply the methods proposed in chapter 3 on ultrasonic scans of industrially relevant specimens kindly provided by Airbus Operations GmbH. The measurements have been taken with a phased array probe with a center frequency of 5 MHz. Parameters and measurement process are given in [2].

For each of two CFRP specimens, which we name B50 and B56, we have three scans, respectively. These specimens have a thickness of 6 mm and a layer thickness of approximately 0.25 mm. Each scan holds 4060 A-scans at the points of a 29×140 grid, each A-scan sampled with $f_A = 100$ MHz for 9.98 μ s. A true BWE is contained in these A-scans, which is located between 4 μ s and 6 μ s. We consider the part from 1 μ s to 4 μ s as intermediate echoes. This part is not influenced by the back-wall, since the ultrasound takes longer to reach the back-wall (and return to the probe afterwards).

For the reference threshold, which is applied on the true BWE to obtain the ground truth, we choose -6 dB for our evaluation. The results for -12 dB are included in appendix A.

The two CFRP specimens differ in their general amount of porosity. While B50 has a good quality, B56 is porous. In order to contain both in one dataset and for the purpose of having distinct training and evaluation parts, we merge half of the scans of B50 and B56 (one half containing 29×70 A-scans). We call the combination of the left halves B50+56l and the combination of the right halves B50+56r. This allows for training on B50+56l and evaluation on B50+56r and vice versa.

4.2 Fourier transform

Before presenting results of regression models, a short motivation shall be given in this section. A common approach in signal processing is using Fourier transform. We use the absolute value of the discrete Fourier transform (FFT)¹ of E ,

$$(F(E))_{k,l} = \left| \sum_{j=0}^{m_s-1} E_{k,j} e^{-i2\pi \frac{j l}{m_s}} \right|, \quad k = 0, \dots, n-1, \quad l = 0, \dots, \left\lfloor \frac{m_s}{2} \right\rfloor. \quad (4.1)$$

Note that we cut off the FFT after $\lfloor \frac{m_s}{2} \rfloor + 1$ values, because the signal is real and thus the remaining absolute values of the FFT are symmetric to the first ones and do not provide further information.

Out of the amplitude spectrum (magnitudes of the Fourier transformed time series, figure 3.1 (b) top) standard features can be determined such as

- the frequency at which the maximum magnitude occurs,

¹The discrete Fourier transform is usually computed by the fast Fourier transform (FFT) algorithm.

- the left and right cut-off frequency at which the signal drops by $-6\text{ dB} \approx -50\%$ compared to the maximum magnitude and
- the bandwidth, the difference between right and left cut-off frequency.

Using these features, only insufficient results with balanced accuracies $< 70\%$ can be achieved (table 4.1).

feature	balAcc [%]	TPR [%]	FPR [%]	AUC	corr
max. magnitude frequency	64.44	56.13	27.25	0.6494	-0.2479
left cut-off frequency	50.27	50.27	49.73	0.5475	0.2338
right cut-off frequency	66.08	66.09	33.92	0.7164	-0.2696
bandwidth	69.25	69.26	30.75	0.7447	-0.3783

Table 4.1: Different quality measures applied on classifications by FFT standard features. All values are averaged over the three scans of B50+56l. The intermediate echo gate is chosen from $1\ \mu\text{s}$ to $4\ \mu\text{s}$ and the reference threshold is -6 dB ; threshold for classification is chosen such that $\text{TPR} \approx \text{TNR}$.

Another approach is to use the sum of magnitudes of several neighbored frequencies as back-wall echo equivalent. We looked therefore for the window that led to best classification results. For $50 + 56\text{l}$ this is 2.67 MHz to 3.67 MHz . With this approach, better classification results can be obtained (table 4.1). These are however still clearly below the goal of reaching $\geq 90\%$ TPR and $\leq 10\%$ FPR.

feature	balAcc [%]	TPR [%]	FPR [%]	AUC	corr
summing of magnitudes from 2.67 MHz to 3.67 MHz	77.16	77.16	22.84	0.8583	-0.5942

Table 4.2: Different quality measures applied on classifications by FFT summing of frequency magnitudes; the summing over frequency magnitudes with best classification has been chosen. All values are averaged over the three scans of B50+56l. The intermediate echo gate is chosen from $1\ \mu\text{s}$ to $4\ \mu\text{s}$ and the reference threshold is -6 dB ; threshold for classification is chosen such that $\text{TPR} \approx \text{TNR}$.

We propose to use linear regression on the amplitude spectrum instead as a new approach in this report.

4.3 Time shifts tested on time-linear regression

Linear regression on time data was developed and evaluated in two former modeling projects within the master degree studies in Industrial Mathematics in Bremen and has produced promising results. However, as mentioned in section 3.1 time-linear regression is expected to be sensitive regarding time shifts, which were not included in those tests. Time shifts in the intermediate echoes are very likely to occur when leaving the lab environment, due to changes in appearance of the parts or changes in distance between probe and sample. In our evaluation we simulated time shifts by adapting the intermediate echo gate for the test time series but not for the training time series. The results are shown in figure 4.1, including those without shift at the bar at 0. To measure the quality we use the balanced accuracy as it summarizes the normalized confusion matrix with a single number, which is also used for choosing the classification threshold on the training data. In the A-scans a main oscillation with a frequency about 5.6 MHz is observed. We try shifts of up to half a period (90 ns) in both directions.

As expected the best results are obtained without shift, while shifts of about half a period result in balanced accuracies of around 50% , which is the expected rate for a random classification, or even below. This can be understood by the fact that coefficients are mapped to time points, at which the signal may have a significantly different magnitude or even the opposite sign, leading to mismatching weights in the linear combination. On the other hand the results for small shifts demonstrate that the time-linear regression is a reasonable method if the distance between the probe and the CFRP layers can be ensured to be constant, and the build-up of the parts to be inspected is very reproducible.

4.4 Performance test of basic FFT-linear regression including time shifts

This section shows a first evaluation of the new approach of linear regression based on FFT data. One could expect the influence of time shifts on this regression to be small, because translations of a signal

$x(t)$ in time only change the phase but not the magnitude of the corresponding frequency spectrum $[\mathcal{F}(x(t))](\omega)$,

$$[\mathcal{F}(x(t \pm t_0))](\omega) = [\mathcal{F}(x(t))](\omega) e^{\pm i\omega t_0},$$

where $|e^{\pm i\omega t_0}| = 1$ for all ω .

Compared to the results for time-linear regression (figure 4.1) the accuracy is higher on average (figure 4.2). The decrease of accuracy after applying time shifts is due to the leakage effect that has a different impact on the amplitude spectrum for each time shift. For small negative time shifts the classification is still accurate in the considered example. This may be due to some shift-specific error of $F(E)$ that is almost uncorrelated with the weights x for these particular shifts. The high accuracy for small negative time shifts that is reached with the specific intermediate echo gate can presumably not be generalized.

Nevertheless the FFT-linear regression is considered as a promising approach.

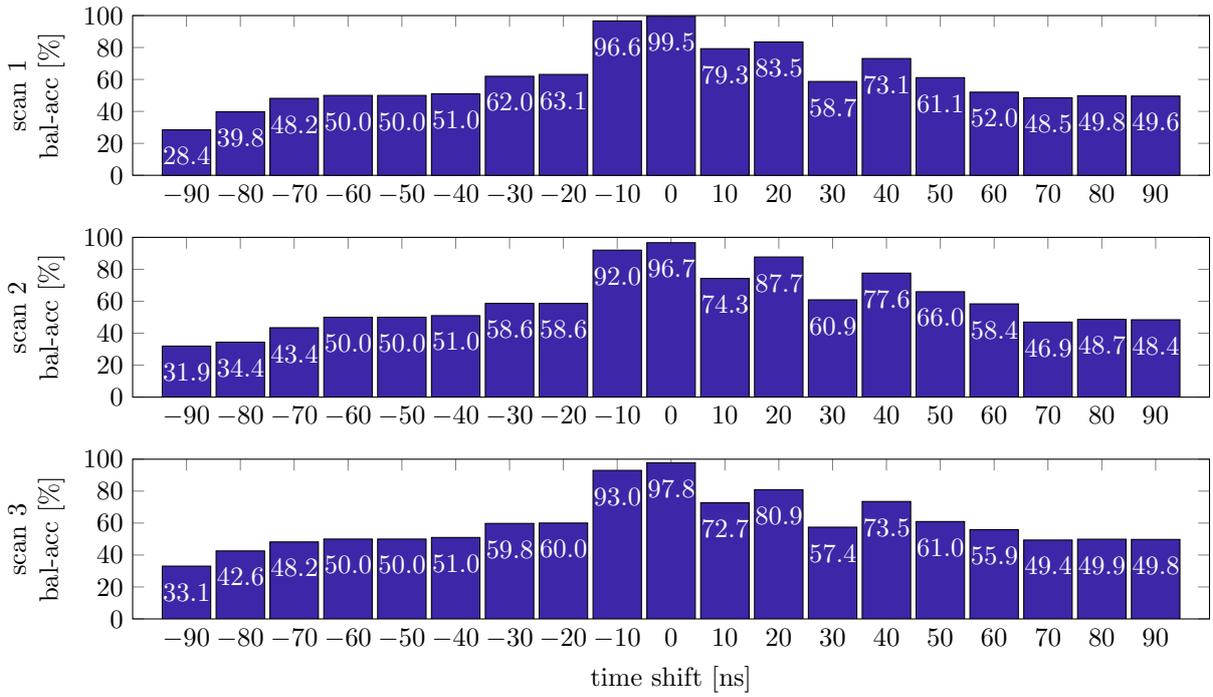


Figure 4.1: Effect of the time shift on the classification result for time-linear regression. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -6 dB.

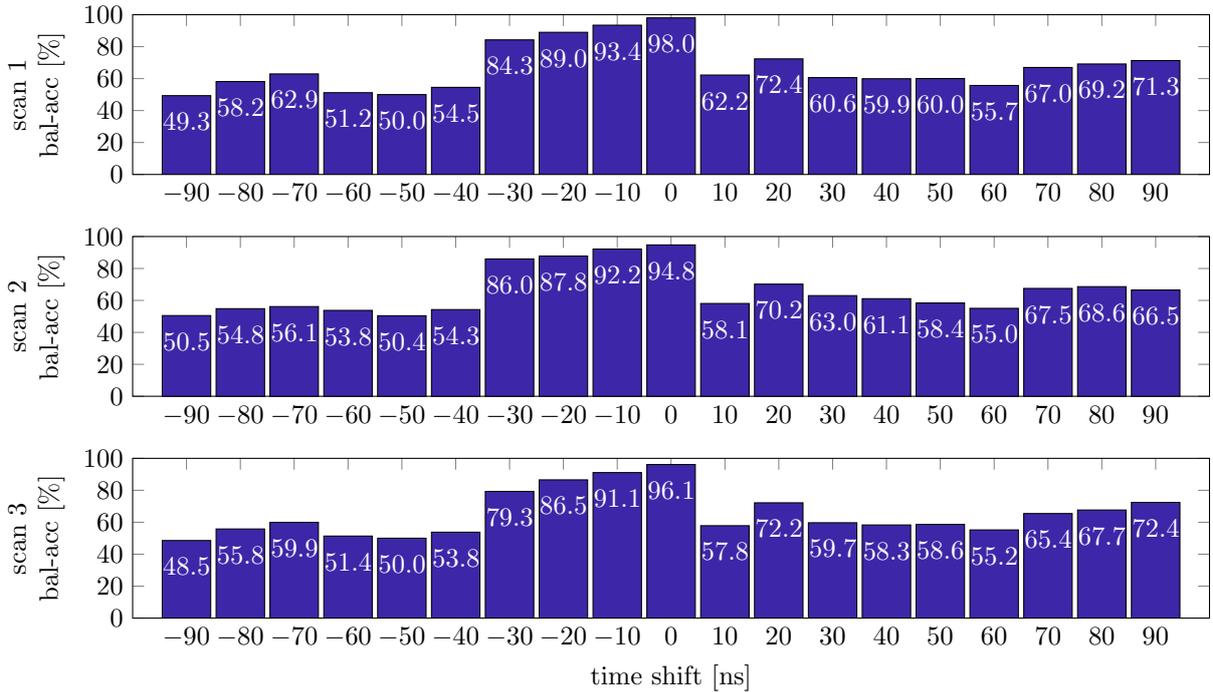


Figure 4.2: Effect of the time shift on the classification result for FFT-linear regression. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -6 dB.

4.5 Comparison of FFT-linear regression methods

Now we focus on the different FFT-linear regression methods obtained by weighting in time domain. We apply different quality measures introduced in chapter 2 on classifications by the FFT-linear regression methods introduced in chapter 3. Both TCG and windowing can be activated optionally, resulting in four methods to compare.

TCG	window	balAcc [%]	TPR [%]	FPR [%]	TNR [%]	FNR [%]	AUC	corr
–	–	96.31	98.31	5.70	94.30	1.69	0.9953	0.8807
✓	–	95.57	98.80	7.67	92.33	1.20	0.9944	0.8756
–	✓	99.03	99.08	1.02	98.98	0.92	0.9995	0.9163
✓	✓	99.03	99.11	1.05	98.95	0.89	0.9996	0.9165

Table 4.3: Different quality measures applied on classifications by FFT-linear regression methods. Here no shifts of the intermediate echoes were simulated. All values are averaged over the three scans of B50+56r after training on all scans of B50+56l. The intermediate echo gate is chosen from 1 μ s to 4 μ s and the reference threshold is -6 dB.

All quality measures verify that windowing improves the results of the basic method. Without windowing, TCG has a negative influence on the performance, but it marginally improves the correlation coefficient when combined with windowing. The overall results in table 4.3 are very satisfying. For the methods applying windowing one can summarize $FPR, FNR < 2\%$, $TPR, TNR > 98\%$ and $AUC > 99.9\%$.

Another way to display the classification results is to directly plot the BWE equivalents against the corresponding true BWE (figure 4.3).

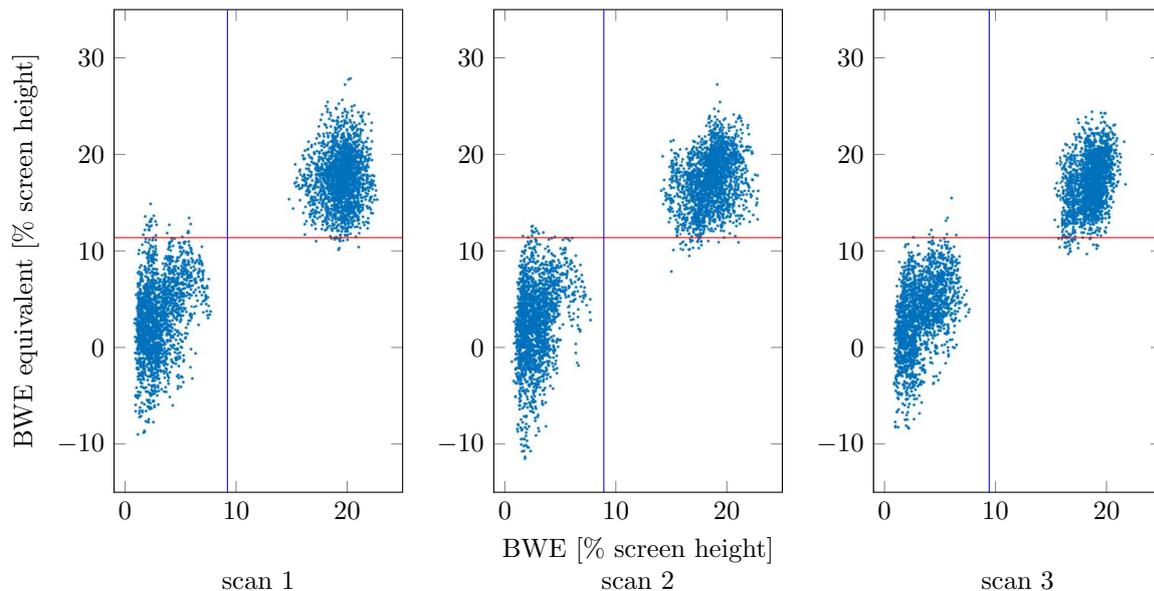


Figure 4.3: Plots of BWE equivalents computed by the FFT-linear regression with windowing and TCG over the corresponding true BWE for the three scans of B50+56r after training on all scans of B50+56l. The reference thresholds are marked by blue lines and the classification thresholds are marked by red lines.

For maximal correlation (cf. chapter 2.3) all points would lie on a straight line. While this is not the case the clouds are still almost separable by the BWE equivalent. Both the plots and the high correlation coefficient of over 0.85 suggest a strong relationship between the BWE and its equivalent, based on the FFT of the intermediate echoes. However it must be expected that the classification would be much less accurate, if also BWEs of medium height would occur.

Comparing the C-scans of the original BWE with the ones of the BWE equivalent computed by the FFT-linear regression with windowing and TCG shows that the errors made by the regression occur at similar locations in all three scans, which means they are reproducible (figure 4.4). Notice that no spatial information is involved when calculating the BWE equivalents. Therefore the clear visual separation between porous and non-porous points has no influence on the prediction results.

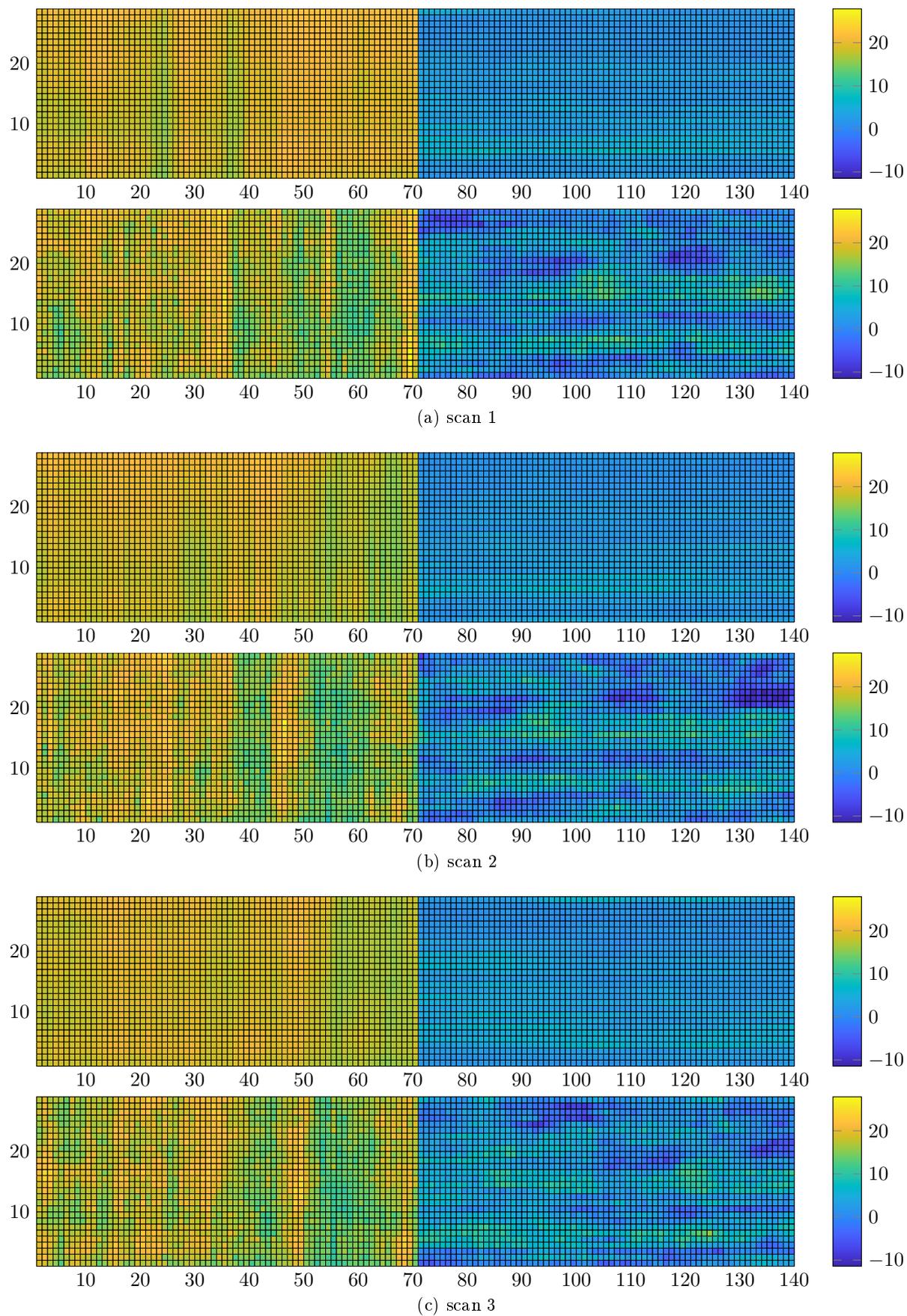


Figure 4.4: C-scans of the original BWE (top) and the BWE equivalent computed by the FFT-linear regression with windowing and TCG (bottom) for the three scans of B50+56r.

4.6 Comparison of Wavelet-based Linear Regression Models

Next we try the Wavelet-based models described in section 3.3, derived from [7].

method	balAcc [%]	TPR [%]	FPR [%]	TNR [%]	FNR [%]	AUC	corr
dwt_db4_lvl5	87.14	81.43	7.14	92.86	18.57	0.9235	0.6615
dwt_db4	93.83	93.94	6.29	93.71	6.06	0.9827	0.6015
dwt_db8	96.05	95.99	3.89	96.11	4.01	0.9907	0.8496
dwt_db10	93.44	91.30	4.42	95.58	8.70	0.9840	0.8048

Table 4.4: Different quality measures applied on classifications by Wavelet-based linear regression methods. Here no shifts of the intermediate echoes were simulated. All values are averaged over the three scans of B50+56r after training on all scans of B50+56l. The intermediate echo gate is chosen from 1 μ s to 4 μ s and the reference threshold is -6 dB.

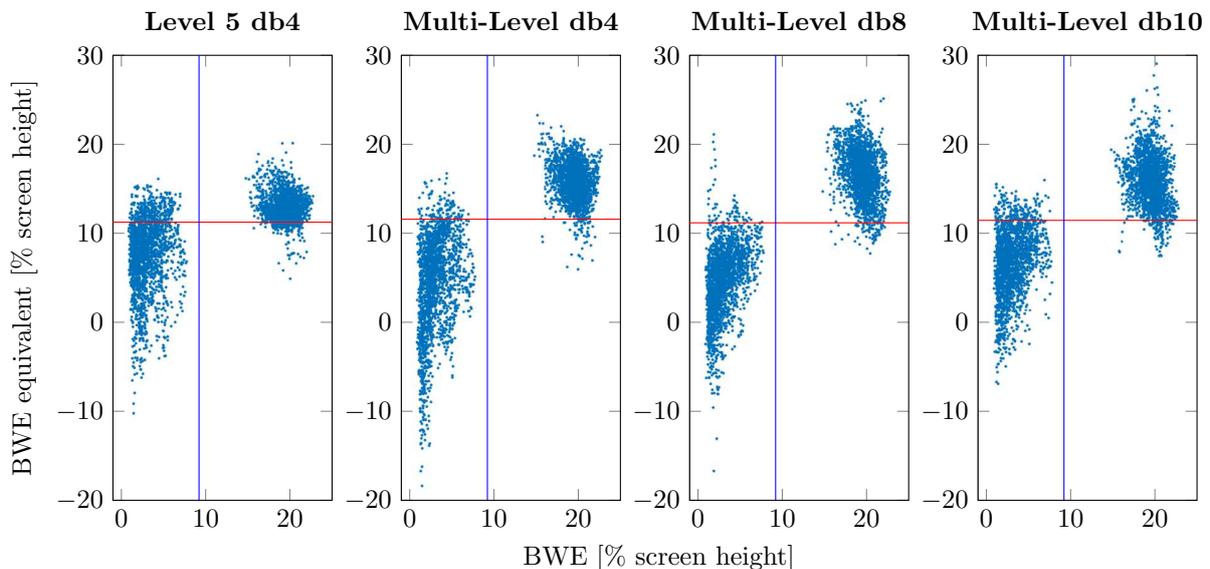


Figure 4.5: Plots of BWE equivalents computed by the four different DWT-based linear regressions over the corresponding true BWE for the first scan of B50+56r after training on all scans of B50+56l. The reference threshold is marked by a blue line and the classification thresholds are marked by red lines.

Notice that the method only using level 5 performs remarkably worse than the other three methods, which include all levels from 1 to 7. A possible explanation would be that for our data and intermediate echo gate the level 5 determined in [7] is not suitable. Nevertheless it could be beneficial to include many levels as input to the regression model, unless one has to expect overfitting. As the train and the test set are disjoint and the classification is quite accurate this does not seem to occur on the data B50+56. For all multi-level-methods we observe $FPR, FNR < 10\%$, $TPR, TNR > 90\%$ and $AUC > 98\%$.

For the first two methods the correlation coefficient is much smaller. Because this does not match with the balanced accuracies, we have a look at the BWE/BWE-equivalents plots (figure 4.5). The plots show that the loss of correlation is not caused by the BWE equivalents of the non-porous points but rather of the porous points. Mainly the BWE equivalents of some porous points reach large negative values, which lie far away from a linear fit through all points. However this is not important for the classification, since negative BWE equivalents always should be classified as porous. Because of this observations and the absence of medium-porous points, the classification performance does not depend strongly on the correlation.

Since some results of the DWT-based linear regressions get relatively close to the results of the previous section, we will proceed our evaluation by applying time shifts to FFT-linear regression models with weighting in time domain and to Wavelet-based linear regression models.

4.7 Time shifts for FFT-linear and Wavelet-based linear regression models

Here we restrict ourselves to the comparison of the best methods delivering the best results of each model group, which are the FFT-linear regression with Hann windowing and TCG, and Multi-Level-DWT with *db8*.

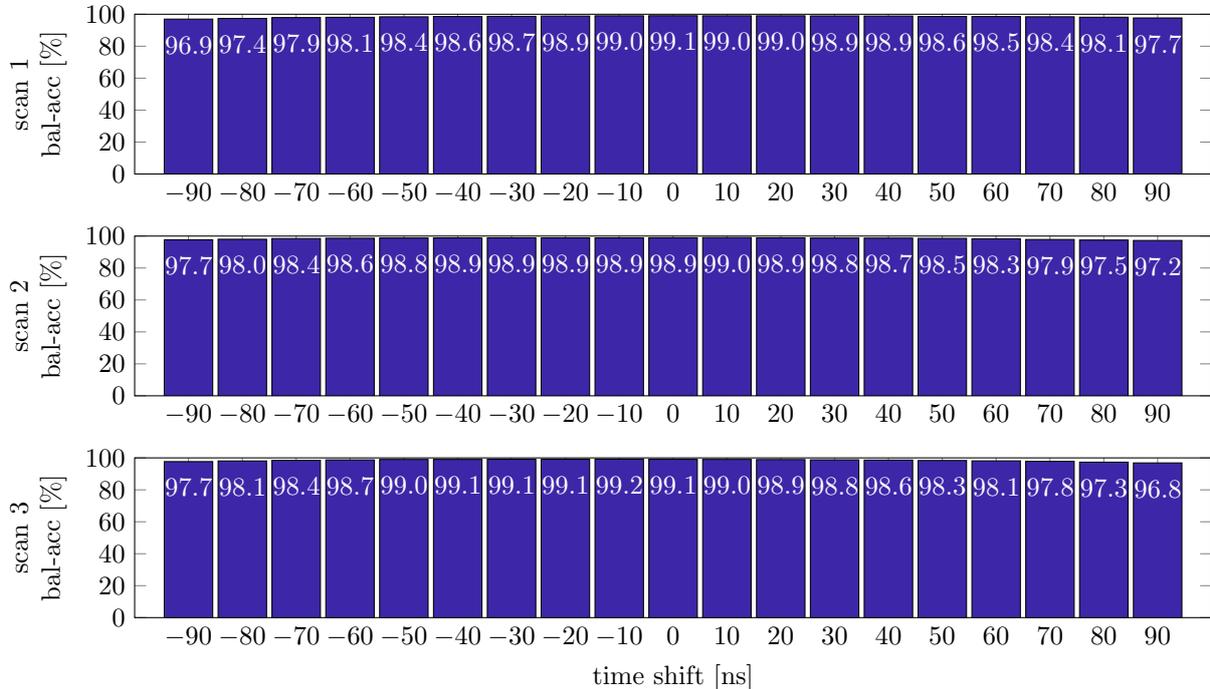


Figure 4.6: Effect of the time shift on the classification result for FFT-linear regression with windowing and TCG. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -6 dB.

Windowing leads to great robustness of the FFT-linear regression regarding time shifts (cf. figure 4.6). The averaged balanced accuracy over all time shifts reaches 98.4% and for small time shifts almost 99%. While we did assume that windowing would improve the performance for large time shifts, these results exceeded our expectations. Note that for the reference threshold -12 dB the method is robust as well (cf. figure A.3), even though the performance is worse like observed for all methods.

The DWT-based linear regression on the contrary cannot handle time shifts, which is, however, no surprise considering its explicit time dependency. It provides decent results for positive time shifts, especially for large shifts, for which about 90% is reached (figure 4.7). But already for small negative time shifts the classification is not much better than a random classification. Since we require robustness regarding time shifts, the FFT-linear regression with windowing and TCG should be preferred over the DWT-based approach.

4.8 Validation on different data sets of a specimen

To validate the good results of the FFT-linear regression with windowing and TCG, now additional data sets of the specimen B50+56, of which the left and right half (B50+56l and B50+56r) were used previously, are taken into account.

We consider four data sets in total. The first two are B50+56l and B50+56r, which we now call “left” and “right”. The third one (“repeat”) is merged from repeated scans of the whole B50 and B56 specimens, where the probe was moved with approximately twice the speed, leading to scans containing 29×70 A-scans, combined 29×140 A-scans. The last data set (“flipside”) contains the same scans as “repeat” of the B50 specimen, but scans of the B56 specimen for which it was turned around such that its front- and back-wall swapped positions.

All balanced accuracies in table 4.5 are remarkably over 90%. The entries in boxes on the diagonal are trained and evaluated on the same data set, and thus clearly not representative for a real application

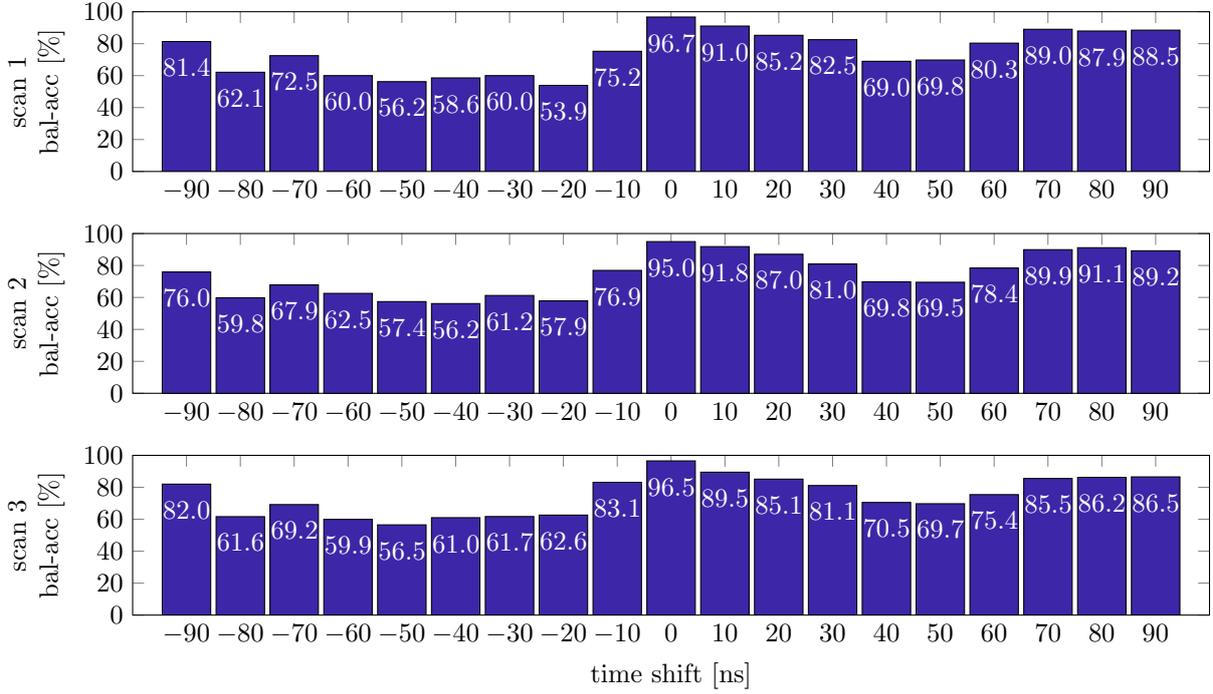


Figure 4.7: Effect of the time shift on the classification result for multi-level-DWT with *db8*. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -6 dB.

balAcc on:	left			right			repeat			flipside			
	scan	1	2	3	1	2	3	1	2	3	1	2	3
left		99.00	99.80	99.88	99.06	98.94	99.09	97.49	98.67	98.89	97.44	98.35	99.09
right		96.08	98.74	98.65	99.85	99.85	99.68	95.49	97.14	96.82	94.38	94.24	96.43
repeat		94.79	97.17	96.03	98.13	98.18	98.20	99.51	99.78	99.70	99.63	99.38	99.98
flipside		92.77	95.05	94.51	93.77	95.54	95.10	99.58	98.87	98.55	99.73	99.73	99.98

Table 4.5: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

scenario. The only slightly worse results are observed when learning from “left” or “right” and evaluating on “repeat” or “flipside”, or vice versa, i.e. in the upper right or lower left 2×2 -submatrices.

The goal formulated in chapter 2 is to reach $\text{FPR} > 90\%$ and $\text{TPR} < 10\%$. Table 4.6 shows that the criterion for the FPR is fulfilled safely and the criterion for the TPR also is met with two minor exceptions (88.76% and 89.56%, when training on “left” and evaluating on “flipside”).

TPR on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	99.60	99.75	99.75	98.62	99.16	99.56	99.95	99.61	99.61	99.85	98.97	100.00
right	98.66	99.36	99.11	99.70	99.90	99.51	99.61	99.90	99.90	97.39	94.09	99.11
repeat	93.18	94.92	92.32	97.64	98.77	97.49	99.70	99.66	99.46	99.95	98.87	100.00
flipside	88.76	91.67	89.56	91.97	95.71	92.76	99.70	97.83	97.14	100.00	99.56	100.00

FPR on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	1.61	0.15	0.00	0.49	1.28	1.38	4.98	2.27	1.82	4.98	2.27	1.82
right	6.49	1.87	1.82	0.00	0.20	0.15	8.62	5.62	6.26	8.62	5.62	6.26
repeat	3.61	0.59	0.25	1.38	2.41	1.08	0.69	0.10	0.05	0.69	0.10	0.05
flipside	3.22	1.57	0.54	4.43	4.63	2.56	0.54	0.10	0.05	0.54	0.10	0.05

Table 4.6: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are TPR (top) or FPR (bottom) in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

The small performance losses may be caused by an unfortunate choice of the classification threshold, which is determined by maximization of the balanced accuracy on the training data. Therefore we look at the AUC, which is independent of the classification threshold. The results in table 4.7 suggest that the decreased accuracy is due to a more inaccurate regression, since the AUC also decreases for those cases.

Here again we tested to shift the intermediate echo gate for the evaluation data while keeping it fixed for the training data. Only very small differences ($< 2.5\%$) are observed between the balanced accuracies of the validation without time shift (table 4.5) and with a time shift of approximately a quarter wave length, ± 50 ns (table 4.8 and 4.9).

AUC on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	0.9985	0.9999	1.0000	0.9994	0.9997	0.9996	0.9994	0.9995	0.9994	0.9992	0.9991	1.0000
right	0.9941	0.9995	0.9994	1.0000	1.0000	1.0000	0.9965	0.9996	0.9990	0.9898	0.9901	0.9973
repeat	0.9913	0.9969	0.9970	0.9985	0.9990	0.9982	0.9999	0.9999	1.0000	1.0000	0.9998	1.0000
flipside	0.9880	0.9931	0.9943	0.9847	0.9918	0.9912	0.9999	0.9989	0.9993	1.0000	1.0000	1.0000

Table 4.7: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	98.44	99.26	99.46	98.37	98.79	99.01	98.84	99.46	99.41	98.72	98.00	99.70
right	96.29	98.94	98.74	99.68	99.66	99.56	97.19	98.89	98.37	95.30	93.62	97.93
repeat	93.26	95.86	94.46	97.46	97.88	97.27	99.61	99.73	99.21	99.75	99.04	100.00
flipside	91.35	93.39	92.96	93.33	95.47	94.33	99.63	98.37	97.83	99.75	99.68	99.98

Table 4.8: Validation with a time shift of -50 ns (after training without shift) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

balAcc on:	left			right			repeat			flipside		
scan	1	2	3	1	2	3	1	2	3	1	2	3
left	98.46	99.56	99.48	98.60	98.55	98.35	95.64	97.02	97.12	95.64	97.00	97.27
right	95.11	98.03	97.34	99.85	99.68	99.33	93.55	95.37	94.48	93.03	93.57	94.36
repeat	95.19	97.66	97.00	97.91	97.91	98.28	99.26	99.56	99.70	99.38	99.31	99.88
flipside	93.84	95.88	94.75	93.99	95.57	95.42	99.56	98.74	98.87	99.68	99.61	99.98

Table 4.9: Validation with a time shift of +50 ns (after training without shift) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

4.9 Restriction to parts of the intermediate echo gate

So far we used the full intermediate echo gate from close after the front-wall echo (1 μs) up to close before the BWE (4 μs). Now we test different gates.

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	95.65	96.03	96.95	96.18	97.00	95.34	96.06	92.73	94.33	96.38	95.22	96.95
right	93.87	95.00	95.27	97.22	97.88	97.64	93.99	90.37	90.34	94.14	91.87	92.36
repeat	92.59	93.35	92.73	94.43	94.85	93.99	97.56	94.38	95.39	96.31	94.14	97.49
flipside	92.60	92.61	92.36	92.66	93.84	93.37	96.23	90.59	92.81	97.12	95.84	98.97

Table 4.10: Validation using only the middle part of the intermediate echoes (1.5 μs to 3.5 μs) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	94.89	96.38	98.87	93.00	93.77	92.61	90.20	93.40	91.55	86.72	89.21	94.41
right	88.26	91.68	95.25	98.37	97.29	96.28	87.96	93.79	91.85	85.27	88.84	94.06
repeat	90.95	93.20	95.47	93.99	92.61	92.81	96.16	97.09	96.06	94.04	94.70	99.11
flipside	87.47	91.25	92.09	89.80	89.38	88.94	97.71	93.60	91.58	97.83	98.30	99.90

Table 4.11: Validation using only the first half of the intermediate echoes (1 μs to 2.5 μs) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	98.21	98.89	98.99	98.25	98.18	98.45	96.23	97.39	96.97	96.60	95.20	97.17
right	96.34	97.29	98.28	99.04	98.89	98.77	95.76	97.73	98.18	94.14	94.14	98.08
repeat	94.68	96.58	96.92	96.80	97.00	97.00	98.84	98.40	98.72	98.20	96.85	99.66
flipside	92.82	94.58	94.06	93.84	94.33	94.29	98.62	96.43	97.09	98.99	97.96	99.75

Table 4.12: Validation using only the second half of the intermediate echoes (2.5 μs to 4 μs) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

Simply shrinking the gate from both sides slightly decreases the performance as expected (cf. table 4.10). When taking only the first half, the accuracy drops significantly (cf. table 4.11). This can be explained by the fact that no information about the part after 2.5 μs is contained anymore. On the other hand the second half is influenced by the first half of the intermediate echoes, since energy losses in the first half (wave reflections at pores) lead to different results for the second half, even if the according part of the sample was completely free of pores. This might be the reason why the results on this half are much better (cf. table 4.12).

4.10 Regression coefficients

Now we are interested in the information learned by the FFT-linear regression with windowing and TCG, which showed the best performance of the methods tried in this report. This information is the vector of weighting coefficients for the frequencies. It is displayed for the full B50 and B56 data set compared to the “repeat” data set (figure 4.8) and compared to “flipside” (figure 4.9).

At first sight one notices that the weights in the frequency range from 7 MHz to 16.5 MHz are almost completely positive. From this these frequencies seem to appear more strongly in non-porous material compared to porous material.

To consider the actual impact of a frequency on the BWE equivalent, not only its weighting coefficient but also its variance across all A-scans has to be taken into account. Therefore we view the product of these two values, which we call the *impact value*,

$$x_j \cdot \sigma^2 = x_j \cdot \frac{1}{n} \sum_{i=1}^n \left(e_{i,j} - \frac{1}{n} \sum_{i=1}^n e_{i,j} \right)^2, \quad j = 1, \dots, m \quad (4.2)$$

(cf. figure 4.10 and figure 4.11). This shows that all frequencies over 7 MHz are almost irrelevant for the BWE equivalent. These include the frequencies up to 16.5 MHz with positive weights and the higher frequencies, whose weighting does not show a concrete structure. Therefore they appear to only adjust to some minor effects rather than meaningful aspects of the data. Instead the frequency range from 1 MHz to 5 MHz seems to have a negative influence on the BWE.

The largest positive impact value for both training on the full B50 and B56 data set and on the “repeat” data set is at 5.33 MHz (figure 4.10). On the contrary for the “flipside” data set this frequency has a slightly negative impact value (figure 4.11). However the neighboring frequency 5.67 MHz has the largest positive impact here. We assume that these two frequencies are strongly coupled, for example due to some blurriness caused by the windowing, and thus are nearly interchangeable. This might also be the case for other (not necessarily neighboring) frequencies, which are coupled by physical effects not further investigated.

Nevertheless the major characteristics of the impact values are similar for all training data sets viewed. Therefore we conclude that a relation between the frequency spectrum and the porosity of CFRP seems plausible and consider the FFT-based approach to be a decent choice.

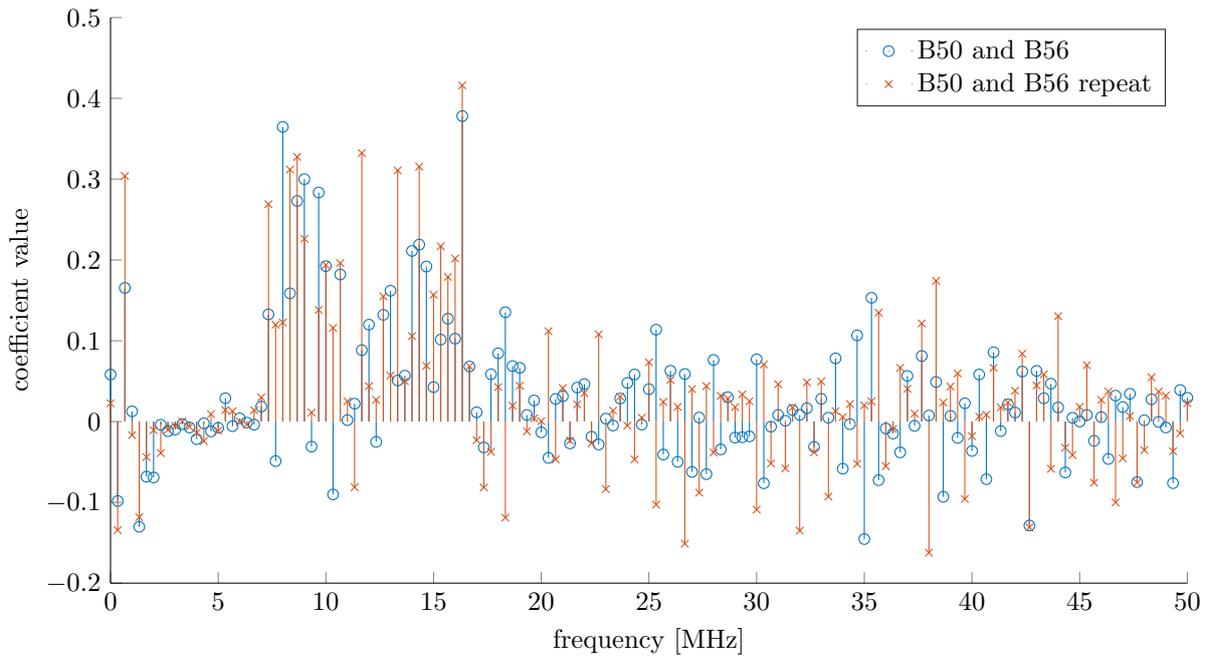


Figure 4.8: Coefficient values of the FFT-linear regression with windowing and TCG. The training was done on the full B50 and B56 data set (circles) and on the “repeat” data set (crosses).

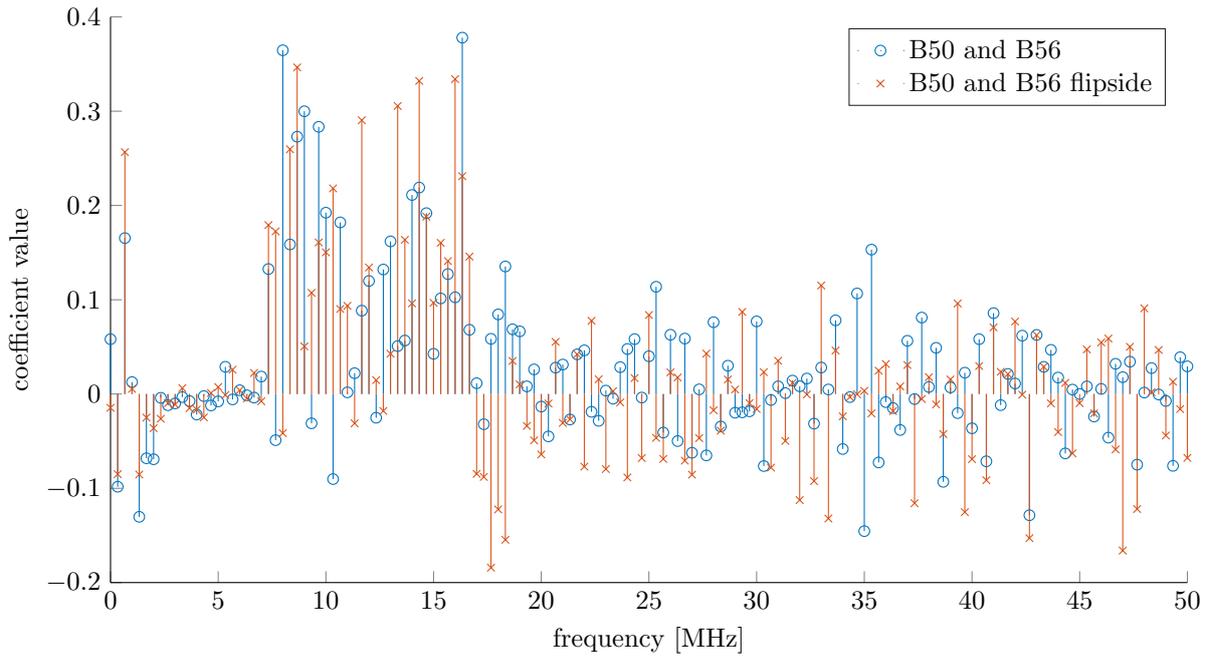


Figure 4.9: Coefficient values of the FFT-linear regression with windowing and TCG. The training was done on the full B50 and B56 data set (circles) and on the “flipside” data set (crosses).

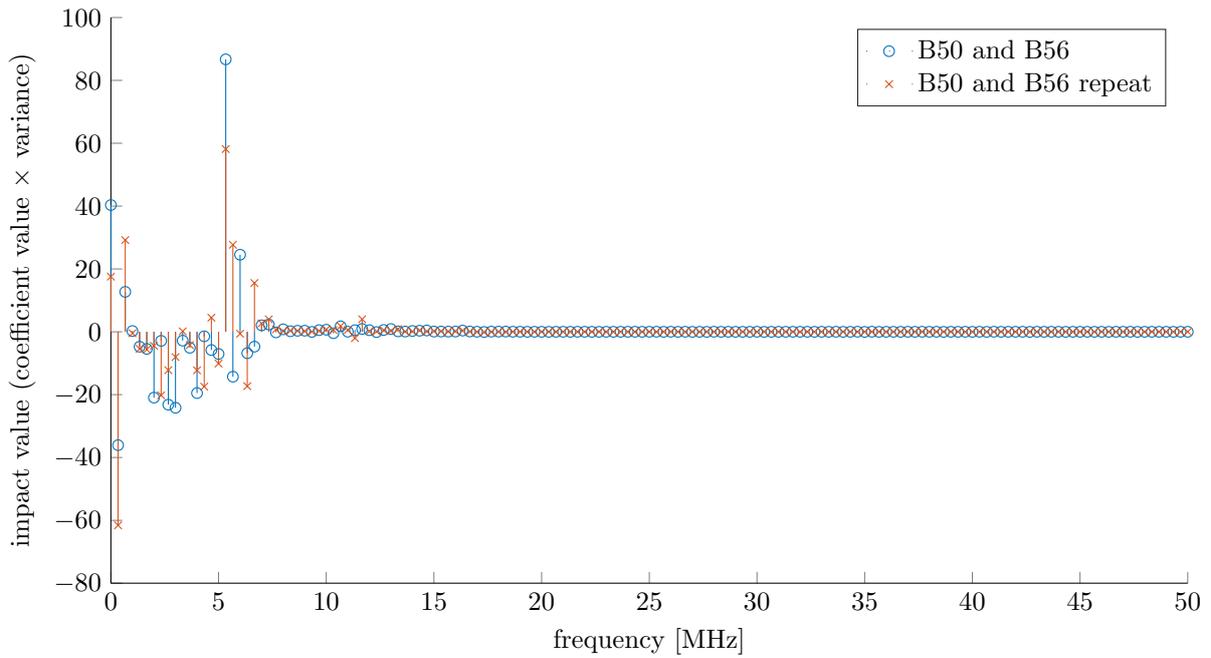


Figure 4.10: Coefficient values of the FFT-linear regression with windowing and TCG, multiplied with the variance of the magnitude of the respective frequency in the training data set (impact value). The training was done on the full B50 and B56 data set (circles) and on the “repeat” data set (crosses).

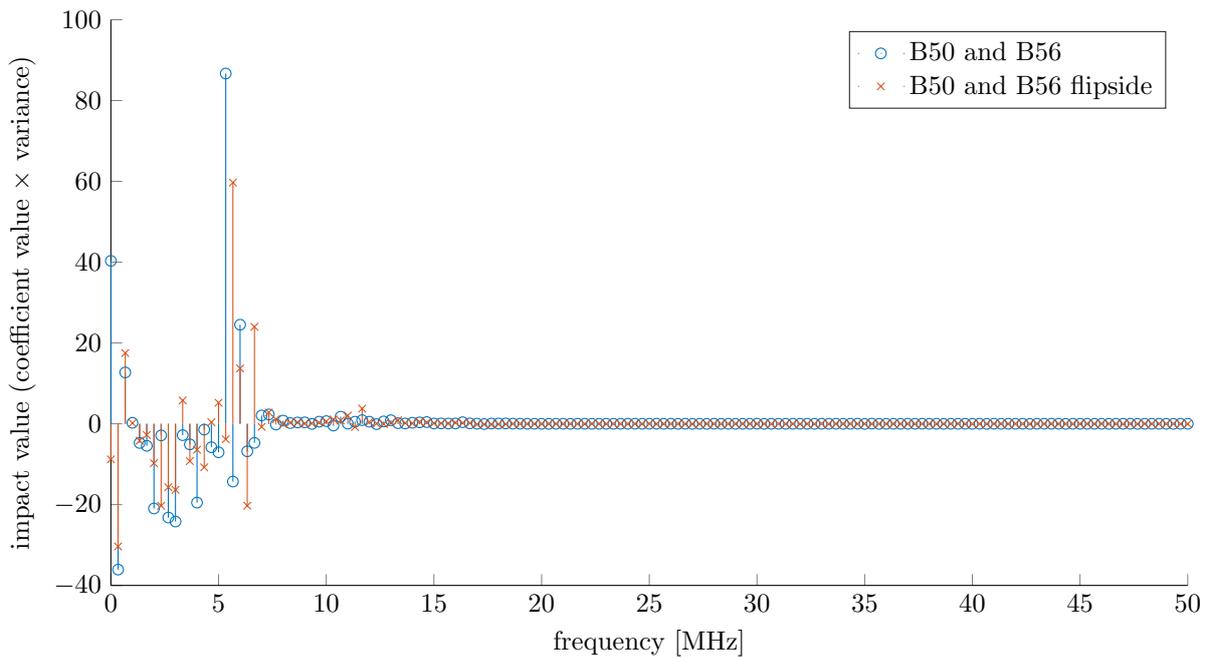


Figure 4.11: Coefficient values of the FFT-linear regression with windowing and TCG, multiplied with the variance of the magnitude of the respective frequency in the training data set (impact value). The training was done on the full B50 and B56 data set (circles) and on the “flipside” data set (crosses).

4.11 Regularization of the regression

We now consider the ℓ_2 -regularized version

$$x = \arg \min_x \|F(E)x - p\|_2^2 + \lambda \|x\|_2^2.$$

of the minimization problem (3.1).

This regularized problem is solved analytically by

$$x = (F(E)^T F(E) + \lambda I)^{-1} F(E)^T p,$$

which is derived like the normal equation (3.2) by differentiating the objective and finding its zero.

By adding the convex ℓ_2 -term, the solution x becomes more stable, which prevents overfitting. In our case we do not expect great benefits, since the tests across the different data sets of the B50 and B56 specimen already showed very good results. Table 4.13 shows the results for the regularized regression.

balAcc on: scan	left			right			repeat			flipside		
	1	2	3	1	2	3	1	2	3	1	2	3
left	98.46	99.66	99.88	99.19	99.43	99.80	97.86	98.42	98.72	97.81	97.56	98.99
right	95.99	99.09	99.56	99.78	99.75	99.51	95.81	97.39	96.28	95.05	95.22	95.91
repeat	94.85	97.68	98.18	97.54	96.90	97.71	99.58	99.56	99.56	99.63	99.31	99.95
flipside	91.77	94.18	94.14	92.17	93.10	93.33	99.33	97.27	97.64	99.78	99.43	99.93

Table 4.13: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -6 dB.

More interesting are the weights obtained with regularization, as shown in figure 4.12 and 4.13. Clearly the weights for the frequencies over 20 MHz are regularized to nearly 0, in contrast to figure 4.8 and 4.9. The weights for frequencies up to 20 MHz are however qualitatively similar with or without regularization.

Figures 4.14 and 4.15 show the corresponding impact values.

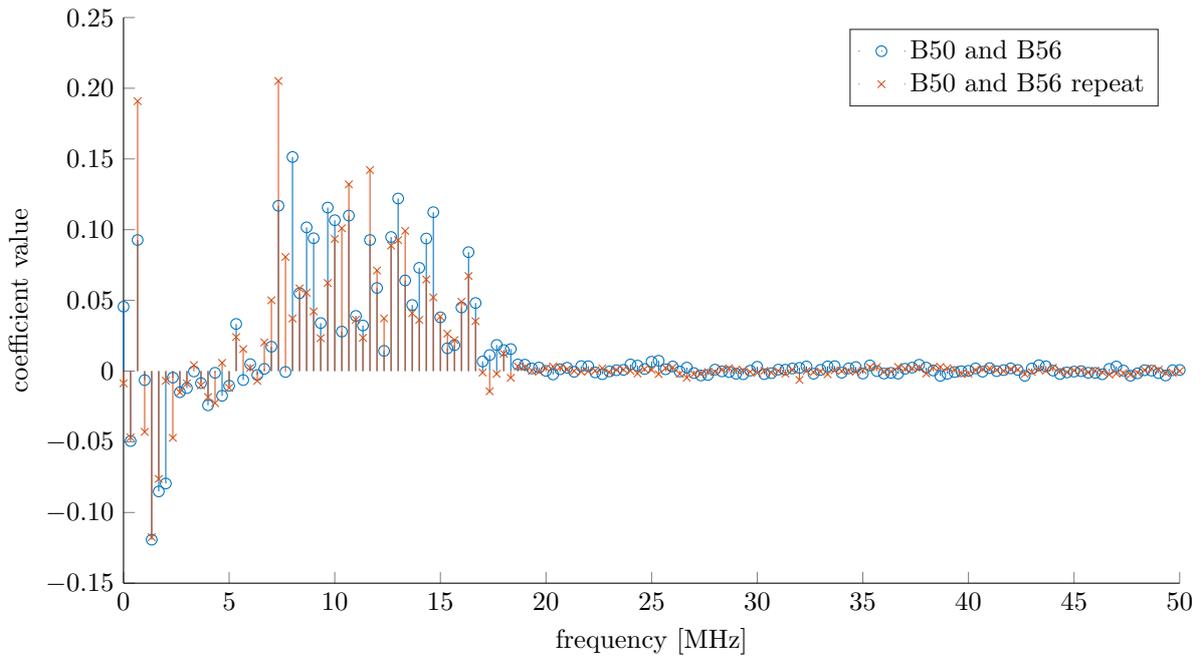


Figure 4.12: Coefficient values of the ℓ_2 -regularized FFT-linear regression with windowing and TCG. The regularization parameter is $\lambda = 10^5$. The training was done on the full B50 and B56 data set (circles) and on the “repeat” data set (crosses).

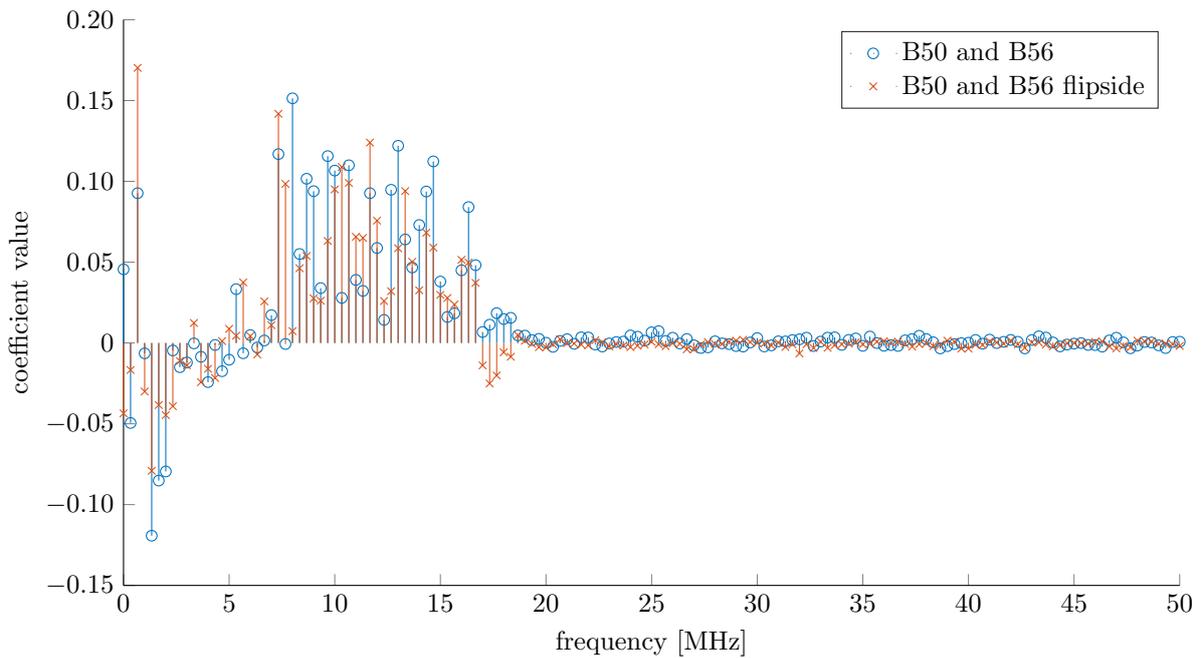


Figure 4.13: Coefficient values of the ℓ_2 -regularized FFT-linear regression with windowing and TCG. The regularization parameter is $\lambda = 10^5$. The training was done on the full B50 and B56 data set (circles) and on the “flipside” data set (crosses).

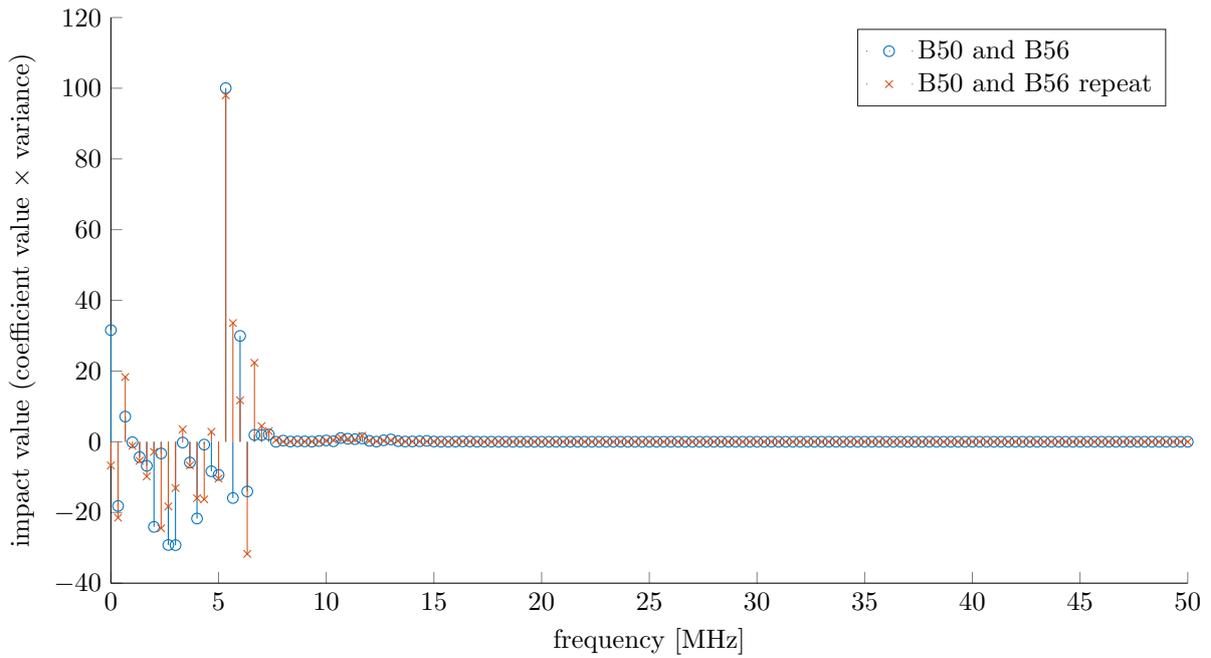


Figure 4.14: Coefficient values of the ℓ_2 -regularized FFT-linear regression with windowing and TCG, multiplied with the variance of the magnitude of the respective frequency in the training data set (impact value). The regularization parameter is $\lambda = 10^5$. The training was done on the full B50 and B56 data set (circles) and on the “repeat” data set (crosses).

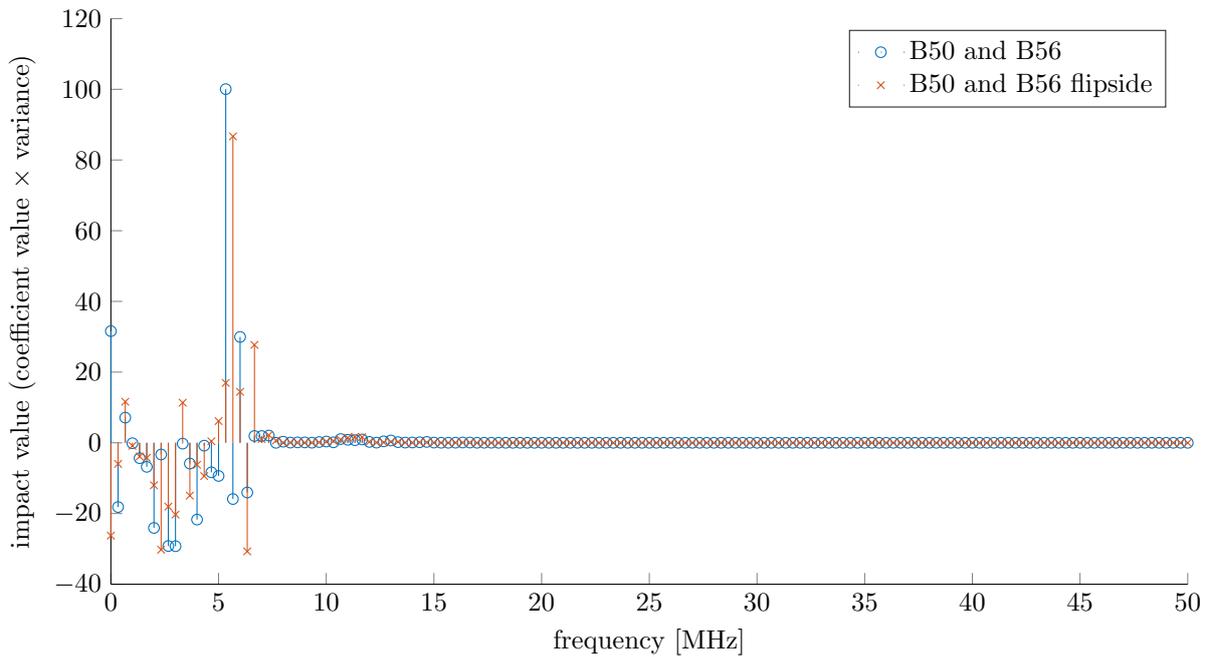


Figure 4.15: Coefficient values of the ℓ_2 -regularized FFT-linear regression with windowing and TCG, multiplied with the variance of the magnitude of the respective frequency in the training data set (impact value). The regularization parameter is $\lambda = 10^5$. The training was done on the full B50 and B56 data set (circles) and on the “flipside” data set (crosses).

Chapter 5

Conclusion

Methods for ultrasonic testing of carbon fiber reinforced polymers for porosity without a back-wall echo have been investigated. These methods attempt to provide a replacement for the industrially used method of back-wall echo reduction, which cannot be applied in certain situations. In particular linear regression of time series data, its Fourier transform and its wavelet transform have been investigated.

Very good results were obtained by the FFT-linear regression with windowing and time corrected gain. This method exceeds the goal formulated in chapter 2 of $\text{TPR} > 90\%$ and $\text{FPR} < 10\%$ in most cases. The performance does not decrease significantly in our time shift robustness test, where time shifts are applied to the time series used for evaluation but not on the time series used for training.

Time-linear regression and wavelet-based linear regression also deliver good results, but do not reach the same quality, especially with time shifts, where they tend to loose accuracy. To make the time-linear regression more robust against time shifts, some kind of automatic phase alignment should be tested. Such an approach was suggested by Kai-Wah Chan and Matthias Rick in a former modeling project.

The FFT-based approach has, in addition to better performance, the advantage that it allows more easily to find a physical explanation of the regression coefficients found. A reasonable next step is the evaluation of the method on further data. It would be helpful to test specimens with a medium degree of porosity, which are hard to produce. Furthermore, other methods directly providing classification without creating continuous back-wall echo equivalents as an intermediate step, such as a logistic regression classifier, could be tested.

Bibliography

- [1] *Airbus A350: Composites on Trial Part I*. <http://scribol.com/technology/aviation/airbus-a350-composites-on-trial-part-i/>. Visited on 10.10.2017.
- [2] C. Brandt. “Recurrence Quantification Analysis as an Approach for Ultrasonic Testing of Porous Carbon Fibre Reinforced Polymers”. In: *Recurrence Plots and Their Quantifications: Expanding Horizons*. Ed. by C. L. Webber Jr., C. Ioana, and N. Marwan. Cham: Springer International Publishing, 2016, pp. 355–377. ISBN: 978-3-319-29922-8. URL: https://link.springer.com/chapter/10.1007/978-3-319-29922-8_19.
- [3] C. Brandt. *Recurrence Quantification Analysis for Non-Destructive Evaluation with an Application in Aeronautic Industry*. 19th World Conference on Non-Destructive Testing (WCNDT 2016), 13–17 June 2016 in Munich, Germany. URL: <http://www.ndt.net/article/wcndt2016/papers/we1i4.pdf>.
- [4] I. N. Bronshtein et al. *Handbook of Mathematics*. Springer Berlin Heidelberg, 2004. DOI: 10.1007/978-3-662-05382-9. URL: <https://doi.org/10.1007/978-3-662-05382-9>.
- [5] A. Hering. *Carbon fabric image*. https://upload.wikimedia.org/wikipedia/commons/2/24/Carbon_fabric_yc_3k_200_twill.jpg. Accessed on 03.04.2019.
- [6] T. R. Hoens and N. V. Chawla. “Imbalanced Datasets: From Sampling to Classifiers”. In: *Imbalanced Learning*. John Wiley & Sons, Inc., 2013, pp. 43–59. ISBN: 9781118646106. DOI: 10.1002/9781118646106.ch3. URL: <http://dx.doi.org/10.1002/9781118646106.ch3>.
- [7] K.-B. Kim, D. K. Hsu, and D. J. Barnard. “Estimation of porosity content of composite materials by applying discrete wavelet transform to ultrasonic backscattered signal”. In: *NDT & E International* 56 (2013), pp. 10–16. ISSN: 0963-8695. DOI: <https://doi.org/10.1016/j.ndteint.2013.01.014>. URL: <http://www.sciencedirect.com/science/article/pii/S0963869513000212>.
- [8] J. Krautkrämer and H. Krautkrämer. *Ultrasonic Testing of Materials*. 4th edition. Berlin: Springer, 1990. URL: <http://link.springer.com/book/10.1007/978-3-662-10680-8#>.
- [9] P. K. Mallick. *Fiber-Reinforced Composites: Materials, Manufacturing, and Design, Third Edition*. Mechanical Engineering. CRC Press, 2007. ISBN: 9781420005981.
- [10] M. Neitzel, P. Mitschang, and U. Breuer. *Handbuch Verbundwerkstoffe: Werkstoffe, Verarbeitung, Anwendung*. Carl Hanser Verlag GmbH & Company KG, 2014. ISBN: 9783446436978. URL: <https://www.hanser-elibrary.com/isbn/9783446436961>.
- [11] D. Sundararajan. *The Discrete Fourier Transform*. WORLD SCIENTIFIC, 2001. DOI: 10.1142/4610. eprint: <https://www.worldscientific.com/doi/pdf/10.1142/4610>. URL: <https://www.worldscientific.com/doi/abs/10.1142/4610>.

Appendix A

Results for lower reference threshold

In the main part of the report only the reference threshold -6 dB was considered. This appendix contains results for the reference threshold -12 dB.

Method comparison

TCG	window	balAcc [%]	TPR [%]	FPR [%]	TNR [%]	FNR [%]	AUC	corr
–	–	90.36	94.86	14.15	85.85	5.14	0.9455	0.8807
✓	–	90.41	94.93	14.10	85.90	5.07	0.9423	0.8756
–	✓	90.96	91.91	9.99	90.01	8.09	0.9662	0.9163
✓	✓	91.08	92.02	9.87	90.13	7.98	0.9671	0.9165

Table A.1: Different quality measures applied on classifications by FFT-linear regression methods. Here no shifts of the intermediate echoes were simulated. All values are averaged over the three scans of B50+56r after training on all scans of B50+56l. The intermediate echo gate is chosen from $1 \mu\text{s}$ to $4 \mu\text{s}$ and the reference threshold is -12 dB. Cf. section 4.5

method	balAcc [%]	TPR [%]	FPR [%]	TNR [%]	FNR [%]	AUC	corr
dwt_db4_lvl5	84.48	80.04	11.08	88.92	19.96	0.8888	0.6615
dwt_db4	88.73	94.64	17.17	82.83	5.36	0.9381	0.6015
dwt_db8	90.27	92.39	11.85	88.15	7.61	0.9568	0.8496
dwt_db10	87.51	86.13	11.10	88.90	13.87	0.9446	0.8048

Table A.2: Different quality measures applied on classifications by Wavelet-based linear regression methods. Here no shifts of the intermediate echoes were simulated. All values are averaged over the three scans of B50+56r after training on all scans of B50+56l. The intermediate echo gate is chosen from $1 \mu\text{s}$ to $4 \mu\text{s}$ and the reference threshold is -12 dB. Cf. section 4.6

Time shifts

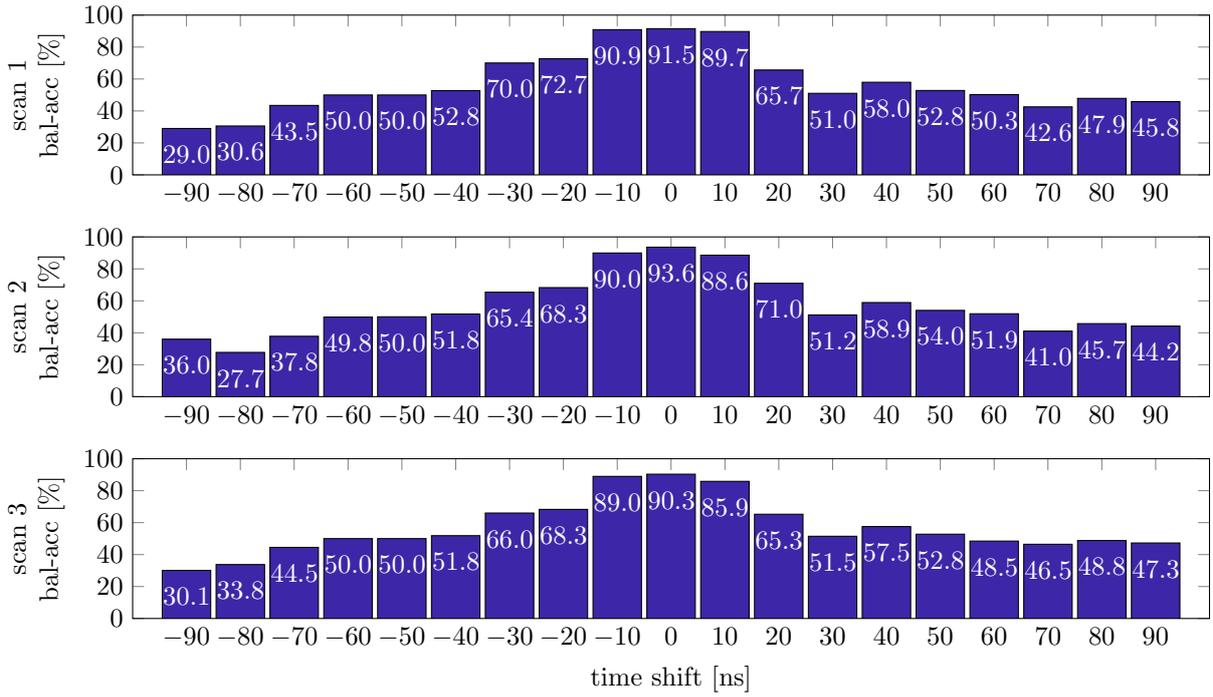


Figure A.1: Effect of the time shift on the classification result for time-linear regression. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -12 dB. Cf. section 4.3

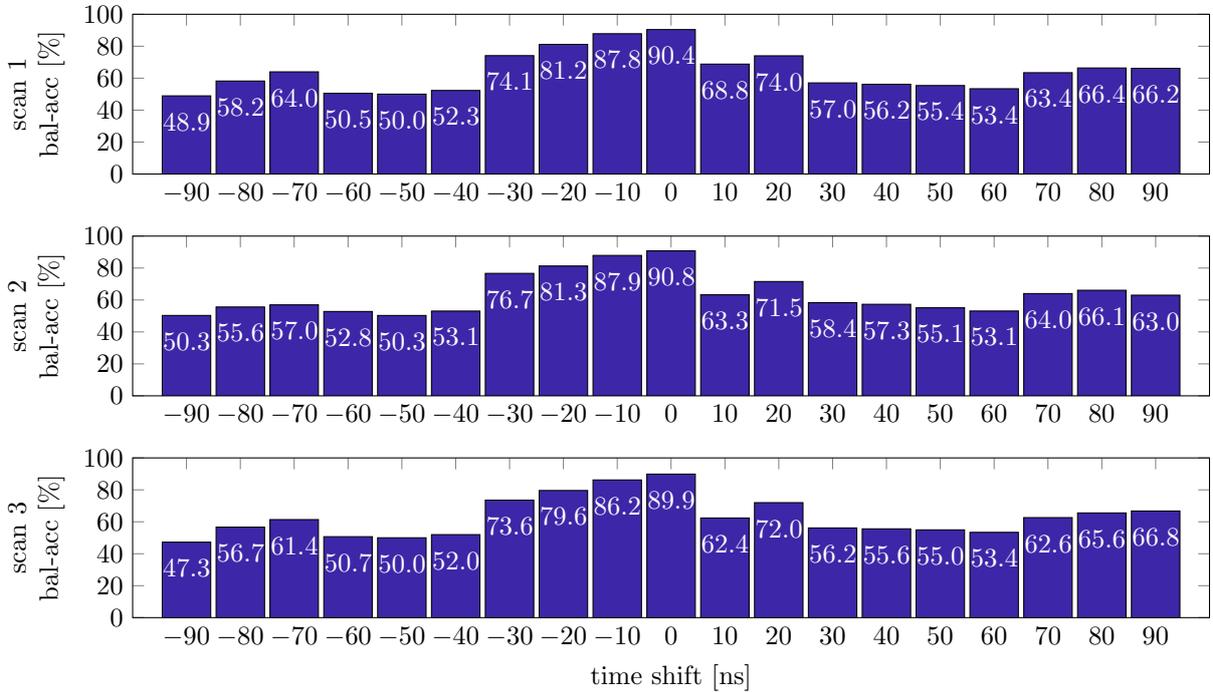


Figure A.2: Effect of the time shift on the classification result for FFT-linear regression. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -12 dB. Cf. section 4.4

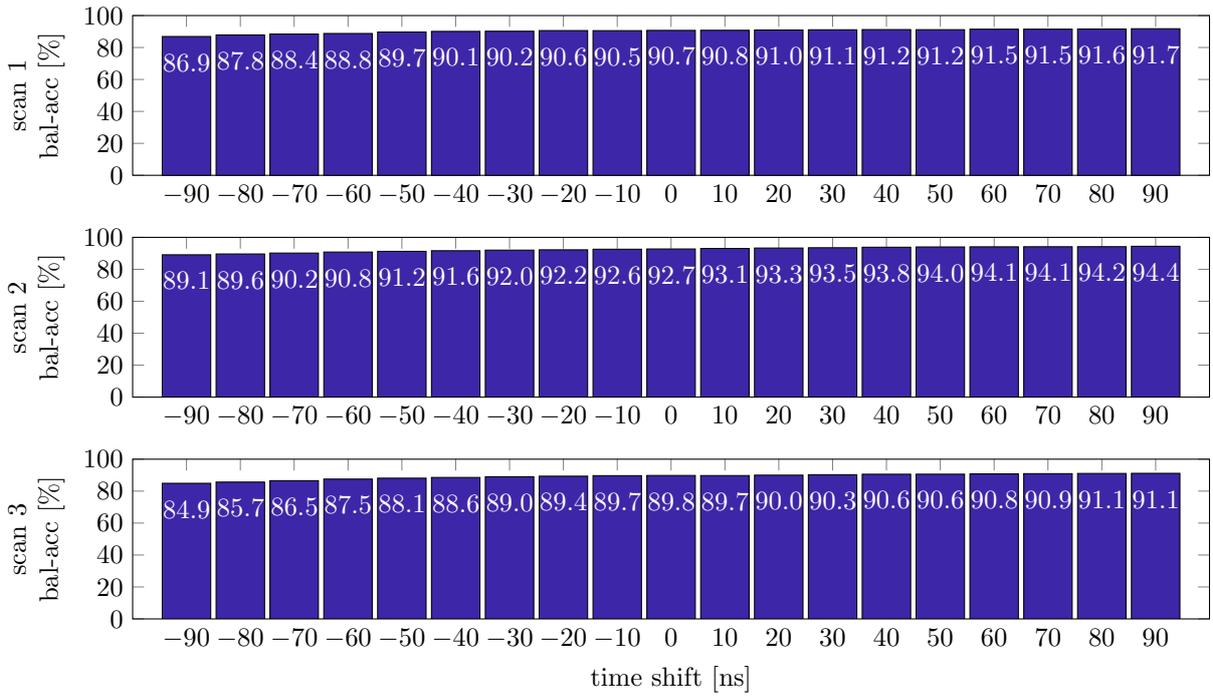


Figure A.3: Effect of the time shift on the classification result for FFT-linear regression with windowing and TCG. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -12 dB. Cf. section 4.7

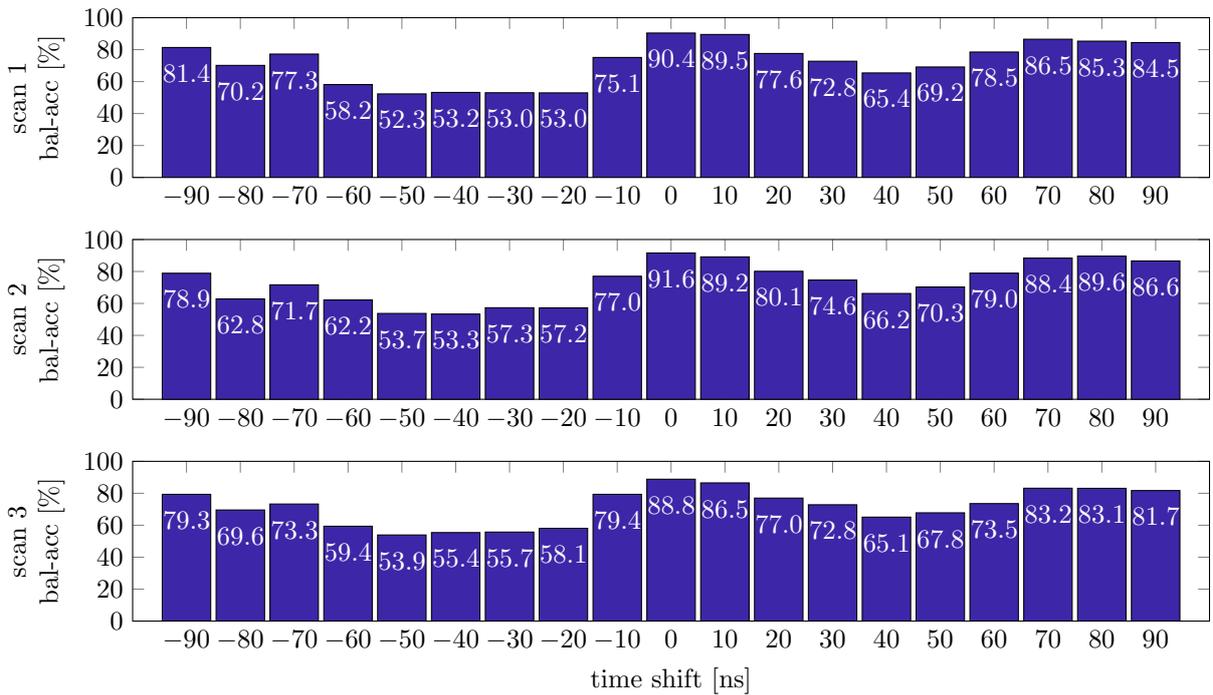


Figure A.4: Effect of the time shift on the classification result for DWT-based linear regression. The regression was trained on all scans of B50+56l and evaluated on each scan of B50+56r. The reference threshold is -12 dB. Cf. section 4.7

Validation on different data sets of a specimen (cf. section 4.8)

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	86.00	86.57	89.49	90.71	92.73	89.79	93.27	92.66	92.04	93.43	91.75	98.14
right	83.77	86.86	88.02	93.52	95.69	91.59	93.25	93.71	92.81	91.71	90.52	96.74
repeat	86.58	88.04	88.66	93.04	94.56	91.00	96.11	94.17	93.63	95.76	95.25	99.63
flipside	86.18	87.21	86.84	88.57	90.57	87.34	95.58	91.79	91.95	95.80	96.42	99.63

Table A.3: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

TPR on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	96.45	93.82	95.18	91.58	92.20	92.29	94.34	95.86	95.55	93.82	87.43	96.73
right	97.10	97.84	96.44	99.53	99.24	98.27	96.39	99.26	98.84	91.33	85.17	95.14
repeat	93.83	93.44	89.32	96.58	96.62	94.14	99.52	98.92	98.21	99.51	95.32	99.95
flipside	89.35	88.34	83.45	86.93	87.84	85.75	98.01	92.86	93.47	99.73	97.84	99.95

FPR on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	24.45	20.69	16.21	10.16	6.74	12.71	7.80	10.54	11.46	6.95	3.94	0.44
right	29.57	24.12	20.40	12.49	7.86	15.09	9.89	11.85	13.22	7.90	4.12	1.66
repeat	20.67	17.36	11.99	10.50	7.50	12.14	7.30	10.58	10.94	7.99	4.82	0.68
flipside	16.99	13.92	9.78	9.78	6.69	11.07	6.85	9.28	9.57	8.13	5.00	0.68

Table A.4: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are TPR (top) or FPR (bottom) in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

AUC on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	0.9238	0.9293	0.9390	0.9637	0.9787	0.9589	0.9619	0.9635	0.9617	0.9760	0.9826	0.9992
right	0.9037	0.9144	0.9241	0.9737	0.9844	0.9628	0.9551	0.9656	0.9644	0.9729	0.9772	0.9969
repeat	0.9340	0.9424	0.9541	0.9655	0.9778	0.9593	0.9764	0.9695	0.9705	0.9768	0.9832	0.9991
flipside	0.9360	0.9426	0.9530	0.9432	0.9673	0.9465	0.9784	0.9673	0.9672	0.9774	0.9846	0.9989

Table A.5: Validation on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

balAcc on: scan	left			right			repeat			flipside		
	1	2	3	1	2	3	1	2	3	1	2	3
left	85.09	85.17	87.17	89.69	91.24	88.11	91.87	91.20	90.37	90.64	88.10	96.23
right	83.63	86.04	87.25	93.45	95.19	91.19	92.95	93.71	92.73	89.46	87.60	95.32
repeat	86.47	87.37	87.33	91.99	93.79	90.08	96.03	93.75	93.03	95.57	94.55	99.56
flipside	84.82	86.01	85.95	87.78	89.80	85.89	95.58	91.53	90.69	95.87	96.39	99.63

Table A.6: Validation with a time shift of -50 ns (after training without shift) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

balAcc on: scan	left			right			repeat			flipside		
	1	2	3	1	2	3	1	2	3	1	2	3
left	85.62	86.69	89.20	91.16	94.00	90.60	93.89	93.50	92.71	94.72	94.31	98.89
right	83.85	86.80	88.42	93.49	95.79	91.75	92.85	93.41	92.38	92.18	92.01	96.95
repeat	86.30	87.67	88.89	93.39	94.88	91.53	96.03	94.18	93.66	95.74	95.68	99.66
flipside	86.27	87.31	87.13	88.59	90.63	88.11	95.37	91.46	91.84	95.77	96.11	99.66

Table A.7: Validation with a time shift of $+50$ ns (after training without shift) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

Restriction to parts of the intermediate echo gate for lower reference threshold (cf. section 4.9)

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	83.96	86.26	88.27	91.79	94.16	89.99	92.01	89.74	90.53	92.50	93.04	98.42
right	82.48	85.78	87.93	91.91	95.00	90.87	91.26	89.87	88.88	91.75	91.90	95.93
repeat	83.73	85.56	87.06	90.96	93.29	89.01	94.46	92.02	91.54	92.08	91.38	97.48
flipside	83.30	85.76	87.00	91.86	94.36	89.91	93.09	89.42	90.60	93.63	93.98	98.67

Table A.8: Validation using only the middle part of the intermediate echoes (1.5 μ s to 3.5 μ s) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	82.39	84.25	89.02	83.70	88.17	85.31	90.28	88.01	86.34	84.95	85.84	96.00
right	78.84	82.66	87.93	92.48	93.75	89.49	86.66	89.71	87.74	83.75	87.29	95.45
repeat	81.85	84.64	88.78	88.58	89.73	88.02	93.42	91.95	90.54	90.53	91.48	98.84
flipside	79.44	82.84	86.32	84.24	86.52	83.78	94.21	89.42	87.48	94.02	95.50	99.56

Table A.9: Validation using only the first half of the intermediate echoes (1 μ s to 2.5 μ s) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.

balAcc on:	left			right			repeat			flipside		
	scan	1	2	3	1	2	3	1	2	3	1	2
left	83.71	86.18	88.45	92.06	94.31	90.66	92.25	92.28	91.49	92.80	92.71	97.42
right	82.95	85.45	87.46	92.91	95.16	90.78	92.38	92.31	91.60	90.81	90.21	97.13
repeat	82.71	85.48	87.68	92.15	94.03	90.00	95.39	93.90	93.00	93.99	94.87	99.29
flipside	83.95	86.39	86.95	90.68	93.95	89.74	95.45	91.80	92.55	95.16	95.36	99.53

Table A.10: Validation using only the second half of the intermediate echoes (2.5 μ s to 4 μ s) on different data sets of the B50 and B56 specimens using FFT-linear regression with windowing and TCG. All values are balanced accuracies in percent. The first cell of each row specifies the data set used for training and the column header specifies the data set and scan number used for evaluation. The reference threshold is -12 dB.