Multivariable Regression Approach for Porosity Determination in Composite Materials

T. Olofsson M. Gustafsson H. Rydén

Abstract

The porosity content in composite material is known to influence the strength of the material. It is therefore of interest to monitor the porosity contents during manufacturing. Since attenuation and porosity are related, traditional porosity determination in composites is performed as attenuation measurements using ultrasonic tone burst through-transmission. In this paper we propose a multivariable regression approach for estimating ultrasound attenuation in composite materials by means of pulse-echo measurements, thus overcoming the problems with limited access that is the main drawback of through-transmission testing. The result from the work shows that we can obtain good approximations of the attenuation values using pulse echo ultrasound. This indicates that it will be possible to replace the through-transmission technique by a pulse echo technique.

Introduction

The increased use of composite materials in aircraft industry the last years has implied a growing need for efficient methods for nondestructive characterization of composite materials. One example is determination of porosity contents in composite specimens during manufacturing. Results have been reported [1], showing that the porosity contents can be estimated with good accuracy by utilizing a linear relation between the frequency dependence of the attenuation, i.e., $P = K\frac{d\alpha}{df} + l$, where P is the porosity content, K and l are constants and where $\frac{d\alpha}{df}$ is the slope of the attenuation curve, $\alpha(f)$.

At CSM Materialteknik AB the question has been raised whether the porosity estimation can be simplified further by utilizing attenuation measurements at a single frequency only. Empirical work show that for a given frequency, which is selcted based on the thickness of the inspected object, there is a correlation between the attenuation and the porosity contents. However, there has still not been any documented results proving that single frequency measurements are sufficient for porosity estimation.

One obvious disadvantage with both the approaches mentioned above is that the attenuation measurements are based on through-transmission, TT, testing which means that we need access to both sides of the specimen and this cannot be guaranteed for many of the complex geometries found, e.g., in the aircraft industry.

Based on the observation above, we maintain that it is of great interest to be able to replace TT measurements by pulse-echo, PE, measurements. The PE technique does not have the above mentioned disadvantage, but there will on the other hand be difficulties in analyzing the reflected signal because of the strong disturbing reflections caused by the complex internal structure of the material.

In this paper we propose a method to measure (estimate) the attenuation in composites by means of PE measurements using a multivariable regression approach.

Theory

Traditional vs regression approach to automatic material characterization

The traditional approach to automatic material characterization is based on *physical* reasoning where a set of features of the signals that we assume to be the most relevant for solving the characterization problem is selected. However, in situations with a complicated relation between the measurements and the material property to be characterized, this approach is *not always* applicable due to limited understanding of the underlying physical relations.

If the signal features already have been chosen, another important problem is how to optimally combine these features in order to obtain the best estimate of the material property. The physical reasoning will give us ideas of how to combine the features but there will be no guarantee that we are using the chosen features in an optimal way. One reason for this is that we have to take into account the uncertainties that always are present in measurement data.

We should also note that most of todays data acquisition systems are capable of producing enormous amounts of data which the traditional approach does not exploit for anything but verification of different ways to extract and combine features. To search in the space of all such combinations is however a tremendous task.

In a regression approach to material characterization, a statistical model which describes the relation between measurements and the material property is formulated and unknown model parameters are estimated from experimental data. This approach is attractive because it does not require a detailed physical model, and because it automatically extracts and optimally combines important features. Moreover, it can exploit the large amounts of data available.

As we have mentioned, the particular characterization task considered in this work is to determine attenuation in composite materials. At our hand we have a data acquisition system that can provide us with data from both PE and TT testing. The approach is to treat the attenuation problem as a multivariable regression problem where our target values, y_n, are the measured attenuation values (at different locations n) and where our input data are the (preprocessed) PE data vectors, $\mathbf{u}_n$. The problem is to find a function $\hat{y}_n = f(\mathbf{u}_n)$, such that $\hat{y}_n \approx y_n$, based on measured data, the so called *training data.*

We can use multivariate regression as a tool for extracting not only a suitable model relating our measurements to the attenuation but also as a tool for finding attenuation sensitive features. The steps involved in the traditional approach of first choosing the relevant features and then combine these optimally are in the regression approach merged into a single step. If the preprocessed data can be interpreted as physical entities, we might also be able to interpret the model parameters in the multivariable regression. For instance, if we limit ourselves to *linear* regression, the estimate will be a weighted sum of the input data elements and the magnitude of the weights can therefore give important information about what features in the input data are relevant for estimation of the attenuation.

Multivariable Regression

The regression problem is here formulated as the optimization problem

$$\min_{w} J = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 = \frac{1}{N}\sum_{n=1}^{N}(y_n - f_w(\mathbf{u}_n))^2 \qquad (1)$$

where y_n is the n-th target value, D is the dimension of the observation vectors $\mathbf{u}_n = (u_{n1} \ldots u_{nD})^T$ and the function f_w is chosen within a family of functions that are parameterized by a parameter vector $\mathbf{w}$, i.e., $f_w(\mathbf{u}_n) = f(\mathbf{w}, \mathbf{u}_n)$. N is the number of observations.

Assume that we choose f_w to be a linear function of $\mathbf{u}_n$, i.e.,

$$\hat{y}_n = f_w(\mathbf{u}_n) = w_0 + \sum_{d=1}^{D} w_d u_{nd} = w_0 + \mathbf{w}^T \mathbf{u}_n \tag{2}$$

where the model parameters are the scalar w_0 and the elements of the vector $\mathbf{w} = (w_1\, w_2 \ldots w_D)^T$. In the linear case the optimal solution to the minimization problem (1) is given by [1]

$$\mathbf{w} = R_u^{-1} R_{yu}^T \qquad \text{and} \qquad w_0 = m_y - R_{yu} R_u^{-1} \mathbf{m}_u \tag{3}$$

where $m_y = \frac{1}{N}\sum_{n=1}^{N} y_n$ and $\mathbf{m}_u = \frac{1}{N}\sum_{n=1}^{N} \mathbf{u}_n$. $R_{yu} = \frac{1}{N}\sum_{n=1}^{N}(y_n - m_y)(\mathbf{u}_n - \mathbf{m}_u)^T$ is the empirical cross-covariance matrix between $\mathbf{u}$ and y and $R_u = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{u}_n - \mathbf{m}_u)(\mathbf{u}_n - \mathbf{m}_u)^T$ is the empirical covariance matrix of $\mathbf{u}$. Eq. (3) and eq. (2) yield the estimates

$$\hat{y}_n = f_w(\mathbf{u}_n) = m_y + R_{yu} R_u^{-1}(\mathbf{u}_n - \mathbf{m}_u) \tag{4}$$

The linear solution (3) is known as the *ordinary least squares*, OLS, solution. An analysis of OLS shows that the variance of the solution (3) can become large if R_u is nearly singular. This problem occurs typically when the input data vector $\mathbf{u}$ has large dimension. Then we can expect redundancy and thus strong correlation between some elements in $\mathbf{u}$ which will result in a nearly singular or singular matrix R_u. As we mentioned earlier, we are interested in finding the most informative features of the measurements, meaning that we ideally want to examine as large an input vector as possible and thus we will have problems with an ill-conditioned R_u.

Some methods that partly cope with the above mentioned problem have been proposed in the literature. The subject has been treated in areas like Chemometrics, Econometrics etc, giving rise for example to the methods Partial Least Squares, PLS, Ridge Regression, RR, and Principal Component Regression, PCR [2]. In this work we have chosen to illustrate the multivariable approach using PCR as our regression tool, mainly because it has a relatively easy interpretation. The basic idea of PCR is described below.

Let $\mathbf{u}$ be a vector valued stochastic variable with dimension $D \times 1$ and with covariance matrix R_u of size $D \times D$. The key idea is to linearly transform all observation vectors, $\mathbf{u}_n$, to new variables, $\mathbf{z}_n = W^T \mathbf{u}_n$, and then solve the optimization problem (1) where we replace $\mathbf{u}_n$ by $\mathbf{z}_n$. We choose the transformation so that the covariance matrix of $\mathbf{z}$ is diagonal and (more importantly) none if its eigenvalues are too close to zero. (Loosely speaking, the eigenvalues close to zero are those that are responsible for the large variance of the OLS-solution). In order to find the desired transformation, a singular value decomposition of R_u is performed yielding

$$R_u = T \Lambda T^T \tag{5}$$

[1] More correctly, the regression problem involves means instead of averages in (1). Furthermore, when the criterion function is quadratic, the general (usually nonlinear) optimal solution is given by $\hat{y}_n = E[y|\mathbf{u}_n]$, i.e., the conditional mean of y given the observation $\mathbf{u}_n$.

where Λ is a diagonal $D \times D$ matrix containing the eigenvalues of R_u, $\{\lambda_1, \lambda_2, ..., \lambda_D\}$, in decreasing order as diagonal elements. T is a matrix containing the eigenvectors, $\{t_1, t_2, \ldots t_D\}$ corresponding to these eigenvalues as column vectors, i.e., $T = (t_1\, t_2\, \ldots\, t_D)$. The new, transformed input vectors, are formed by

$$\mathbf{z}_n = T_m^T(\mathbf{u}_n - \mathbf{m_u}) \tag{6}$$

where T_m denote a $D \times m$ matrix that only contain the eigenvectors corresponding to the largest eigenvalues, i.e., $T_m = (t_1\, t_2\, \ldots\, t_m)$. Note that $\mathbf{z}$ is a zero-mean stochastic variable with dimension $m < D$. We choose the parameter m as the index where the decreasing sequence $\{\lambda_1, \lambda_2, ..., \lambda_D\}$ goes below a certain threshold value, θ, i.e. $\lambda_m \geq \theta, \lambda_{m+1} < \theta$, thereby getting dangerously close to zero. Setting the threshold is a subjective choise and that is one of the disadvantages with PCR. Unfortunately we cannot determine whether the information that we throw away is useless for the characterization task, we just know that our solution will have less variance. By solving the optimization problem (1) with $\mathbf{u}_n$ replaced by $\mathbf{z}_n$ we obtain the estimates

$$\hat{y}_n = m_y + R_{yz}R_z^{-1}\mathbf{z}_n = m_y + R_{yu}T_mR_z^{-1}T_m^T(\mathbf{u}_n - \mathbf{m}_u) \tag{7}$$

The last equality is found by using the relations $R_{yz} = R_{yu}T_m$ and $\mathbf{z}_n = T_m^T(\mathbf{u}_n - \mathbf{m}_u)$. Eq. (7) finally gives us the weights, $\mathbf{w} = R_{yu}T_mR_z^{-1}T_m^T$, for our regression model.

Experiment

Through transmission measurements

During the attenuation measurements, Transducer 1 was excited with a narrowband tone burst with center frequency 18 MHz, see Figure 1 for a schematic setup. The amplitude of the sound pressure was measured at Tranducer 2 by means of an amplitude peak detector. A reference amplitude, A_{ref}, was measured outside the object as shown at the right hand side of Figure 1. The object was scanned in the xy-plane and for every position, (x, y), the attenuation, $\alpha(x, y)$, was calculated as the quotient (in db) between the amplitude at Transducer 2, $A(x, y)$, and A_{ref}, i.e., $\alpha(x, y) = 10 \log_{10} \frac{A(x,y)}{A_{ref}}$.

Pulse echo measurements and preprocessing

The PE data was obtained by repeating the scanning of the object, now measuring the received echo at Transducer 1. For every position, (x, y), an A-scan was obtained from which we extracted the back wall echo by means of a time gate. This back wall echo is denoted $\mathbf{s}(x, y)$. Note that $\mathbf{s}(x, y)$ is a time signal that can be written $\mathbf{s}(t, x, y)$ where t is the time index. One example of such a back wall echo is shown in Figure 2.

Because of the double sound path involved in PE measurements of the back wall echo, we approximate the corresponding attenuation at a certain frequency to be twice as large as the attenuation that would be obtained by an ordinary TT measurement. We propose to use the logarithm of the absolute value of the Fourier transform of the back wall echo as input data, i.e

$$\mathbf{u}(x, y) = \ln\left(|\mathcal{F}(\mathbf{s}(t, x, y))|\right) \tag{8}$$

where $\mathcal{F}(\mathbf{s}(t, x, y))$ is the Fourier transform of $\mathbf{s}(t, x, y)$ (taken separately for each component) with respect to time. Since the amplitude spectrum for a real-valued signal is symmetric, only half of the spectral coefficient are needed. This choice of preprocessing method should, provided the approximation mentioned above is valid, yield a linear relation between our preprocessed data and the target values. With a linear relationship, a linear regression approach using PCR is natural.

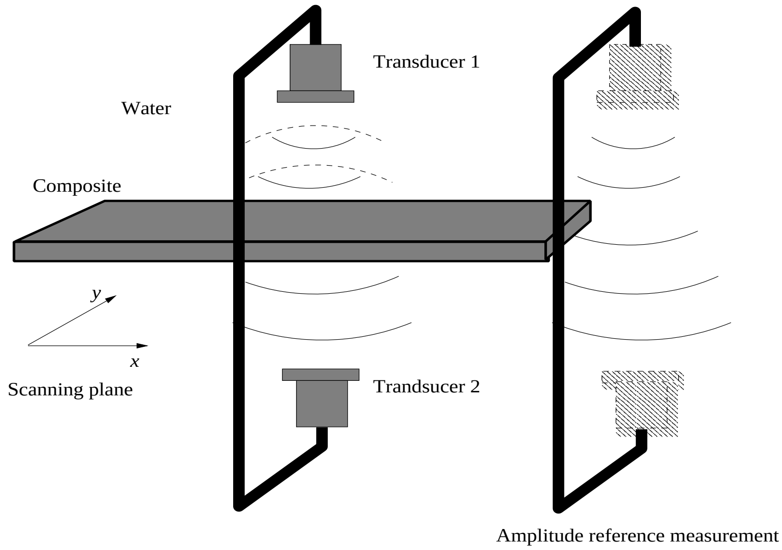

Figure 1: Schematic view of the measurement setup used in the experiment

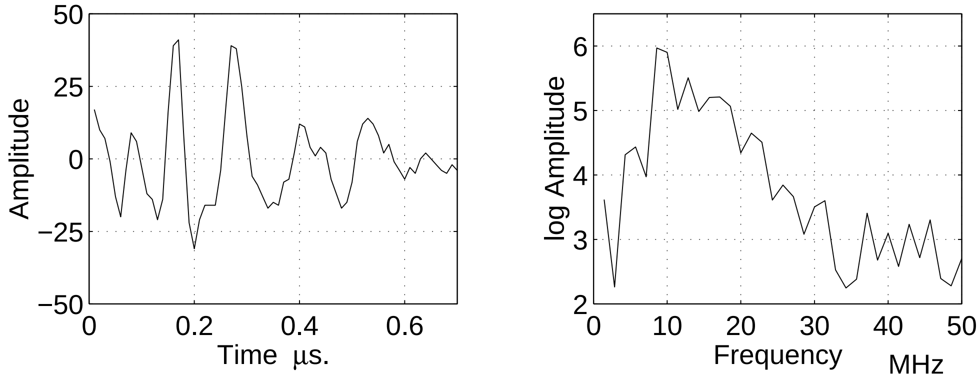

Figure 2: Example of back wall echo (left) and preprocessed back wall echo (right).

Results

The results presented below were obtained using a 2 mm thick carbon fiber reinforced epoxy composite laminate with 16 layers. The laminate was quasi isotropic with fiber orientations: 0°, 90° and ±45°. The laminate had an average porosity content of approximately 1.7%. The object was divided in a *training area* and an *evaluation area*. The model parameters were determined by data solely from the training area. Both ultrasound tranducers used in the experiment had a center frequency of 21 MHz and a 6 dB bandwidth of 70%.

The back wall echoes were sampled at 100 MHz and the length of these were 70 samples, yielding a size of the input data vectors, $D = 35$. An example of such an echo is shown in Figure 2 together with its log spectral amplitude.

In our experiment we used thresholing value $\theta = 0.025$, yielding $m = 25$ and a ratio between the smallest and the largest eigenvalue $\frac{\lambda_m}{\lambda_1} = 0.01$.

In Figure 3 we see how the logarithm of the spectral amplitude effects the estimation results. For each component in input data vector, u_d, we have defined the feature relevance, $F_R(d)$, as the weight in the regression model multiplied by the standard deviation

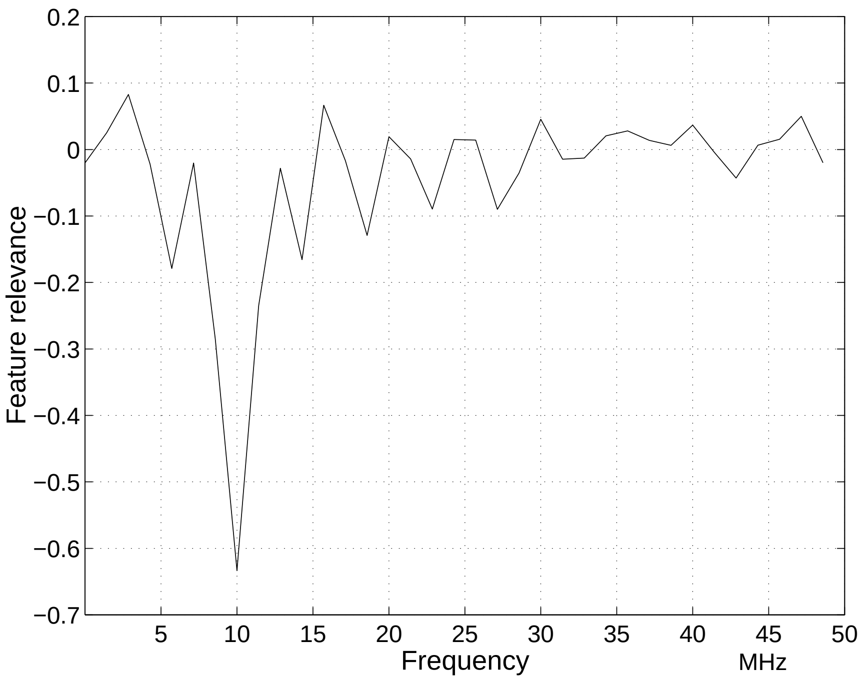

Figure 3: Feature relevance. The weight parameters for every component in the input vector multiplied with the standard deviation for that component are plotted. This is a measure of the significance of this feature (in this case, the logarithm of the power in a small frequency region.)

of the corresponding log amplitude spectral value, i.e., $F_R(d) = w_d std(u_d)$. The multiplication is because components with large standard deviation will have more influence on the estimate than those with less. By examining the plot we find that the important features for estimating TT attenuation are the log spectral amplitudes around 10 MHz. Note that the frequency resolution is inherently bad because of the short A-scan segment that have been used for calculating the Fourier transform.

The results in Figure 3 illustrate how we by using the regression approach can find relevant features in an automatic manner. We can also interpret the extracted features physically. Maybe we would expect the log amplitude of the frequencies around 18 MHz to be the most relevant feature but the results in this example indicate that we instead should put more weight to the frequencies around 10 MHz, at least when performing the PE measurements and the preprocessing as was described above. One interpretation of this result is that since the higher frequencies are attenuated more heavily then the lower, we loose much energy (or loosely speaking, information) in these frequencies and therefore the lower frequencies will bear comparatively more information. However, it is important to note that the extraced weights are closely related to the preprocessing method used and, equally important, to how the measurements are performed.

In Figure 4 the measured attenuation values (TT) and the corresponding estimates are plotted against each other. Ideally (with error free estimates) all sample points should lie on the straight line through the origin with unit slope. Clearly there is a strong correlation between the estimates and the true values.

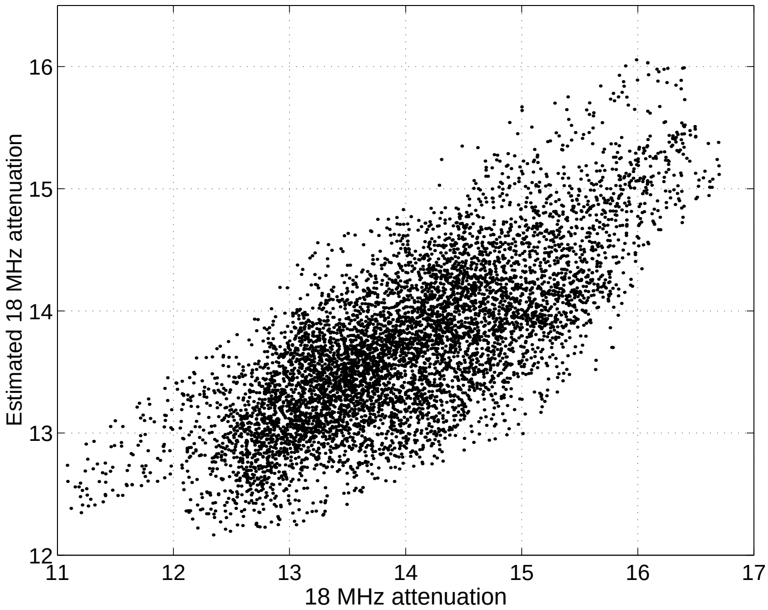

Figure 4: Scatterplot. Measured attenuation values on x-axis and estimated on y-axis. All samples are from the evaluation area

In Figure 5 a C-scan produced by the estimates and corresponding measured C-scan are presented. By comparing the C-scans in Figure 5 we see that the PE estimates in most cases exhibit the global variations that are present in the TT measurements. Note that we have performed a (spatial) lowpass filtering on both C-scans in order to emphasize these global variations. The lowpass filtering is mainly to help in a visual comparison that otherwise would be difficult due to the strong horizontal, vertical and diagonal lines caused by the fiberorientations at different layers that would mask much of the imagages. Lowpass filtering can also be motivated with the original (porosity estimation) problem in mind. Porosity contents is mainly a statistical property that only have meaning when considering a (small) area of the C-scan.

It is worth to mention that the approximation is almost as good in the evaluation area as in the training area. In other words, we seem to have found a regression model with good generalization properties.

Conclusions and future work

We have proposed a multivariable regression approach for the estimation of ultrasound attenuation in composite materials by means of PE-measurements, thus overcoming the problem of limited access that is the main weakness of TT testing. Estimating the attenuation is a step towards determining the porosity contents, something that is of interest, for instance, in composite manufacturing.

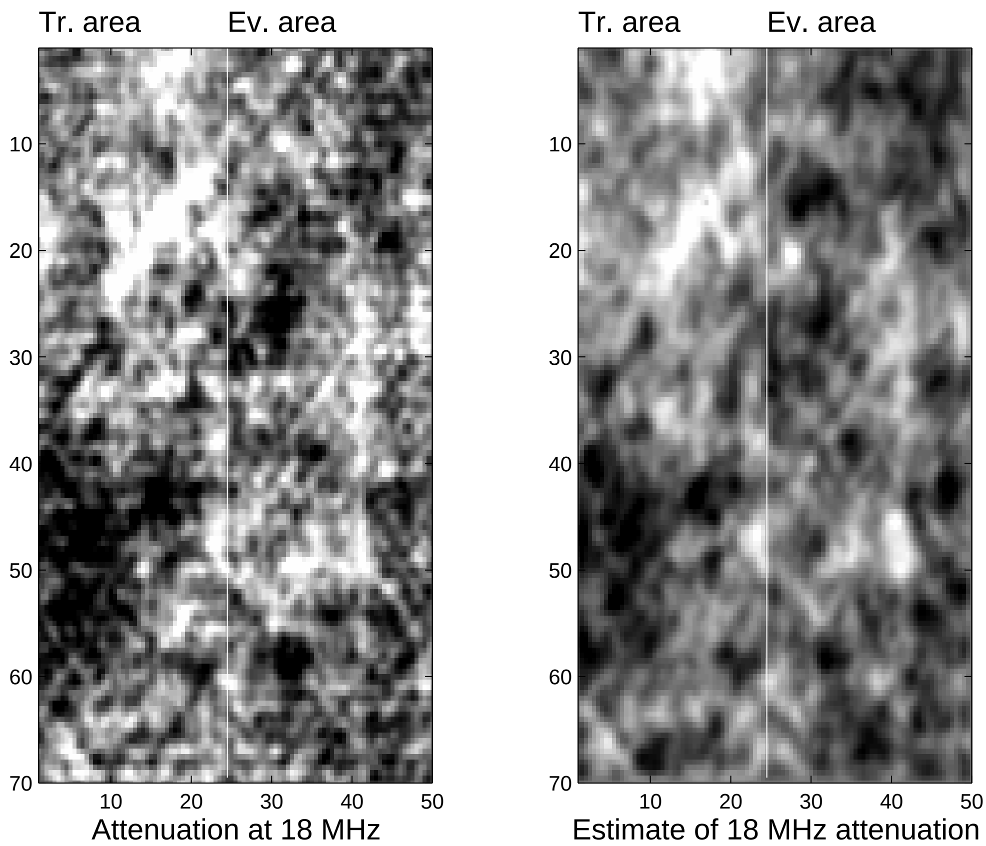

Figure 5: Approximation results. Measured 18 MHz attenuation presented as a C-scan (left) and corresponding estimated attenuation based on PE-data (right)

The regression approach can be used in an iterative procedure to develop a material charactarization system. By investigating the model parameters we can determine the features in the input data which are most significant for solving the estimation problem. This information can be fed back to modify the testing procedure to get more accurate measurement of these suggested significant features.

In this preliminary work we have investigated composite objects with a simple geometry. In future work the proposed approach will be applied to more complicated objects, in particular glued structures. Since we for such objects expect to have a less distinct back wall echo, we have reason to believe that the preprocessing method that was used in this work has to be somewhat modified.

It would also be of interest to investigate if the attenuation estimates can be further improved by extending our input data vectors. Since attenuation (and porosity) is spatially correlated, we should expect improvements when including data from A-scans in a neighbourhood around the point of interest. This is also a topic for future work.

References

[1] F.A. Reed, T. Batzinger, R.W. Reed, and S. Jönsson. Porosity measurement of composites using attenuation methods. *Review of Progress in Quantitative Nondestructive Evaluation*, 12, 1993.

[2] H. Schmidli. *Reduced Rank Regression.* Physica-Verlag, 1995.