STATISTICAL ANALYSIS OF FIBER COMPOSITE INTERPHASE INVERSE PROBLEM

by

Steven A. Cimaszewski

Bachelor of Mechanical Engineering
Villanova University (1992)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1994

Author .....
Department of Mechanical Engineering
May 6, 1994

Certified by .....
James H. Williams, Jr.
SEPTE Professor of Mechanical Engineering
Thesis Supervisor

Accepted by .....
Professor Ain A. Sonin
Chairman, Department Graduate Commitee

STATISTICAL ANALYSIS OF FIBER COMPOSITE INTERPHASE INVERSE PROBLEM

by

Steven A. Cimaszewski

Submitted to the Department of Mechanical Engineering
on May 6, 1994, in partial fulfillment of the
requirements for the degree of
MASTER OF SCIENCE IN MECHANICAL ENGINEERING

Abstract

Interphases — regions between the composite matrix and its fibers — influence the performance of composite materials. Ultrasonic nondestructive evaluation (NDE) is an attractive technique for characterizing the physical and geometrical properties of fiber composite interphases. Multivariate correlation and multiple regression techniques are used to describe relationships between interphase parameters and NDE output wavefields.

The interphase thickness parameter is predicted through linear and nonlinear multiple regression. It is shown that knowledge of one or more interphase parameters such as density, Poisson's ratio and modulus of elasticity can help to improve estimates of interphase thickness. The methods are not specific to the inverse problem under study; on the contrary, they can be applied to the general NDE inverse problem.

Thesis Supervisor: James H. Williams, Jr.
Title: SEPTE Professor of Mechanical Engineering

Acknowledgments

I thank Shiva and Hyunjune for sharing their research experience with me. Shiva was always a prolific innovator and Hyunjune possessed detailed insights into this problem which proved invaluable. Liang-Wu provided much-appreciated tips to facilitate word processing.

I thank Professor James H. Williams, Jr. for his encouragement that I work through the details of my research until they were sharp. To him I attribute the efficacy of my writing.

Thanks to all the people here in Cambridge who have made life interesting: Doug, Laurie, Ravi, Bill and the Roland factor.

I gratefully acknowledge the Department of Defense for its award of the National Defense Science and Engineering Graduate Fellowship, without which this work would not be possible.

Finally, thanks to my parents who have always supported my interests and encouraged me to become whatever it was that I felt in my heart.

I dedicate this work to my fiancée Yvonne Ann, who has been a source of love and support from the beginning.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Foreword

Composites are at the forefront of materials technology today. They have had copious research interest for over two decades and still are of primary importance to materials and engineering professionals today. New applications for composite materials are being discovered as the technology "matures" and emphasis is being shifted away from the defense interest so paramount in the recent past.

These materials have great potential yet to be exploited, since some areas of composite science and manufacture are still misunderstood. Composite performance can be improved and cost can be reduced as advances in research of automated manufacturing, joining technologies, and curing science become available. Innovations in these and other parallel disciplines will certainly begin to eliminate some of the obstacles to future composite progress.

One unique characteristic of composite materials is their ability to assume different properties utilizing the same constituent material. This characteristic, called *tailorability*, allows the design of the end product to be intimately related to the design and selection of the material. In the past, parameters such as constitutive material types, volume fractions, reinforcing patterns, and fiber coatings have been changed to tailor properties in fiber properties.

It has been shown that there exists a distinct material layer, called the *interphase*, between the fiber and the matrix which is critical to the overall properties of fiber composites [1, 2]. Other terms for this intermediate layer are mesophase, interlayer, or reaction zone. The interphase layer is a result of physiochemical reactions which occur during composite processing, and is used as a micromechanics model of the boundary layer between the two phases of fiber and matrix. Or, the interphase may be designed into the composite. Interphase physical and geometrical properties and features directly influence the stiffness, strength and failure mode of the composite, which is the reason it has been the subject of investigation [1, 3].

If it were possible by conventional techniques to evaluate this interphase, composite performance could be better predicted from such knowledge of the interphase. However, interphase properties are difficult if not impossible to determine without specialized laboratory techniques. Some of these techniques, for instance thermo-gravimetric analysis (TGA) or metallographic evaluation, do not permit service usage of the material subsequent to testing. For this reason, nondestructive evaluation (NDE) techniques have been suggested for implementation in this problem. Various measurement techniques are available, such as β-backscattering, X-fluorescence method, and acousto-ultrasonic methods. Problems arise in applying these methods universally to interphase evaluation because the applicability is limited by characteristics of the interphase layer such as conductivity, permeability, and absorption [4]. However, acousto-ultrasonic testing methods do not suffer these shortcomings and are deemed appropriate for composite material property evaluation. Acousto-ultrasonic methods typically utilize input waves in the frequency range 20 kHz – 100 Mhz. It has been shown that properties such as modulus can be determined from acousto-ultrasonic testing, with very good accuracy, in materials such as concrete, ceramics, and composites [5].

1.2 The NDE Inverse Problem

NDE is essentially an inverse problem: using a (presumably) nonunique set of NDE outputs, one is to determine the unique set of inputs which created the output set. For the problem at hand, the NDE researcher has equipment capable of sending a particular wave through a material and of measuring stresses and displacements at the surface(s) of that material. Thus, one can only be sure of the relationship between material state and measurable stresses when one knows both. Typically a researcher will test the techniques on specimens of known material state and then use this information as a database to study unknown material states.

In the case of interphase properties, it can be shown that the problem is characteristic of an *inverse-source problem*, or inverse problem [6]. In order to solve this inverse problem, linear and nonlinear regression techniques are proposed. The proposed method will create the relationship between measured output stresses and material state, so that it can be used to learn about unknown samples. This in turn will find one solution to the inverse problem, the most likely one. More complex methods can then incorporate this technique to find a more complete solution to the general inverse problem.

1.3 Motivation

The field of NDE is a small but very active field of research. It has been growing and will continue to grow over the next decade into the twentieth century. Papadakis [7] predicts ultrasonic testing to be a part of this growth. Specifically, it has been projected that ultrasonics will be used to determine material property correlations. Another point from this same author [7] is the importance that artificial intelligence (AI) will have in the development of NDE. Ultrasonics will have to be interfaced with AI to produce new "solution packages." Some strides in this direction already taken include: automatic learning networks, expert systems, fuzzy logic, and approximate reasoning. Such disciplines will increase the utility of ultrasonic techniques in the future, and will be needed to extract information from the signal processing used for interpreting NDE data.

Data analysis is named as critical to the successful future of NDE. Vary [8] writes that there is a need for models which identify the degree to which various microstructural factors interact and govern mechanical properties and structural response. Such models would guide and focus interrogation, for unraveling microstructure-property interrelations would also require more advanced tools such as computational simulation of wave propagation. All of these techniques would serve as the bases needed for interpretation of NDE data.

1.4 Literature Review

Regression is not a new topic. The term "regression" originated with the work of Francis Galton [9]. The studies of inheritance inspired by the work of Charles Darwin led Galton to believe that everything could be studied quantitatively. One of Galton's studies involved the linear trend between the heights of fathers and their sons (see Fig. 1-1). The slope of the trend line in this particular study was positive but less than one, so Galton called the relationship a "regression toward the mean." The term "regression" was then applied to any linear trend. It was an unfortunate term, however, because the slope of a least-squares trend line need not be less than one. Be that as it may, its applications are as widely varied as the social sciences, economics and engineering. It continues to be developed through the present day.

For the purpose of predicting material property parameters, regression techniques have been utilized before albeit not to a great degree. One study performed by Grotz and Lutz [10] investigated the relations between material characteristics and electromagnetic quantities, particularly eddy current signals. An analysis of piston pins was completed wherein it was desired to predict surface hardness, case-hardening depth, and core tensile strength from regression utilizing eddy current signal components as predictors.

Schneider, Schwarz and Schultrich [4], developed a method to determine the thickness and the elastic modulus of surface layers simultaneously from surface wave dispersion. Here the inverse solution of the surface wave dispersion equation in a homogeneous,

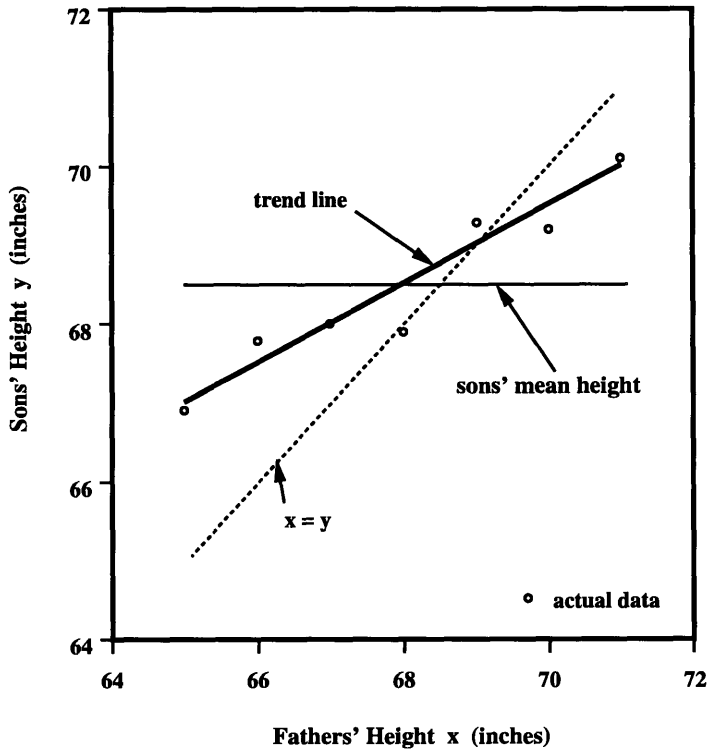

Figure 1-1: Representative plot of fathers' and sons' heights, according to Galton.

isotropic material coated with a homogeneous, isotropic layer was carried out by nonlinear regression. The results demonstrated the power of the multidimensional algorithm for material property inferences and the value of ultrasonic testing for material property estimation.

Regression analysis was applied to hardenability prediction. Two papers, the first written by Lund [11] and the second by Thomas, et al. [12] discussed the use of forms of regression analysis to address the issue of data scatter. Both implemented the compositional elements of metals as the regressor variables in the regression equation. The benefit of this approach was that a material property (hardenability) could be linked with a microstructural characteristic (composition), thereby enabling one to ascertain exactly what effect a change will have, in this case a change in composition. It was reported that regression techniques are the most effective solution to the problem, having succeeded in explaining the variation better than any other available methods.

In the composites area, multivariate statistical analyses were used for failure predic-

tion. Hill [13] discussed predicting the burst pressures for filament-wound composite pressure vessels using acoustic emission data. Specifically, acoustic emission amplitude and energy measurements during a low proof test were used as the regressors in the analysis. In addition, variables such as cure mode, burst temperature, preimpregnated material batch, and pressurization scheme were included in the analyses. While the methodology appears sound and the published results seem useful, there are certain intricacies of the analyses which preclude the conclusions. In any particular analysis, no more than eight points were used. This data set is not likely to be representative, random, nor sufficiently large. Depending on the number of degrees of freedom, this will almost certainly provide a good result due to inherently low variation in data.[1] Another intricacy was that for certain models data were **omitted** and identified as outlier. This was done on the basis that the said data points did not fit the model. In this case a different model should have been considered. Important however was verification of the correlation between certain acoustic emission parameters and composite defects. The work further demonstrated the connection between material parameters and acoustic emission test data.

Perhaps the most skillful and thorough use of regression analysis in the literature was performed by Capener [15]. The project was to ascertain the ability of ultrasonic testing to evaluate potential flaws due to intergranular stress-corrosion cracking (IGSCC) in boiling-water reactor piping. Regression analyses of the ultrasonic testing depth measurement upon transformed variables of the actual depth and thickness dimensions of the flaw were performed. The end result was a statistically significant model which described the ultrasonic data as a function of the actual flaw depth and thickness. Although this model would not provide the desired actual dimensions explicitly, it can be inferred that such information would be available through comparison with other testing methods. The conclusion was that ultrasonic testing has weaknesses for estimating flaw size in this application, and other methods and techniques should be included into the design of ultrasonic performance testing. No predictive value was obtained from this model,

[1]For instance, consider the case of fitting three data points. A second order fit is able to model these exactly as a curve, in the form $y = \alpha + \beta_1 x_1 + \beta_2 x_1^2$. See reference [14].

but statistics uncovered the problem of assuming this testing was sufficient for flaw characterization. The numerical details in this particular study were well documented, and the procedure offered a brief explanation for many statistical parameters.

A study was published by Harris [16] to refine the selection of parameters to consider for inclusion into a model. Stepwise regression optimally reduced the unprocessed batch of variables to a refined set. The author [16] reported that inspection success rate was 400 percent higher when those variables identified by the stepwise analysis were present in inspection. Grabec, Sachse and Grabec [17] used non-parametric multidimensional regression to evaluate the surface roughness for certain manufacturing processes. Alternatively they [17] offered a system which includes neural networks to discriminate acoustic emission (AE) signals generated during the process in order to characterize the surface roughness. This very recent work is symbolic of the position that artificial intelligence methods will assume in the future.

There are interesting examples of research available from many disciplines which utilize statistics. For instance, Zhang [18] used multivariate statistical techniques for low cycle fatigue life prediction, and Saniie [19] utilized them for ultrasonic grain signals classification. Roth, et al. [20] discussed a complete analysis of ultrasonic velocity-pore fraction relations in the context of statistics for polycrystalline materials. These are mentioned only as a sample of the wide cross section which has been researched using statistics.

Thus, the previous work has shown that regression analysis is very capable of solving problems of the inverse, predictive nature, especially for interests in material properties. Also, these methods have been applied to NDE data in the laboratory. Coupled with the motivation discussed in a previous section, regression analysis appears promising for application to the problem of predicting interphase properties.

1.5 Objective

In the case of fiber composite interphase properties, a direct correlation between interphase properties and the output stress wavefields resulting from traditional acousto-ultrasonic testing has not been proven. In earlier work [21], it has been shown that ultrasonic data can reveal characteristics about the interphase, but under general circumstances no direct correlation is available. The goal of this thesis is to strengthen confidence in the relationship between ultrasonic stress wavefields and interphase properties, and also to develop a system to describe such relationships.

Since the relationships between interphase properties and ultrasonic output wavefields are quite complex, multiparameter estimation techniques are suitable for uncovering those relationships. Multiparameter techniques can explain trends in terms of actual variable quantities, such as the components of a wavefield. Certain methods are also good for prediction; it is an aim of this work to investigate these methods. A data set as large as the one to be considered here has not been addressed in the literature; this data set represents the **entire domain** of the interphase problem, not a specific range of consideration. While this task may be more formidable, information learned through it will be potentially more valuable.

Given a particular ultrasonic output stress wavefield, one should be able to quantitatively identify characteristics of the composite interphase which affected it. Various schemes for this assessment based on multivariate regression techniques, and their performance, are presented.

Through the above problem, it is desired to demonstrate the feasibility of statistical methods for solving the general NDE inverse problem.

Chapter 2

Solution Strategy for the Interphase Inverse Problem

2.1 Problem Definition

The material selected for study of interphase property prediction is a metal matrix composite (MMC). The matrix is aluminum (type AA 520) with alumina (Al_2O_3) fibers and a fiber coating of zirconia (ZrO_2) [22, 23]. The material properties of each of these constituents are found in Table 2.1 [24-29]. Several sources have been utilized to confirm the accuracy of these properties.

A single fiber elastic model [30] is used to approximate the effect of changes of interphase properties on NDE ultrasonic output. Briefly, the three-constituent MMC is modeled as a single cylindrical fiber scatterer surrounded by an interphase layer, embedded in an infinite matrix material. An excitation in the form of a steady-state plane longitudinal (P) or in-plane shear (S) stress wave is assumed at the boundary and the corresponding stress state is determined within the three distinct regions of fiber, interphase and matrix. The constituents are assumed to be elastic, and the interfaces between the constituents are perfectly bonded. The single fiber elastic model is sketched in Fig. 2-1.

The interphase is identified by its physical and geometrical properties. They are thickness, density, Poisson's ratio, and modulus of elasticity. Since the interphase properties

Table 2.1: Constituent Material Properties.

	Matrix: AA 520 aluminum	Interphase: zirconia	Fiber: alumina
Elastic modulus	66 GPa	97 GPa	360 GPa
Density	2.6 g/cm^3	6.3 g/cm^3	3.7 g/cm^3
Poisson's ratio	0.31	0.33	0.25

of a particular composite are fixed, for purposes of study the properties are considered to vary. In this way an unknown interphase can be evaluated, presumably without knowledge of the actual interphase property values. In particular, the interphase thickness is varied over a large range, in order to study thickness effects on the output wavefield. The interphase parameters are varied about the nominal values of the interphase properties listed in Table 2.1 to create the ranges of consideration.

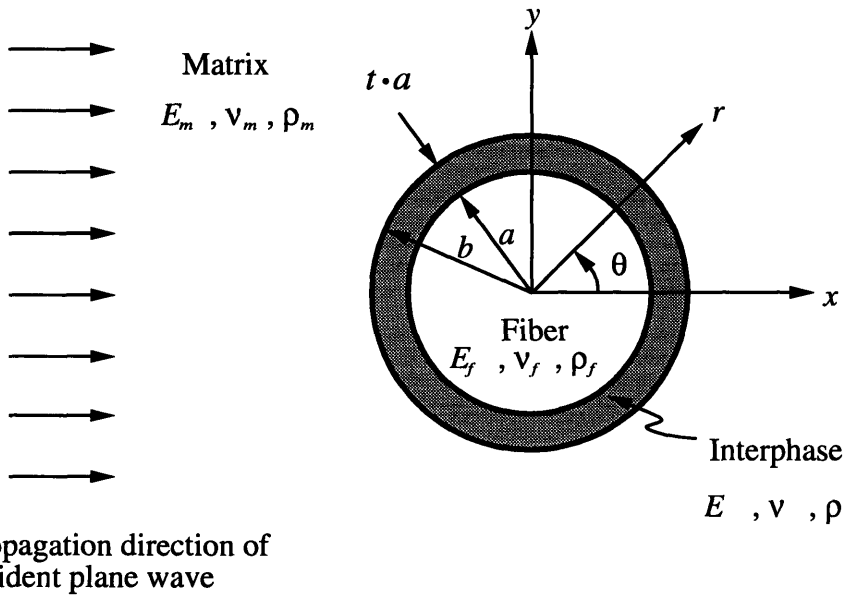

Figure 2-1: Two-dimensional model of single fiber scatterer.

An output wavefield is created for each set of interphase parameters considered. The stress components of this wavefield are shown in Fig. 2-2. Two locations of particular interest are labeled A and B in Fig. 2-2. Location A has r,θ coordinates of $(20a,0)$ and location B has r,θ coordinates of $(20a,\pi)$. The wavefield components at location A have the additional subscript '1', and those at location B have the additional subscript '2'.

Thus, the stress components at location A are σ_{rr1}, σ_{tt1} and σ_{rt1} and the stress components at location B are σ_{rr2}, σ_{tt2} and σ_{rt2}. Displacement wavefield components u_{r1} and u_{t1} exist at A and u_{r2} and u_{t2} exist at B. Also, the wavefield components at location A are called the *transmitted* wavefield components and the wavefield components at location B are called the *scattered* wavefield components. The scattered wavefield is the total wavefield minus the incident wavefield, also called the back scattered wavefield. All wavefield components are amplitudes normalized by the excitation wave amplitude.

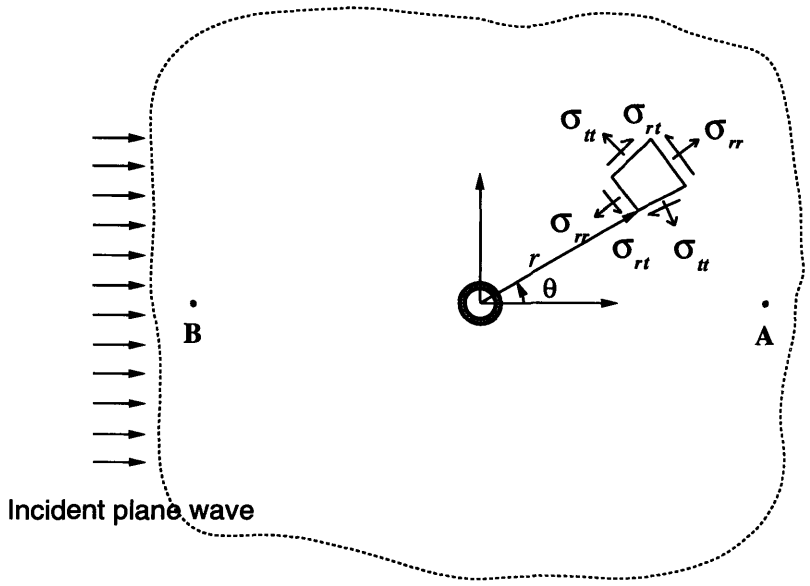

Figure 2-2: Wavefield stress components.

2.2 Data Generation

Data used for purposes of analysis were created according to the procedure outlined in Appendix A. This process was done once at the beginning of the work and subsequently the stored information was retrieved as required.

The actual data set consists of 18,480 *records*. Each record consists of a particular interphase state and the accompanying stress and displacement field. The interphase state in each record contains the following interphase physical and geometrical properties: thickness t, density ρ, Poisson's ratio ν, and modulus of elasticity E. The thickness

t is normalized by the radius of the fiber, a. The parameters ρ and E are normalized by the nominal matrix density value ρ_m and the nominal matrix modulus value E_m, respectively, found in Table 2.1. The wavefield in each record includes the following stress and displacement components: σ_{rr1}, σ_{tt1}, u_{r1}, σ_{rr2}, σ_{tt2}, u_{r2}, σ_{rt1}, u_{t1}, σ_{rt2}, u_{t2}.

The data set is configured in the following form:

$$t,\ \rho,\ \nu,\ E,\ \sigma_{rr1},\ \sigma_{tt1},\ u_{r1},\ \sigma_{rr2},\ \sigma_{tt2},\ u_{r2},\ \sigma_{rt1},\ u_{t1},\ \sigma_{rt2},\ u_{t2}$$

A small sample of the data set can be found in Appendix E.

2.3 Approach

2.3.1 Multivariate Correlation

In order to determine whether there is a linear relationship between two scalar variables, *simple* or *bivariate correlation* is used. The two variables under consideration may or may not be simply related or may not even possess the same units. Correlation analysis merely intends to describe the change of one variable with respect to the variation of another variable, that is the degree to which the quantities are linearly related. It is concerned with measuring the *relationship* or *strength of association* among variables [31].

Correlation provides a method to determine the independence of the variables in a data set, and here is utilized to show whether a meaningful relationship exists between certain interphase parameters and output wavefield components. Also, correlation analysis allows for the simplification of the subsequent regression analyses by reducing the number of necessary known variables (*regressors*) [32]. The measure of correlation between two

variables, say, x and y, is called the *correlation coefficient* r, defined as [14]

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.1}$$

where x_i and y_i are the ith observations of the variables x and y in the data set, respectively, and $\bar{x}$ and $\bar{y}$ are the average values of the variables x and y, respectively.

The correlation coefficient can range from –1 to +1. Thus, a perfect positive correlation between two variables is equivalent to a correlation coefficient of +1, while a perfect negative correlation is equivalent to a value of –1. A value of zero indicates no correlation between two variables.

Multivariate correlation analysis, as the name implies, is concerned with the correlations that exist among several variables and is the general case of bivariate correlation. The only requirement is that all observations of data are made with respect to each and every variable under consideration. A *correlation matrix* is used as a convenient representation of the correlation coefficients among the variables.

For the problem at hand, all correlations are to be found between t, ρ, ν, E, σ_{rr1}, σ_{tt1}, u_{r1}, σ_{rr2}, σ_{tt2}, u_{r2}, σ_{rt1}, u_{t1}, σ_{rt2}, u_{t2}. The correlation analysis can identify relationships in the interphase inverse problem and thereby can allow improved regression models for interphase thickness. The mathematical details of correlation analysis are found in Appendix B.

2.3.2 Multiple Regression

The principal means for study in this research is regression. This is a method of fitting a particular function to a set of (x,y) points in order to use the function for subsequent interpolation. Specifically, the function is generated by a least squares minimization process, such that the sum of the squares of the differences between the actual data points y_i and the predicted values $\hat{y}_i$ is minimized ($y_i - \hat{y}_i$ is called the *residual* at a

point). The function will depend on the data set considered, but is not sensitive to small changes in the contents of the data set. This fact makes regression techniques ideally suited for scientific prediction. Multiple regression extends this concept to data sets of more than two variables. The model for multiple regression has the form [9]

$$\hat{y} = \alpha_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k \tag{2.2}$$

where $\hat{y}$ is the predicted value of the variable of interest, α_o is the intercept or constant in the equation (which is often called *bias*), $\beta_{j,(j=1,2,3,...,k)}$ are the coefficients corresponding to the variables x_j, and $x_{j,(j=1,2,3,...,k)}$ are *predictor variables* or *regressors*, namely, some parameters which can be measured, known or estimated. The x_j can be transforms of a variable, such as x^2, $\sqrt{x}$, or $3x + 4$. The multiple regression procedure is explained in detail in Appendix C.

It has been determined that a multiple regression technique can provide a good model of the interphase inverse problem solution surface. For the sake of study, the interphase thickness t is considered as the variable y to be predicted (the *criterion variable*). The output stress components σ_{rr1}, σ_{rr2}, σ_{rt1}, σ_{rt2} are the bases for the predictor variables.

Regression models are developed here for the data set and subsets thereof, to show that subsidiary information about the interphase can contribute to improved interphase thickness prediction. Subsets are created by considering certain interphase parameters fixed, as depicted schematically in Fig. 2-3. For example, in case **A**, a subset is considered which has ρ constant at some known value ρ_1, and all other parameters are free to vary. Or, in case **E**, ρ, ν, and E are fixed at some known values ρ_1, ν_1, and E_1, respectively, so t is the only interphase parameter permitted to vary.

The criterion variable for all analyses is t and the predictor variables (or regressors) for all analyses are the output wavefield components, or a transform thereof. The linear multiple regression analysis utilizes σ_{rr1}, σ_{rr2}, σ_{rt1}, σ_{rt2} as regressors x_1, x_2, x_3, x_4. The nonlinear multiple regression analysis includes all regressors from the linear analysis, and also considers σ^2_{rr2}, σ^2_{rt2} as regressors x_5, x_6. These models simulate the potential

problem of an NDE researcher who seeks to learn about a specific interphase while being able to measure only these output wavefield components.

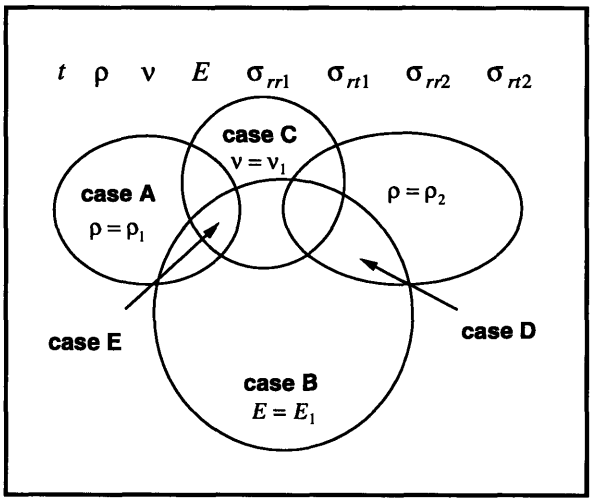

Figure 2-3: Data set showing possible subset cases for evaluation.

The procedure to form a model for the interphase thickness as a function of the output wavefield and to develop the necessary statistical parameters for unbiased comparison is as follows. Stepwise regression (Appendix D) is used to optimize the predictive capability of the model for any particular case. Next, conventional multiple regression implements that model and the relevant parameters needed for unbiased comparison between models are created. Also, the predictive capability of every model must be described. Two statistical quantities are considered for this purpose: coefficient of multiple determination R^2 and the prediction interval on the predicted values.

Coefficient of Determination R^2

The *coefficient of multiple determination* is defined as [32]

$$R^2 = \frac{SSR}{S_{yy}} \tag{2.3}$$

where SSR is equal to $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ and S_{yy} is equal to $\sum_{i=1}^{n}(y_i - \bar{y})^2$. SSR is known as the *sum of squares due to regression*, while S_{yy} is also called the *total sum of squares*.

Thus, R^2 is a ratio of the variability in y explained by the regression model to the total variability in criterion variable y. For instance, a model which explains y totally has R^2 equal to 1, while a model with no ability to explain y has a value of zero.

Prediction Intervals

Multiple regression yields coefficients which are used to create a model of a particular data set. The regression equation will then output one value for any given combination of input predictor values. For predicting the single observed value y_o, the best estimate is the point on the regression line at x_o, $\hat{y}_o = \alpha_o + \sum_{j=1}^{k} \beta_j (x_o)_j$.

Since any solution surface is likely to include fluctuation about the actual data points, analytic techniques used for prediction often yield an interval estimate for the solution. This is true in regression, and the interval is the *prediction interval.* The prediction interval for a value predicted by regression is defined as [9]

$$PI_{1-\alpha} = \hat{y} \pm t_{\frac{\alpha}{2},n-k-1} \sqrt{s_{y,x}^2 (1 + \frac{1}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} p_{ij} (x_i^* - \bar{x}_i)(x_j^* - \bar{x}_j))} \qquad (2.4)$$

where $(1 - \alpha)$ is the percent confidence desired, $t_{\frac{\alpha}{2},n-k-1}$ is the *student's t* statistic [9] which in the limit as $n \rightarrow \infty$ approaches the *normal* z statistic [9], x_i^* is the specific known value of predictor x_i, and p_{ij} are the elements of the solution coefficient matrix from the minimization of residuals [9]. $s_{y,x}^2$ is the variance about the trend line, and is computed as [9]

$$s_{y,x}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - k - 1} \qquad (2.5)$$

Concisely stated, the prediction interval is that region of the criterion variable range within which future observations of new data are found, with a specified confidence. It is an expression which considers the variability of the data, the regression performed, the number of observations in the data set and the values of the predictor variables. Details of both R^2 and prediction intervals are available in Appendix C.

Chapter 3

Results of Statistical Analyses

3.1 Correlation Analysis Results

As a preliminary step for understanding trends which exist in the problem, multivariate correlation analysis is implemented (see Appendix B). Correlation analysis enables one to ascertain relationships between any two quantities of interest. A correlation coefficient is found for all combinations of variables in the data set.

The correlation matrix in Table 3.1 contains the results from the multivariate correlation analysis. The correlation between any two quantities is found by the intersection of the corresponding column and row. For example, the correlation found in Table 3.1 between t and $\dot{\sigma}_{rr2}$ is -0.781. All correlation analyses are performed with SAS^{tm}, the Statistical Analysis System software package [33].

As a check on the values in Table 3.1, it can be seen that there are correlations of zero between the four interphase parameters t, ρ, ν, E; this is expected since the interphase parameters are specified independently, and there should not be any linear relationship between them. Also, all diagonal elements have a value of one. Initially, it is beneficial to reduce the number of variables in the set of output which have to be measured. This simplifies both the procedure for gathering data as well as the complexity of the model. We look to the correlation table for insight into similarities among the variables.

Immediately striking is the perfect correlation between σ_{rr1} and u_{r1}, and the almost

Table 3.1: Correlation Table.

	t	ρ	ν	E	σ_{rr1}	σ_{tt1}	u_{r1}	σ_{rr2}	σ_{tt2}	u_{r2}	σ_{rt1}	u_{t1}	σ_{rt2}	u_{t2}
t	1.000	0.000	0.000	0.000	0.289	0.291	0.289	-0.781	-0.781	-0.781	0.061	0.063	-0.792	-0.792
ρ		1.000	0.000	0.000	0.560	0.557	0.560	-0.373	-0.373	-0.373	0.440	0.440	-0.227	-0.227
ν			1.000	0.000	-0.239	-0.238	-0.239	0.083	0.083	0.083	0.097	0.097	-0.057	-0.057
E				1.000	-0.384	-0.385	-0.384	0.067	0.067	0.067	-0.502	-0.501	0.259	0.259
σ_{rr1}					1.000	0.999	1.000	-0.593	-0.593	-0.593	0.553	0.553	-0.487	-0.487
σ_{tt1}						1.000	0.999	-0.593	-0.592	-0.592	0.555	0.556	-0.493	-0.493
u_{r1}							1.000	-0.593	-0.593	-0.593	0.553	0.553	-0.488	-0.488
σ_{rr2}								1.000	1.000	1.000	-0.262	-0.264	0.769	0.768
σ_{tt2}									1.000	1.000	-0.262	-0.264	0.768	0.768
u_{r2}										1.000	-0.262	-0.264	0.769	0.768
σ_{rt1}											1.000	1.000	-0.435	-0.434
u_{t1}												1.000	0.436	-0.436
σ_{rt2}													1.000	1.000
u_{t2}														1.000

perfect correlation between σ_{rr1} and σ_{tt1}. This observation indicates that only one of the variables σ_{rr1}, σ_{tt1} and u_{r1} needs to be retained to contain the unique information present in the group. Since there is a perfect correlation between σ_{rr1} and u_{r1}, it is most wise to retain one of these two variables. Also, the most readily available experimental apparati, such as piezoelectric transducers, measure stresses at the surface of a material; therefore, σ_{rr1} is the single best choice to represent this group of three highly correlated quantities, σ_{rr1}, σ_{tt1} and u_{r1}.

Analogous to the above case is that of the scattered components σ_{rr2}, σ_{tt2}, and u_{r2}. The three are perfectly correlated to one another and hence all three need not be kept for analysis. The scattered component of the radial stress (σ_{rr2}) is retained using the same logic as before.

Other variables in Table 3.1 which are highly correlated are the stress and displacement components resulting from shear wave excitation, σ_{rt1}, u_{t1}, σ_{rt2}, u_{t2}. The transmitted components σ_{rt1} and u_{t1} have a correlation coefficient of one as do the scattered components σ_{rt2} and u_{t2}. For these two pair the stress components are kept and the displacement components are disregarded since all the information they contain as a set is already represented by the stress components.

As a prefix to the regression analysis to be explained later, the correlation relationships between the interphase quantities t, ρ, ν, E and σ_{rr1}, σ_{rr2}, σ_{rt1}, σ_{rt2} are studied. Ideally, a perfect correlation between the interphase parameters and the output stresses would provide knowledge of the interphase parameters simply by measuring the output stresses; however, the potential nonuniqueness of this NDE problem suggests that perfect correlation will not be the case. Upon examination of the values in the Table 3.1, one can see that there are no perfect correlations, but some strong negative correlations between t and σ_{rr2} (-0.781), t and σ_{rt2} (-0.792); and fair positive correlation between ρ and σ_{rr1}, ρ and σ_{rt1} (0.560 and 0.440 respectively). Modulus of elasticity E shows a similar degree of correlation as ρ, having fair negative correlation to both σ_{rr1} and σ_{rt1} (-0.384 and -0.502). Conversely, there is no output stress component considered which correlates well with ν. Thus, these correlation coefficients reinforce the concept from earlier work

Table 3.1: Correlations Between Interphase Parameters and Output Stresses.

	t	ρ	ν	E
σ_{rr1}	0.289	0.560	-0.239	-0.384
σ_{rr2}	-0.781	-0.373	0.083	0.067
σ_{rt1}	0.061	0.440	0.097	-0.502
σ_{rt2}	-0.792	-0.227	-0.057	0.259

[21] that there are relationships between the interphase parameters and the components of the output wavefield, most notable of which are those which include the interphase thickness t.

From this information it is evident that interphase thickness shows the most promise for a regression model using the output stresses as regressors. It is important to note that this analysis has enabled the number of required measurable variables to be reduced from ten to four. The complexity of the model is therefore reduced quite significantly, allowing for a more efficient model, requiring less computational time. Also, utilizing an efficient method allows the scope of the analysis to be increased if the need arises. Table 3.1 summarizes the significant correlation relationships between the interphase parameters and the output stresses.

3.2 Regression Analysis Results

3.2.1 Case of Unknown Interphase State

A linear multiple regression model is first considered for predicting the interphase thickness t. This is to say that all predictors x_j are linear as are all parameter coefficients β_j. The predictors considered for inclusion in the model by the stepwise analysis are σ_{rr1}, σ_{rr2}, σ_{rt1}, σ_{rt2}. The criterion variable is t. This and all subsequent regression analyses are performed with SAS™.

This first model produces quite successful results, explaining 79.1% of the variability in the criterion variable t, shown in Table 3.2. This model is a linear, multiple regression model, the first model to be considered in any regression analysis. No knowledge of

interphase parameters is assumed. Evident in this result is the good correlation (see Table 3.1) between the two scattered components σ_{rr2} and σ_{rt2} and the criterion variable t.

In this work, the adjusted coefficient of multiple determination R^2_{adj} (eqn. C.5) is not utilized for analysis. Due to the similar number of regressors in all models considered, the values of R^2_{adj} do not deviate significantly from the unadjusted R^2 values. Also, Mallows' statistic C_p (eqn. C.6) remains within a narrow range, and it does not indicate a sizable difference in the models created. These facts suggest that all models in this study are of equivalent worth.

Nonlinear multiple regression models are formed from eqn. 2.2 by considering nonlinear transforms of output wavefield components as predictors x_j. In addition to those predictors considered in the linear analysis, σ^2_{rr2} and σ^2_{rt2} are considered for inclusion in the nonlinear model. The nonlinear model of interphase thickness increases the determination coefficient to a very satisfactory value of 88.7%. An R^2 value such as this indicates a reliable fit with which the interphase thickness can be predicted, with acceptable precision.

3.2.2 Cases with Interphase Information Known

Of interest is the issue of how prediction ability improves if the unknown material state can be characterized before analysis. For example, an unknown interphase can be characterized by identifying the value of one of its material properties, such as ρ. Fig. 3-1 illustrates the data set, and possible subsets defined by fixing a parameter or several parameters of the interphase material state. For any subset case, regression of t onto the output wavefield components produces a model for that case represented by a function $f_{j,(j=1,2,3,...)}$ with prediction intervals represented by $\delta_{j,(j=1,2,3,...)}$. Comparisons among these models permit the study of how interphase parameters can affect the ability to predict t.

Analyses are completed for each of the discrete values in the domain of the interphase density. First, the data set is transformed into several smaller ones which contain three unknown variables, t, ν and E and one known value ρ. Regression is used on each of

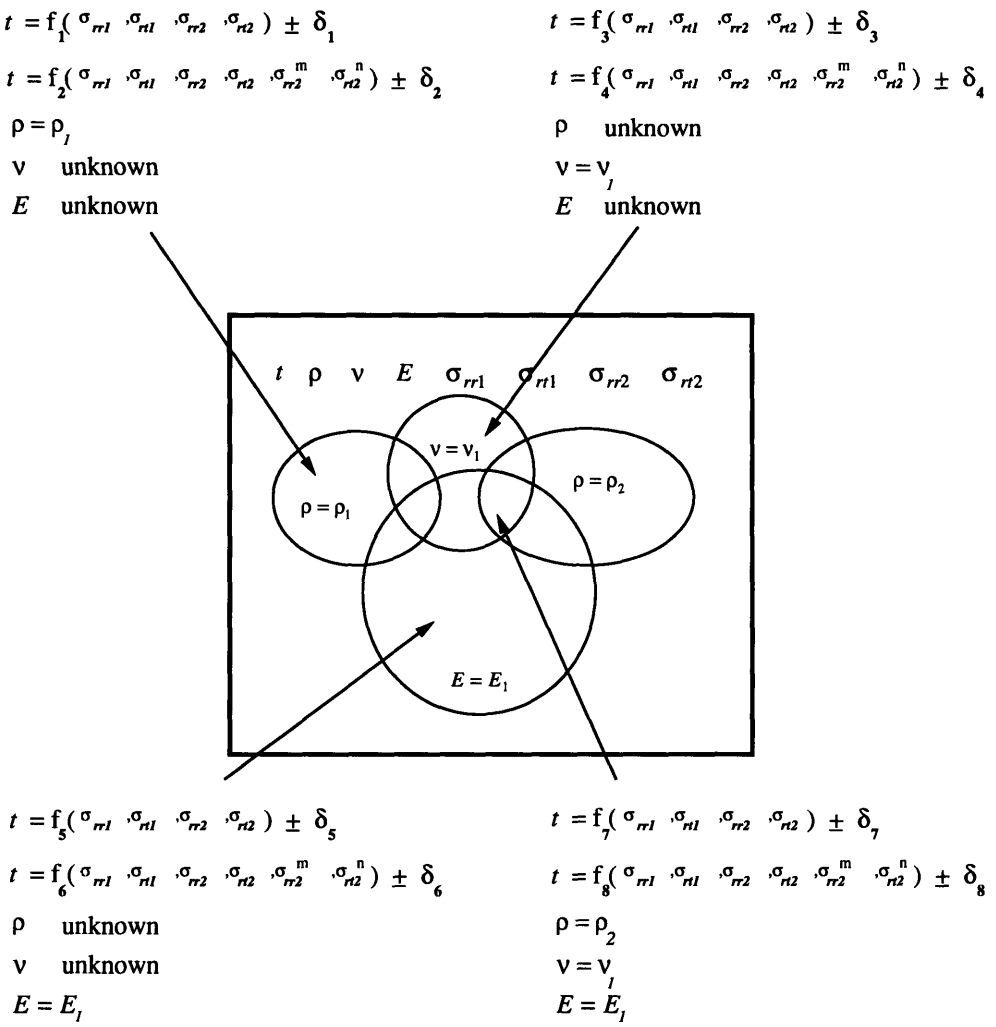

Figure 3-1: Subset cases with interphase information available.

these smaller data subsets to generate regression models for each. In doing so, the effect of ρ is removed from the analysis; the results prove to be very satisfactory. A situation such as this mimics the case where a researcher knows the value of interphase density, but does not know other properties. Table 3.2 contains the determination coefficients from the regression analysis cases for each particular ρ. The results are plotted in Fig. 3-2.

Considered next is the situation with elastic modulus known in the interphase region. For cases with the interphase modulus fixed and all other parameters allowed to vary, a summary of results is available in the middle portion of Table 3.2, shown graphically in Fig. 3-2.

Table 3.2: Determination Coefficients for Regression Models of Interphase Thickness t.

ρ	ν	E	R^2: linear	R^2: nonlinear
unknown	unknown	unknown	0.791	0.887
1.2115	unknown	unknown	0.962	0.983
1.4538	unknown	unknown	0.966	0.990
1.6962	unknown	unknown	0.965	0.965
1.9385	unknown	unknown	0.972	0.985
2.1808	unknown	unknown	0.960	0.982
2.4231	unknown	unknown	0.950	0.980
2.6654	unknown	unknown	0.930	0.970
2.9077	unknown	unknown	0.880	0.950
3.1500	unknown	unknown	0.820	0.920
3.3923	unknown	unknown	0.740	0.870
3.6346	unknown	unknown	0.620	0.770
unknown	unknown	1.0284	0.600	0.770
unknown	unknown	1.1754	0.680	0.815
unknown	unknown	1.3223	0.840	0.890
unknown	unknown	1.4692	0.860	0.904
unknown	unknown	1.6161	0.830	0.908
unknown	unknown	1.7631	0.850	0.927
unknown	unknown	1.9100	0.867	0.941
unknown	unknown	2.0569	0.873	0.942
unknown	unknown	2.2038	0.880	0.940
unknown	unknown	2.3507	0.894	0.941
unknown	unknown	2.4977	0.910	0.946
unknown	unknown	2.6446	0.922	0.951
unknown	unknown	2.7915	0.929	0.953
unknown	unknown	2.9384	0.930	0.951
unknown	0.10	unknown	0.789	0.896
unknown	0.16	unknown	0.787	0.901
unknown	0.22	unknown	0.783	0.908
unknown	0.28	unknown	0.787	0.905
unknown	0.34	unknown	0.807	0.887
unknown	0.40	unknown	0.830	0.880

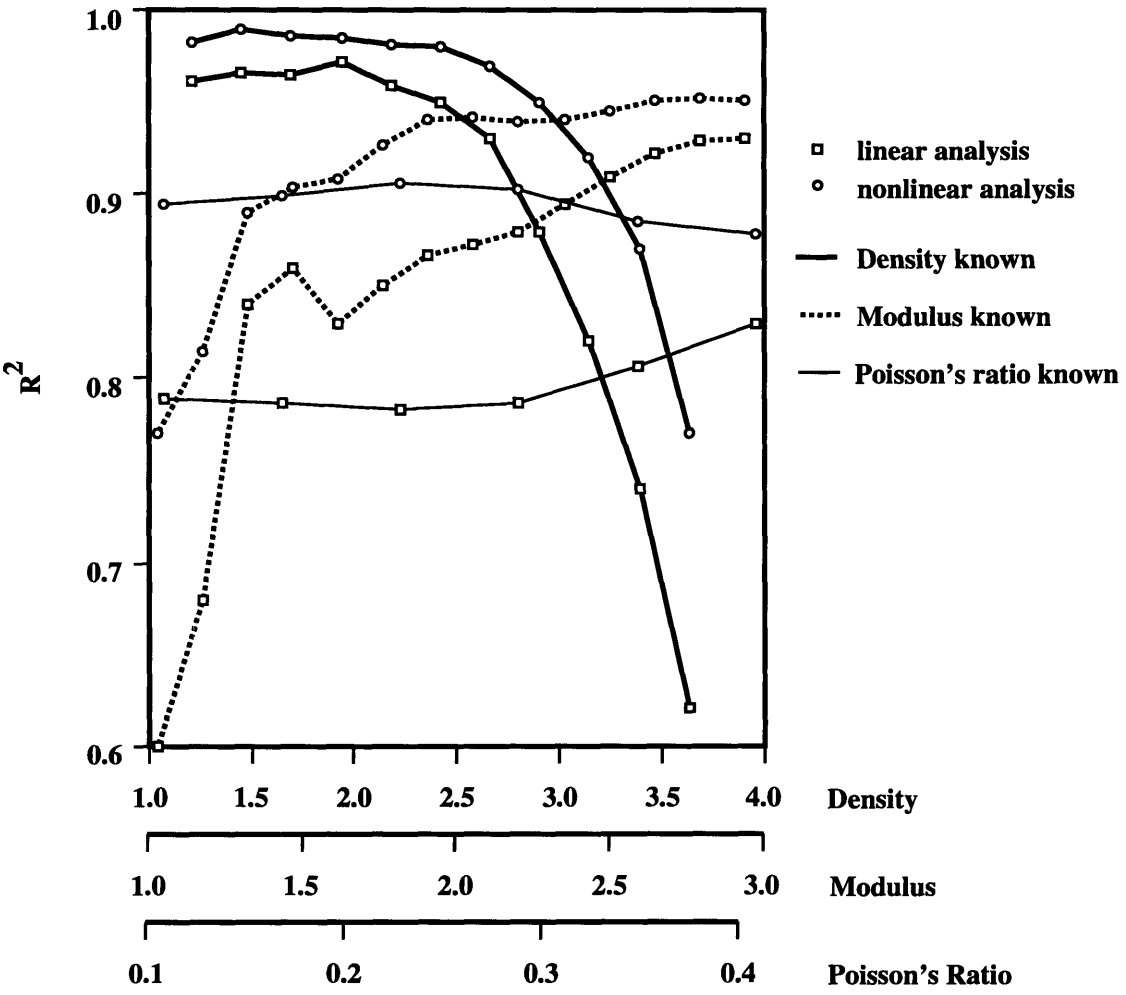

Figure 3-2: Determination coefficients for known interphase material state cases.

Interestingly, should a researcher have knowledge of the interphase Poisson's ratio value, no real benefit can be asserted for explaining interphase thickness. Table 3.2 summarizes the results for analysis cases with known interphase Poisson's ratio. The lack of a distinct trend in Fig. 3-2 also makes the absence of improved results apparent.

Of the most striking results are those cases in which the interphase material state is known (ρ, ν, E are known) and it is desired to predict interphase thickness. Although testing every possibility is impractical (and most probably implausible), many are analyzed. Table 3.3 lists the particular cases investigated, and the results of the analysis.

Table 3.3: Determination Coefficients for Known Interphase Material State Cases.

ρ	ν	E	R^2: linear	R^2: nonlinear
1.2115	0.28	1.0284	0.998	1.00
1.2115	0.40	1.1754	1.00	1.00
1.4538	0.28	2.6446	0.999	1.00
1.6932	0.16	1.4692	0.987	1.00
1.6932	0.16	2.7915	0.998	1.00
1.6962	0.34	1.1754	0.994	1.00
1.9385	0.40	1.3223	0.985	1.00
2.1808	0.28	2.0569	0.996	1.00
2.1808	0.34	1.0284	0.985	0.999
2.4231	0.40	1.9100	0.997	1.00
2.6654	0.10	1.1754	0.989	0.989
2.6654	0.28	1.3223	0.983	0.983
2.6654	0.40	1.1754	0.988	0.999
2.9077	0.34	2.4977	0.996	0.998
3.1500	0.34	1.9100	0.956	0.993
3.3923	0.16	2.9384	0.985	0.996
3.3923	0.28	2.0569	0.977	0.989
3.3923	0.40	1.6161	0.815	0.903
3.6346	0.10	1.4692	0.833	0.894
3.6346	0.16	1.7631	0.815	0.828

Table 3.4: Grouped R^2 Values for Known Interphase Material State Cases.

R^2	Number of Cases	
	linear analysis	nonlinear analysis
0.800 – 0.900	3	2
0.900 – 0.950	0	1
0.950 – 0.980	2	0
0.980 – 0.990	7	3
0.990 – 1.000	8	14

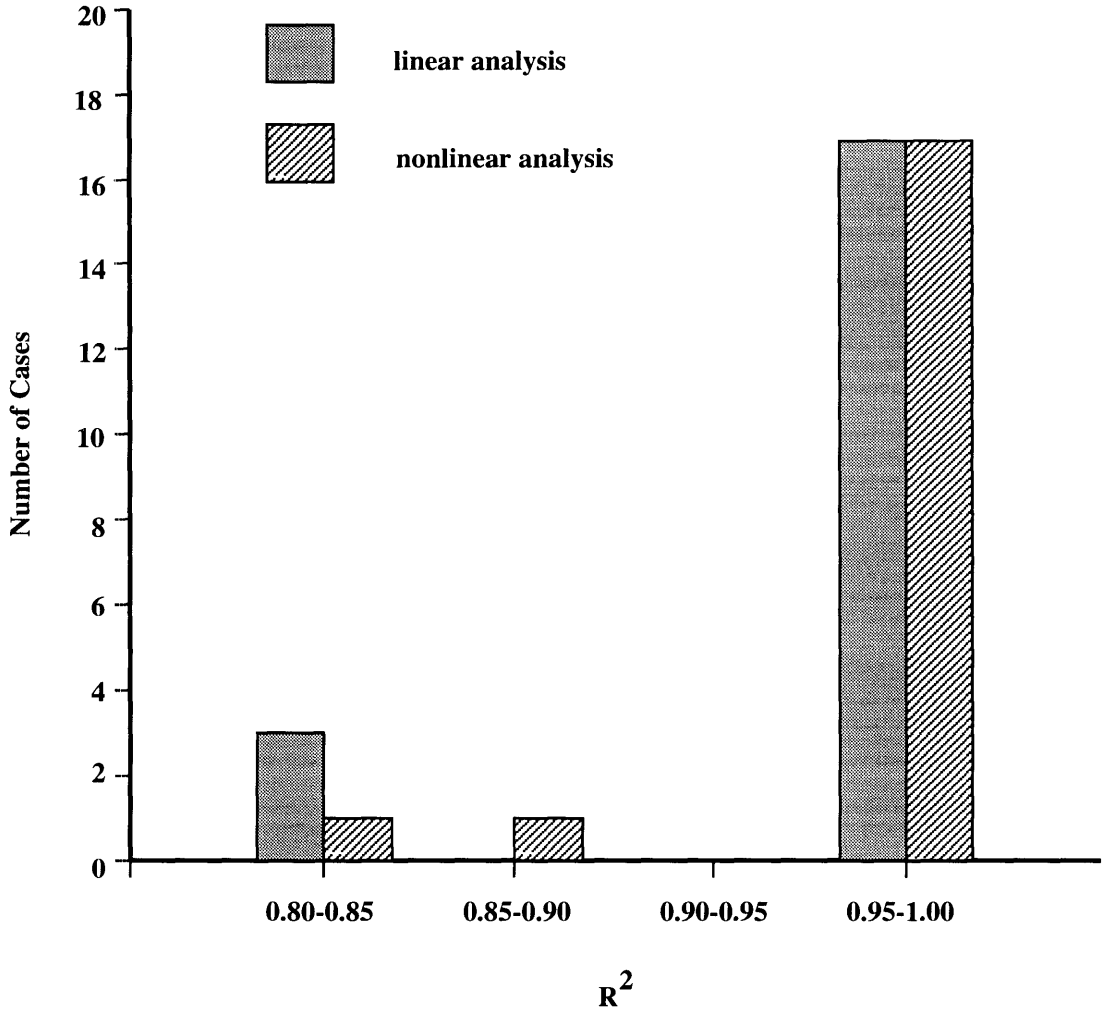

Figure 3-3: Summary plot of cases with known interphase material state.

Table 3.4 summarizes the coefficients of multiple determination and their frequency from the cases of known material state. This table is organized in the following fashion: for a total of twenty cases considered, the number appearing in the second column is the number of cases which have the coefficient of multiple determination listed in the same row as the value. For instance, of the twenty linear analyses which have been completed, three have an R^2 value between 80% and 90%, and eight have values from 99% to 100%. This same information is plotted in Fig. 3-3.

3.2.3 Evaluation of Estimated Values: Prediction Intervals

To evaluate the precision of the interphase thickness estimates, prediction intervals are required. The size of the prediction interval is used to describe the prediction ability for comparison purposes. A smaller prediction interval implies that the value is more tightly bounded and more useful for competently describing the interphase thickness of the fiber composite. The prediction interval is specified as a range around the predicted value. A 95% confidence level is utilized to determine all prediction intervals. Table 3.5 includes prediction intervals for selected regression cases.

The size of the prediction interval from the linear analysis of the original data set is ± 0.0255. Implementing a nonlinear analysis on the original data set enables the size of prediction interval to be reduced by 20%, to ± 0.0190. This is not a trivial result and instead implies that the resolution increases by using a nonlinear analysis. The trend continues in further analyses.

Cases with known interphase density have resulted in increased determination coefficients for the regression models of t. Three such cases of known interphase density are tabulated in Table 3.5. The first tabulated value of $\rho = 1.4538$ corresponds to a very successful regression case, the second value of $\rho = 2.6654$ is a good analysis case and the last value of $\rho = 3.3923$ is a case unimproved from the case with no interphase information known, as evidenced by its R^2 value.

A known modulus also is able to permit improved interphase thickness prediction. This situation is summarized for three cases of known interphase modulus in Table 3.5.

Table 3.5: Prediction Interval Size for Regression Models of Interphase Thickness t.

			linear analysis		nonlinear analysis	
ρ	ν	E	R^2	interval	R^2	interval
unknown	unknown	unknown	0.791	±0.0255	0.887	±0.0190
1.4538	unknown	unknown	0.966	±0.0105	0.990	±0.0056
2.6654	unknown	unknown	0.930	±0.0153	0.970	±0.0107
3.3923	unknown	unknown	0.740	±0.0294	0.870	±0.0209
unknown	unknown	1.0284	0.600	±0.0363	0.770	±0.0276
unknown	unknown	1.9100	0.867	±0.0209	0.941	±0.0140
unknown	unknown	2.9384	0.930	±0.0151	0.951	±0.0128
1.2115	0.28	1.0284	0.998	±0.0001	1.00	±6.57e-5
2.6654	0.10	1.1754	0.989	±0.0083	0.989	±0.0035
3.3923	0.16	2.9384	0.985	±0.0103	0.996	±0.0056

The prediction intervals behave in the same manner as those of the known interphase density case, with respect to the coefficient of multiple determination: low values of R^2 correspond to large intervals, and vice-versa. Modulus value cases tabulated are examples of poor, good, and excellent predictive results.

The remainder of Table 3.5 includes cases of known material state. The intervals from these cases indicate excellent results, as expected from their high coefficients of multiple determination. Such cases emulate the idealized situation wherein an investigator is certain of the material parameters of the interphase under consideration, and only needs an estimate of interphase thickness.

3.3 Discussion

For most analysis cases of known interphase density, nonlinear regression analyses provide determination coefficient values which are well into the ninety percent range, as are over half of the R^2 values from linear analyses. Low R^2 values from such analyses appear to surface when the interphase density value is known, and is high in the range considered. Accordingly, a loss of the ability to explain the variation of interphase thickness accompanies these particular cases. Utilizing a nonlinear model permits good results for all cases but two, that is eighty percent of the cases here. However, if we look

to the physical meaning of these ρ values, we can determine more.

The cases with known interphase density values providing less than desirable results (that is, $R^2 < 0.90$) are the cases corresponding to the four highest density values in the range considered. Even if nonlinear analysis is utilized, the cases corresponding to the two highest values in the interphase density domain are those which yield poor results. However, these interphase density values are quite unlikely to be encountered in actual practice. The lowest value of concern is $\rho = 2.9077$, which corresponds to an actual interphase density value of 7.56 g/cm^3. In comparison, recall the density of alumina is 3.7 g/cm^3 and that of zirconia is 6.3 g/cm^3, each less than the value for interphase density assumed in this case. Similarly, for the remaining three questionable cases the values used for the known interphase density are 8.19, 8.82, and 9.45 g/cm^3, respectively. These larger density values are even more unlikely to be encountered in the actual composite, by the same reasoning.

If this is indeed the case, then those particular values of interphase density can be omitted from the domain as extremely unlikely to be encountered. Hence, what remains are very favorable results for the analysis cases in which interphase density is known.

A known interphase modulus of elasticity is also successfully utilized to improve the models for predicting interphase thickness. It is seen from Table 3.2 that for this situation, high modulus values provide the best basis for interphase thickness prediction. In fact, if 90% is demanded as the minimum acceptable coefficient of multiple determination, only the cases of the three lowest modulus values are unacceptable. This leaves almost eighty percent of the interphase elastic modulus domain which is very accurately modeled with regression.

Linear models of the data subsets with known ν do not appear to significantly improve the results in comparison to the initial linear model of the entire data set. Recall the determination coefficient for the original instance was 79.1%. One can see from Table 3.2 that having knowledge of the interphase Poisson's ratio does not permit much improved R^2 for the interphase thickness regression model. Analysis of the case corresponding to the last value in the domain of Poisson's ratio yields a coefficient of determination value

of 83.0%, from Table 3.2. This may indicate that prediction ability is better for this one case, but that may not be a statistically significant difference. The effects of known interphase Poisson's ratio on the ability to predict interphase thickness appear unclear from these results.

Nonlinear models for cases of known interphase Poisson's ratio do not appear to ameliorate the situation. In comparison to the R^2 value of 88.7% from analysis of the case with no information known, the R^2 values ranging from 88.0% to 90.8% for known ν cases are not substantive enough to justify the added effort. It is clear from both the linear and nonlinear cases that the effect of known interphase Poisson's ratio on the ability to predict interphase thickness is insignificant, or perhaps random at best. This conclusion is also evidenced in Table 3.1 from the correlation relationships between ν and the output wavefield components. Correlations between ν and any component tend to be small, less than 23%, typically less than 10%. Therefore having knowledge of ν in this NDE problem does not help to narrow down the number of potential solution wavefields for a given interphase material state, and thus cannot contribute to improved estimation of interphase thickness t.

From the equations of motion and constitutive equations for wave mechanics which have been used in the solution to this problem [34, 35], one can see that Poisson's ratio appears in terms like $(1\text{-}\nu)^2$, $(1\text{-}\nu^2)$ or $(3\text{-}\nu)$. Typically, these become higher order terms in the dynamic analysis and hence as ν changes over the range considered in this study, there is not much effect on the output wavefield, as corroborated from the correlation table. With this rationale in mind, the above results relating to ν are well founded.

Table 3.3 contains the excellent results for regression cases of known material state. The lowest R^2 value encountered during the analysis was actually 81.5%, for particular material states to be described. Next, one might notice the predominance of virtually perfect determination coefficients — especially for the nonlinear cases considered. The implication here is that if one has a known interphase material state, such as in the example of a fiber sizing, a value for interphase thickness is available with exceptional accuracy. This *same* task for the determination of fiber sizings has been accomplished in

the past through typically destructive methods (e.g. thermo-gravimetric analysis (TGA) [36] or metallographic evaluation), and has been of interest for researchers across several disciplines.

Also, the poorest results from the analysis cases of known interphase material state occur for particular interphase material states wherein the interphase density is high. Note that is also the situation for the cases of known interphase density, summarized in Table 3.2. Similarly, if the cases of known interphase material state with high interphase density can be eliminated by the logic utilized before, there is no exception to the capability of the regression methods which are provided in cases of known interphase material state.

Known interphase information allows improved thickness estimation. In fact, when any outside information is available to the analysis, its inclusion can only improve the results unless the outside information is insignificant to the trends of the problem. Consider, for example, the case of a known interphase density value; the regression model makes no provision for including this known information as a regressor. However, it is by reducing the number of degrees of freedom in the data set that the results have changed, because some possible wavefields are eliminated as solution possibilities for the particular model under consideration. That is, the dimensionality of the output wavefield data is reduced with the fixed interphase parameter by considering a specific subset of entire data set. Consider Fig. 3-4, which depicts a scheme for the interphase inverse problem. The domain of the forward problem can map onto the solution space as indicated by the dotted arrows. The inverse problem is to draw conclusions about the **interphase domain** based on the **output wavefield**, as represented by bold arrows.

The interphase inverse problem with known interphase density is of smaller dimension than the interphase inverse problem with no known interphase information. The additional information of known interphase density allows the solution space to be reduced, thereby removing the effects of interphase density from those which are to be modeled with regression. Hence, there is less variation in the stress solution space which is to be accounted for by the regression model. Similar logic applies to the cases of known

elastic modulus, cases of known Poisson's ratio, and the cases of known material state. In the cases of known Poisson's ratio, the solution space is not reduced whereas in the cases of known material state, the solution space is reduced quite considerably.

The primary concern is to demonstrate the relationship between the output stress wavefield components and the properties of the interphase. It is entirely possible that there is a need for increased resolution to study effects which are unexplainable at this time, such as the effect interphase Poisson's ratio has on the wavefield. However it has been shown that there are effects which are readily resolved utilizing information about the problem. Preprocessing in other forms such as principal component analysis would likely have a positive effect also.

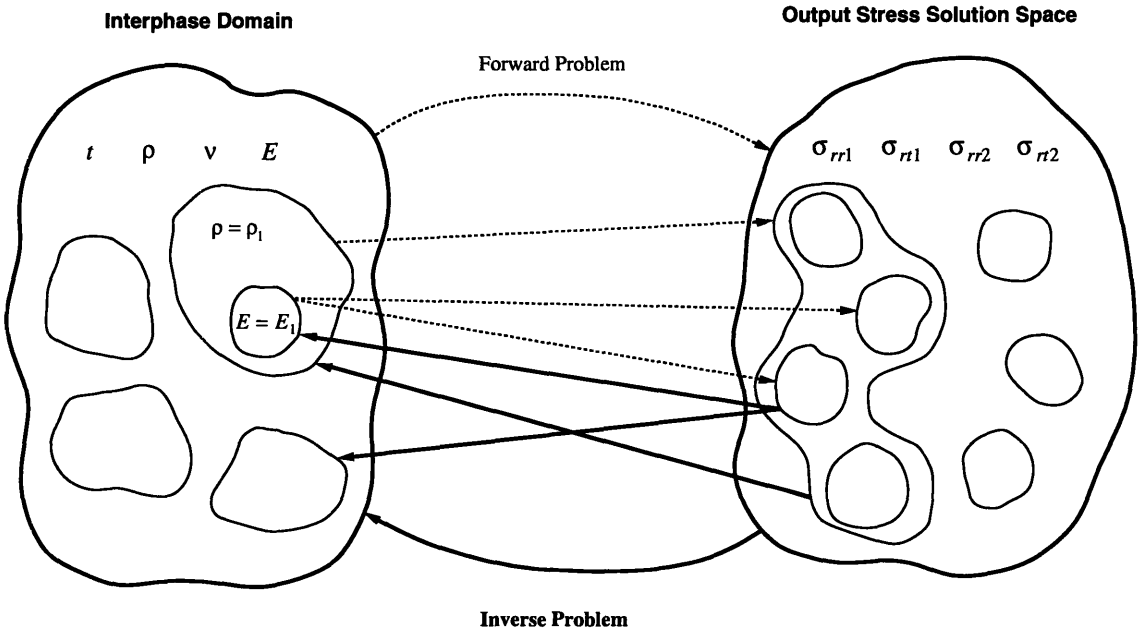

Figure 3-4: Solution space.

3.4 Error Interpretation

One can consider the regression results in terms of error, and how this error was affected. There are two types of error involved in the analysis. One source of error is inherent in the procedure of prediction intervals, and is called a *Type I* error. A Type I error is the error of misclassification, that is, the percentage of predicted values for new data which really do *not* fall within the prediction interval. This is the error evidenced in the significance level which is fixed at 5% (1.00 - 0.95), and its only recourse is to increase the size of the prediction intervals thereby reducing the effectiveness of the algorithm. This error can not ever be eliminated completely, but it is small, and not cause for concern.

By assuming knowledge of interphase parameters, the ability to decrease the size of the prediction interval is demonstrated. However, the size of the prediction interval in itself *is* error. Ergo, one can only assume the entire prediction interval to be the error for a particular point estimate of interphase thickness, to provide the most conservative estimate for a range on the predicted interphase thickness value.

Thus there are two distinct types of estimation error to be concerned about here: the Type I error, and the error due to the spread of the data. The former can be controlled, at the expense of affecting the certainty of an estimate. The latter can be decreased by utilizing information which is pertinent to the trends in the problem, in order to improve the fit of the model to the data. This is the method which is utilized to improve estimation in this work.

Chapter 4

Future Work

4.1 Introduction to Artificial Intelligence

The field of NDE is beginning to acknowledge the need for automation of its processes. The primary reason for this need is to keep nondestructive evaluation techniques competitive with and cost-effective in comparison to other testing techniques, especially in the manufacturing industry. Interest in artificial intelligence (AI) has grown in the context of NDE because of its ability to be applied to general problems, and the great potential it has for solving problems usually too complex for a machine to model. Also, AI can automate much of the work which is presently done with human skill, and thereby can improve speed and reduce error.

Artificial intelligence research is quite active, and there is much room for growth in this technology. Accordingly, NDE has not yet gleaned the benefits from this research work. Many prospective applications have been published, and a sample is discussed below. This is certainly an indication that the field of NDE will gain significant contributions from AI, as they are developed.

4.2 Overview of AI Terminology

Artificial intelligence is a general term used to describe all processes of integrating

human knowledge into computer systems. An equivalent term for this is *knowledge engineering*. A *knowledge based* system is one type of artificial intelligence system which contains some kind of *knowledge base* and an *inference mechanism*. The distinction for these systems is that they are a set of state-driven processes rather than hard-coded ones. *State-driven processes* are those in which decisions about how to process data are part of the knowledge of the system. Many variants of AI technology have developed in recent years, and each of the major types will be discussed below.

Expert Systems have proven to be the most versatile of all knowledge based systems. An expert system is an information system that can pose and answer questions relating to information borrowed from human experts. This is accomplished by a software system that mimics the deductive and inductive reasoning of experts. Since its knowledge base comes from human experts, its facts and rules are not totally certain nor consistent. An inference procedure permits automatic extraction of pertinent rules from the description of the data provided to it. Any method of drawing conclusions from facts which seem to be correct is called *plausible reasoning*. Plausible reasoning is usually heuristic, but can also be mathematically sound. Expert systems presently find use in capital intensive areas, since in such areas it is easy to recapture the initial high capital expenditures required for startup. Examples include oil exploration and drilling. Detailed explanation is available in references [37, 38].

As already explained, expert systems require a collection of facts and rules and a way of making inferences from those facts and rules. A *rule* is a pattern-invoked program, activated whenever certain conditions hold in the data. These examine the current state of facts and in turn can change the current state of facts. A *production system* is made up of a set of rules, also called *production rules*, which operate in cycles according to the input to the system. In each cycle, the conditions of each production rule are matched against the current state of facts. One benefit of this architecture is modularity. This allows decisions to be made without complete information, unlike algorithmic code solutions.

For the user, this translates into the effect of helping to focus more on the problem itself than on the implementation of the underlying software. This is possible because

expert systems are *self-regulated systems*, intelligent systems free to act according to internalized goals. A movement governed by an internalized goal (such as retrieving the current structure of facts) is called *pullback*. It is contrasted with feedback wherein a system is externally regulated with respect to a margin of error in reference to an external goal.

Fuzzy logic is the system of manipulation for *fuzzy sets*. To say a word is "fuzzy" is to say that sometimes there is no definite answer as to whether the word applies to something. Note that this is due to an aspect of the meaning of the word, not due to the **state** of our knowledge. Fuzzy sets in contrast to crisp sets can be thought to include the entire universe as members. However, characteristic of fuzzy sets is that each element has a degree of membership value, which is in the range of zero to one. As such, since many members will have a degree of membership of zero, the set of all elements which remain with nonzero membership are often referred to as a *fuzzy subset*. Fuzzy subsets are very valuable when trying to encode natural spoken language into a system, because of the open textured nature of language i.e. the possibility of omitting some unintended meaning for a word. Thus it is ideal for expert knowledge which is often imprecise and intricate [39].

Generating much excitement as of late are neural networks, also known as learning networks, parallel distributed processing, or auto-associative processing. *Artificial neural networks* (ANN) are different from knowledge based systems because there is no knowledge base nor inference mechanism. Instead, the function of the knowledge base is supplanted by *training data*, which the software uses to map the domain under consideration. Subsequent navigation of this domain may allow classification of an unknown and completely unique data point through complex fits. Many efforts have been directed into utilizing this technology for problem solving. They require no *a priori* information or built-in rules; rather, they acquire knowledge of a system through the presentation of examples.

4.3 Literature Review

Imaging has long been important in quality assessment, and is constantly being updated with new technology. In work by Papadakis [40, 41], a method is proposed to utilize quality-related imaging information without actually forming the image. This is done to prevent much secondary processing, which can be avoided with the use of computerized interpretation of the imaging data. As an added benefit, human interpretation of output is unnecessary, and hence can be removed from the loop. The proposed automated process for a manufacturing system would utilize a computer to gather the NDE output, control the imaging function, and create control charts. Artificial intelligence algorithms would then interpret the data and ideally offer correction to process flaws. This system would be flexible to other operations, and even offer time-sharing to balance several processes at once. Most significantly, the availability of expert systems and learning networks would reduce error and diagnostic time required once evaluation signals a faulty process.

A large fraction of all AI research has been directed towards the neural network branch of the science. Neural networks have proven to be effective for certain types of pattern recognition problems. An article by Hill, Israel, and Knotts [42] discusses the use of backpropagation neural networks to predict ultimate strengths in aluminum-lithium welds. Acoustic emission flaw growth activity was monitored in the weld specimens from the onset of tensile loading until failure. The data used to train the network, however, were only that from initial loading to 25 percent of the expected ultimate strength. The fully interconnected network had one hidden layer, and the worst case prediction was found to be 2.6 percent in error. It was shown previous to this study that AE data taken during proof loading were correlated with ultimate strengths in both composites and in metals. The advantage in this case was the ability of the neural network to provide an automated technique for sorting out the AE associated with the various mechanisms and determine the functional form of the desired relationships.

The inverse problem of eddy current data has been addressed with neural networks. Mann, Schmerr and Moulder [43] explored the application of neural networks for the

inversion of eddy current data in determining flaw sizes. Training experiments were run using data generated with theoretical solutions, as well as actual experimental data. Three parameters were sought, the depth, thickness and length of a surface flaw. Their conclusion is that the neural networks show great promise in being able to solve the problem, and more work is needed to refine the application. No effort in this work was aimed at optimizing the structure, however fair accuracy was obtained. This paper like that of Hill [42] is good evidence that neural networks are capable of modeling the inverse problem for particular NDE data, in this case eddy current data.

Udpa and Udpa [44] also concentrated on the classification of eddy current defect signals in their work. Their view was that pattern recognition algorithms could be utilized for signal interpretation, and hence the classification of a defect as a "solution" to the inverse problem. Interestingly they employed preprocessing of the data through Fourier descriptors in an effort to derive a parametric model. In turn, this data set was the one used for further classification needs. The results they present are quite successful, showing that the neural network was correctly able to classify the processed defect data into one of four possible defect categories. They compare the results to traditional classification techniques, e.g. K-means clustering.

Thomsen and Lund [45] looked into the ability of an ANN to model a nonlinear classification problem on ultrasonic power spectra. As in the work by Hill, Israel and Knotts [42], correlation of the ultrasonic stress wave factors with composite fabrication quality already existed, so the neural network was implemented to perform an accurate classification of the ultrasonic signals. The results reported classification success rates of 86 to 96 percent, depending on category. Neural networks were in this respect said to be superior to the conventional stress wave factor analysis.

Common to all authors was the acknowledgment that neural networks have major drawbacks. For classification problems, the most considerable drawback is the network's inadequacy in handling any data type which was not presented in the training data. So unless a particular feature is able to be comprised of other, recognizable features, the neural network will be unable to handle the disparity [45]. Other problems specific

to back-propagation methods (the type of neural network used in all the works cited) are its susceptibility to becoming "trapped" in local minima and thereby not accurately modeling the true global minima of a set [43]. Also, training can be quite lengthy, and care is needed to insure the network does not lose its ability to generalize [6, 43]. More fundamentally, the holistic nature of neural networks makes them difficult to understand logically, since neural processing is a complex solution not explained in terms of the problem of interest. Also, it is unreasonable to expect to be able to determine if the found solution is indeed the correct one [6].

Expert systems can circumvent these particular problems. Two examples of their application to NDE are found in the works of Chapman et al. [46] and Shankar, Williams, and Avioli, Jr. [47]. Chapman discusses the application of a commercial expert system software package to accomplish the pattern recognition and decision analysis for the inspection of cracks in an aircraft compressor disk. The results were compared with a human expert's interpretation of the eddy current signals, as well as the analysis provided by SEM. It was found that the system performed significantly better in detecting cracks than the experienced observer, and almost as well as the SEM analysis. An added benefit was the real-time classification of the signals, which is realistically limited in speed only by the acquisition equipment. Features of eddy current data are readily decomposed for crack detection analysis, and are well-suited to expert system methods.

Rule-based decision logic has already been implemented in ultrasonic weld examination of boiling water reactors [47]. The system described is quite powerful because it is configured in a question-and-answer format. The knowledge base consists of over 200 rules, confined to consultation on ultrasonic data and specimen information. These rules were developed to emulate operators while integrating data from testing. The inference engine of the system is of the backward-chaining type, i.e. it attempts to match known feature classes with available data. The research did not seek to demonstrate system performance per se, but instead only to evaluate functionality, accuracy of knowledge base, and direction for improvement. It has succeeded in these respects for discriminating weld IGSCC from benign weld features, in this limited application. The most useful result was

the creation of a valid decision logic to assist in the detection of cracks from conventional ultrasonic signal parameters, which could be utilized in other similar problems.

4.4 Proposed Work

Considering both the previous areas which have been researched as well as the predicted future of NDE, a proposal is presented for solving the inverse interphase property problem. The concept is to utilize a hybrid method involving several artificial intelligence techniques in concert with a reliable statistical descriptor. This permits several benefits over any of the methods alone.

Where statistics appear unable to handle the inverse interphase problem, artificial intelligence methods appear to be quite applicable. That is, statistics are generally unable to solve a pattern recognition problem, although they can reveal trends in some data and provide predictive capabilities. But here is where knowledge based systems can become quite useful; through progressive manipulation of the information in a problem, several systems combine effects to produce the best solution to the problem.

Expert systems while promising are yet undeveloped for NDE purposes. While this is expected to be alleviated as time passes, at the present it is an obstacle. Furthermore, these systems may only be as valuable as the individuals used to generate them; no new understanding is likely to come out of this technology. To their advantage, expert systems incorporate the speed, accuracy and efficiency of computers while still incorporating insight supplied by human expertise and experience.

The proposal is depicted in Fig. 4-1 and functions as follows. A sample "signature" is created for the material via conventional acousto-ultrasonic techniques. This signature is a discretization of output characteristics provided by the processing of the ultrasonic signal in the acquisition equipment. A **pattern recognition algorithm** then attempts to classify the signature into any of several classes predetermined to the procedure. Upon successful identification, this produces information which is used by the **expert system** to specify which statistical models are to be implemented on the signature. The **statistical**

analyzer then computes the desired interphase properties, based on the models identified for consideration.

A **second expert system** is utilized to process the output from the statistical analyzer, and sort the possible interphase states of which the unknown could be. This second expert system assigns probabilities to each solution based on information provided by the statistical analysis and the first expert system. It outputs the solution, statistical data, and the presumptions used to generate that solution.

Thus, multiple solutions are obtained with a spectrum indicating the likelihood of each, as a method of dealing with the inverse problem. Humans can make final analytic decisions from this information, or specify further analysis.

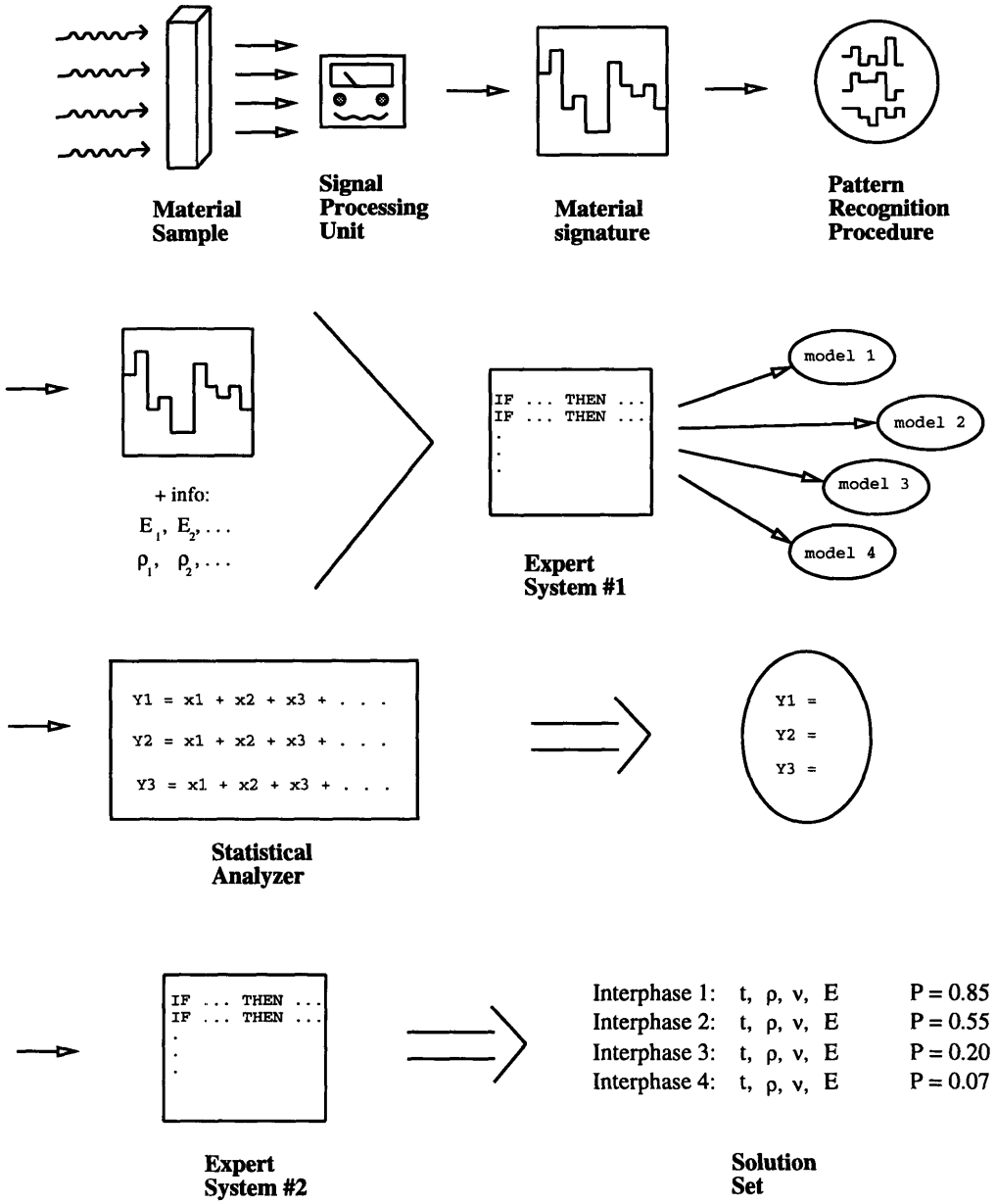

Figure 4-1: Artificial intelligence implementation for inverse interphase problem.

Chapter 5

Conclusions

It was shown that there exists considerable correlation (approximately 79%) between interphase thickness t and the scattered output wavefield components. Noticeable correlation of forty to sixty percent exists between the output wavefield components and the interphase density ρ and the interphase elastic modulus E. Correlation between any output wavefield component and interphase Poisson's ratio ν was limited to 24%, at best.

The results from regression analyses ranged from fair to exceptionally useful. Interphase thickness was described on the basis of the output wavefields. Improved estimates of interphase thickness were attained when other interphase parameters were known or were estimated.

Regression techniques enabled the prediction of t when output wavefield components were known. Precision of prediction increased when selected information about the interphase, such as interphase elastic modulus or interphase density, was known. The coefficient of multiple determination for such analyses ranged from 79.1% to 100%. The 95% prediction intervals on interphase thickness t ranged from ± 0.0035 to ± 0.0190. Smaller intervals were the result of added information about the composite interphase.

Higher methods were proposed which incorporate artificial intelligence techniques to address the nonunique aspect of the inverse interphase problem. These are used in concert with the statistical analysis techniques described in detail herein.

The methods presented here proved to be quite useful for the single fiber model data.

Besides their application to the completely unknown interphase state, it is important to realize that these methods are more powerful than previous methods for predicting interphase properties from this model [30] because they can be applied to other situations, such as a multiple fiber model or actual test data. No assumptions about the configuration or origin of the data set were made in the solution approach. The parameters which were used as predictor variables can be any NDE parameters which are measurable in the laboratory. The criterion variable for prediction can be any property or characteristic of interest. In this way the concepts presented here can be introduced into uses outside interphase property prediction, to experimental NDE in general.

References

[1] P. S. Theocaris. *The Mesophase Concept In Composites*. Polymers, properties and applications. Springer-Verlag, Berlin, 1987.

[2] J. Kardos. The role of the interface in polymer composites: Some myths, mechanisms and modifications. In H. Ishida and G. Kumar, editors, *Molecular Characterization of Composite Interfaces*, volume 27 of *Polymer Science and Technology*, part 1, pages 1–11. Plenum Press, New York, 1985.

[3] H. E. Kautz and B. A. Lerch. Preliminary investigation of acousto-ultrasonic evaluation of metal-matrix composite specimens. *Materials Evaluation*, 49(5):607–612, May 1991.

[4] D. Schneider, T. Schwarz, and B. Schultrich. Determination of elastic modulus and thickness of surface layers by ultrasonic surface waves. *Thin Solid Films*, 219(1–2):92–102, October 1992.

[5] A. Vary. Ultrasonic nondestructive evaluation microstructure and mechanical properties interrelation. Technical Memorandum 86876, NASA, October 1984.

[6] W. Sachse and I. Grabec. Intelligent processing of acoustic emission signals. *Materials Evaluation*, 50(7):826–834, July 1992.

[7] E. P. Papadakis. Future of ultrasonics. *Materials Evaluation*, 49(9):1180–1184, September 1991.

[8] A. Vary. NDE of the universe: New ways to look at old facts. *Materials Evaluation*, 51(3):380–387, March 1993.

[9] S. Dowdy and S. Wearden. *Statistics for Research*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., New York, second edition, 1991.

[10] K. Grotz and B. Lutz. Electromagnetic multiparameter determination of material characteristics. *Materials Evaluation*, 49(1):40–43, January 1991.

[11] T. Lund. Measurement and prediction of Jominy hardenability. *Scandinavian Journal of Metallurgy*, 19(5):227–235, 1990.

[12] K. Thomas, E. A. Geary, P. Avis, and D. Bishop. The prediction of properties in heat treated engineering steels. *Materials & Design*, 13(1):17–22, February 1992.

[13] E. v.K. Hill. Predicting burst pressures in filament-wound composite pressure vessels by using acoustic emission data. *Materials Evaluation*, 50(12):1439–1445, December 1992.

[14] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley series in probability and mathematical statistics. John Wiley and Sons, Inc., New York, second edition, 1981.

[15] E. L. Capener. Regression analysis to assess ultrasonic inspection results. *Materials Evaluation*, 46(5):671–678, April 1988.

[16] D. H. Harris. Effect of human information processing on the ultrasonic detection of intergranular stress-corrosion cracking. *Materials Evaluation*, 48(4):475–480, April 1990.

[17] I. Grabec, W. Sachse, and D. Grabec. Intelligent processing of ultrasonic signals for process control applications. *Materials Evaluation*, 51(10):1174–1182, October 1993.

[18] A. Zhang. A procedure for low cycle fatigue life prediction for various temperatures and strain rates. *Journal of Engineering Materials and Technology, Transactions of the ASME*, 112(4):422–428, October 1990.

[19] J. Saniie, T. Wang, and X. Jin. Ultrasonic grain signals classification using autoregressive models. In *Ultrasonic Symposium Proceedings, vol. 2*, Ultrasonics Symposium, pages 1155–1158, Piscataway, NJ, October 1989. IEEE Ultrasonics, Ferroelectrics & Frequency Control Society, IEEE Service Center.

[20] D. J. Roth, D. B. Stang, S. M. Swickard, M. R. DeGuire, and L. E. Dolhert. Review, modeling, and statistical analysis of ultrasonic velocity-pore fraction relations in polycrystalline materials. *Materials Evaluation*, 49(7):883–888, July 1991.

[21] V. A. S. Ayyadurai, S. A. Cimaszewski, and J. H. Williams, Jr. Unsupervised classification of fiber composite interphases. In A. Vary, editor, *Second International Conference on Acousto-Ultrasonics*, Topical Conference Proceedings Book, pages 29–38, Columbus, OH, June 1993. The American Society for Nondestructive Testing, Inc.

[22] Du Pont Lanxide Composites, Inc. Preliminary engineering data. Corporate literature, Newark, DE, April 1993.

[23] J. E. Hack, R. T. Pepper, and G. C. Strempek. Elevated temperature behavior of Al_2O_3 fiber reinforced aluminum composites. In I. Ahmad and B. R. Noton, editors, *Advanced Fibers and Composites for Elevated Temperatures*, pages 42–54, Warrendale, PA, 1980. AIME, Metallurgical Society of AIME.

[24] F. S. Galasso. *Advanced Fibers and Composites*. Gordon and Breach Science Publishers, New York, 1989.

[25] G. Simmons and H. Wang. *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook*. M.I.T. Press, Cambridge, Massachusetts, second edition, 1971.

[26] American Society for Metals, Reference Publications. *ASM Metals Reference Book.* American Society for Metals, Metals Park, Ohio, 1981.

[27] American Society for Metals, Reference Publications. *Metals Handbook Volume 2: Properties and Selection of Nonferrous Alloys and Pure Metals.* American Society for Metals, Metals Park, Ohio, ninth edition, 1979.

[28] H. W. Rauch, W. H. Sutton, and L. R. McCreight. *Ceramic Fibers and Fibrous Composite Materials*, volume 3 of *Refractory materials.* Academic Press, New York, 1968.

[29] American Society for Metals, Handbook Committee. *Engineered Materials Handbook Volume 1: Composites.* American Society for Metals International, Metals Park, Ohio, 1987.

[30] H. Yim. *Acousto-Ultrasonic Nondestructive Characterization of Interphases in Fiber Composites.* PhD dissertation, Massachusetts Institute of Technology, Department of Mechanical Engineering, May 1993.

[31] M. L. Berenson, D. M. Levine, and M. Goldstein. *Intermediate Statistical Methods and Applications.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1983.

[32] T. H. Wonnacott and R. J. Wonnacott. *Regression: A Second Course in Statistics.* Robert E. Krieger Publishing Company, Malabar, Florida, 1981.

[33] SAS Institute. Statistical analysis system, version 6.07. Cary, North Carolina, 1989.

[34] J. D. Achenbach. *Wave Propagation In Elastic Solids*, volume 16 of *Applied Mathematics and Mechanics.* North-Holland, Amsterdam, 1984.

[35] K. F. Graff. *Wave Motion In Elastic Solids.* Dover Publications, Inc., Mineola, New York, 1991.

[36] R. S. Sposili. A study of the effect of sizing chemistry on interfacial strength of graphite/polymer-matrix composites. Master's thesis, Columbia University, Henry Krumb School of Mines, Metallurgical Engineering, February 1994.

[37] D. A. Waterman. *A Guide to Expert Systems*. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1986.

[38] D. A. Waterman and P. Klahr, editors. *Expert Systems: Techniques, Tools and Applications*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1986.

[39] C. V. Negoita. *Expert Systems and Fuzzy Systems*. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, 1985.

[40] E. P. Papadakis. Imaging methodologies: Making decisions before seeing the images. *Materials Evaluation*, 51(2):120–124, February 1993.

[41] E. P. Papadakis. Image interpretation for quality decisions at production line speeds. *Materials Evaluation*, 51(7):816, July 1993.

[42] E. v.K. Hill, P. L. Israel, and G. L. Knotts. Neural network prediction of aluminum-lithium weld strengths from acoustic emission amplitude data. *Materials Evaluation*, 51(9):1040–1045, September 1993.

[43] J. M. Mann, L. W. Schmerr, and J. C. Moulder. Neural network inversion of uniform-field eddy current data. *Materials Evaluation*, 49(1):34–39, January 1991.

[44] L. Udpa and S. S. Udpa. Eddy current defect characterization using neural networks. *Materials Evaluation*, 48(3):342–353, March 1990.

[45] J. J. Thomsen and K. Lund. Quality control of composite materials by neural network analysis of ultrasonic power spectra. *Materials Evaluation*, 49(5):594–600, May 1991.

[46] C. E. Chapman, A. Fahr, A. Pelletier, and D. R. Hay. Artificial intelligence in the eddy current inspection of aircraft engine components. *Materials Evaluation*, 49(9):1090–1094, September 1991.

[47] R. Shankar, R. Williams, and M. J. Avioli, Jr. Knowledge-based ultrasonic examination assistant. *Materials Evaluation*, 49(10):1316–1321, October 1991.

[48] S. K. Kachigan. *Multivariate Statistical Analysis: A Conceptual Introduction*. Radius Press, New York, 1982.

[49] J. A. Jaffe. *Mastering the SAS System*. Van Nostrand Reinhold, New York, 1989.

[50] R. P. Cody and J. K. Smith. *Applied Statistics and the SAS Programming Language*. North-Holland, New York, second edition, 1987.

[51] SAS Institute. *SAS User's Guide: Basics*. SAS Institute, Cary, North Carolina, version 5 edition, 1985.

[52] G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. Wiley series in probability and mathematical statistics. John Wiley and Sons, Inc., New York, 1989.

[53] R. J. Freund and P. D. Minton. *Regression Methods: A Tool for Data Analysis*. Statistics, Textbooks and Monographs. M. Dekker, New York, 1979.

Appendix A

Data Generation

Data used for purposes of analysis have been created from a double precision program designed to run on UNIX based workstations, written and developed by Yim [30]. The FORTRAN code utilized the single fiber elastic model depicted in Fig. A-1 to approximate composite material behavior. The three-constituent model contains a single cylindrical scatterer, surrounded by the interphase layer of interest, embedded in an infinite elastic material. An excitation in the form of a steady-state plane longitudinal (P) or in-plane shear (S) stress wave is assumed at the boundary and the corresponding stress state is determined within the three distinct regions of fiber, interphase and matrix. The constituents are assumed to be elastic, and the interfaces between the constituents are perfectly bonded. The form of solution is that of Bessel series, and the output consists of stress and displacement components at particular locations in the matrix. The stresses and displacements at the specified locations are calculated for a set of the interphase parameters.

Input to the model include: type of excitation wave (P or SV), wavenumber of excitation wave, locations to measure stresses and displacements, physical and geometrical interphase properties (see Fig. A-2), particular stress and displacement components to be output, and form of output data (real or complex). The matrix and fiber properties are selected, and fixed.

The physical and geometrical interphase properties are: thickness (t) normalized

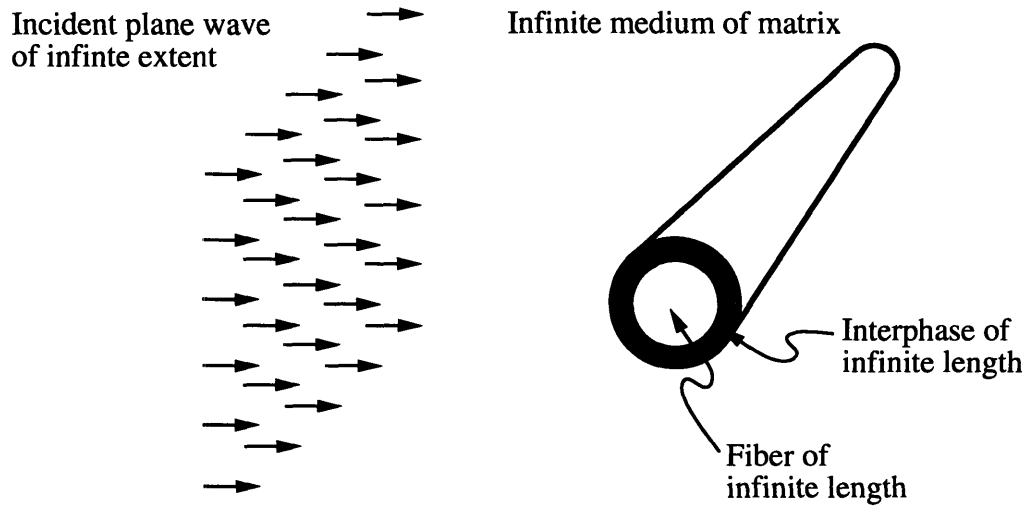

Figure A-1: Three-dimensional view of single fiber scatterer.

by the fiber radius a, density (ρ) normalized by the density of the matrix material, Poisson's ratio (ν), and elastic modulus (E) normalized by the elastic modulus of the matrix material.

Since the interphase properties of a particular composite are fixed, for purposes of study the properties are considered to vary. In this way an unknown interphase can be evaluated, presumably without knowledge of the actual interphase property values. In particular, the interphase thickness t is varied over a large range, in order to study thickness effects on the output wavefield. The interphase parameters are varied about the nominal values of the interphase properties to create the ranges of consideration. Also, the single fiber elastic model can only be analyzed for discrete interphase property values, so the ranges of interphase properties are discretized. The ranges over which the interphase properties vary are listed in Table A.1, along with the particular discrete values of the domain at which the output has been calculated. The ranges of consideration are based on the nominal values of the interphase properties, namely ρ_i, ν_i, E_i, and the fiber radius a.

An output wavefield is created for each set of interphase parameters considered. The stress components of this wavefield are shown in Fig. A-3. Two locations of particular interest are labeled A and B in Fig. A-3. Location A has r,θ coordinates of ($20a$,0)

Table A.1: Ranges and Discrete Values of Domain for Output Calculation.

	t	ρ	ν	E
range:	0.005 – 0.100	$0.5\rho_i/\rho_m$ – $1.5\rho_i/\rho_m$	0.1 – 0.4	$0.7E_i/E_m$ – $2.0E_i/E_m$
discrete values:	0.005	1.2115	0.10	1.0284
	0.010	1.4538	0.16	1.1754
	0.015	1.6962	0.22	1.3223
	0.020	1.9385	0.28	1.4692
	0.025	2.1808	0.34	1.6161
	0.030	2.4231	0.40	1.7631
	0.035	2.6654		1.9100
	0.040	2.9077		2.0569
	0.045	3.1500		2.2038
	0.050	3.3923		2.3507
	0.055	3.6346		2.4977
	0.060			2.6446
	0.065			2.7915
	0.070			2.9384
	0.075			
	0.080			
	0.085			
	0.090			
	0.095			
	0.100			

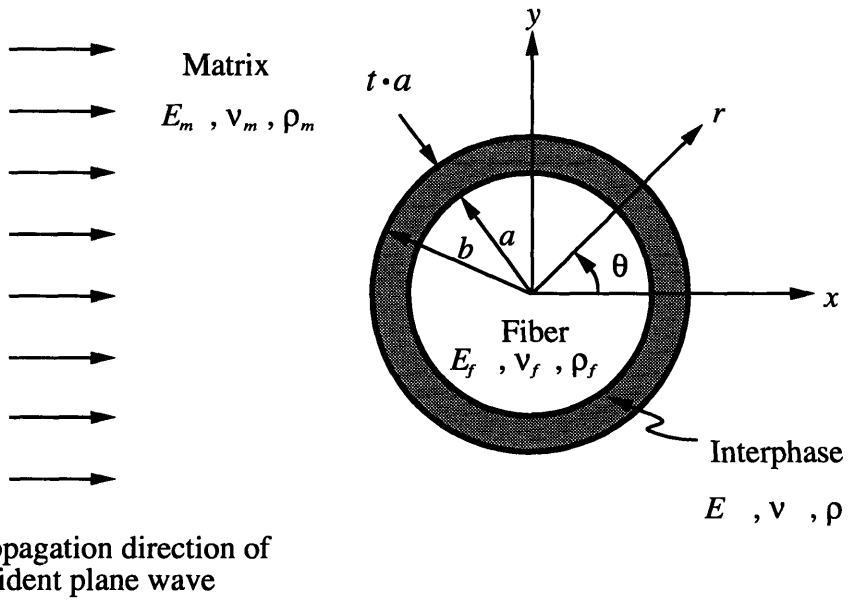

Figure A-2: Two-dimensional view of single fiber scatterer.

and location B has r,θ coordinates of $(20a,\pi)$. The wavefield components at location A have the additional subscript '1', and those at location B have the additional subscript '2'. Thus, the stress components at location A are σ_{rr1}, σ_{tt1} and σ_{rt1} and the stress components at location B are σ_{rr2}, σ_{tt2} and σ_{rt2}. Displacement wavefield components u_{r1} and u_{t1} exist at A and u_{r2} and u_{t2} exist at B. Also, the wavefield components at location A are called the *transmitted* wavefield components and the wavefield components at location B are called the *scattered* wavefield components. The scattered wavefield is the total wavefield minus the incident wavefield, also called the back scattered wavefield. All wavefield components are amplitudes normalized by the excitation wave amplitude. These output stress components are defined as follows:

σ_{rr1}: Amplitude of the normal stress at location A, normalized by the amplitude of the incident P-wave.

σ_{tt1}: Amplitude of the hoop stress at location A, normalized by the amplitude of the incident P-wave.

σ_{rr2}: Amplitude of the normal stress at location B, normalized by the amplitude of the incident P-wave.

σ_{tt2}: Amplitude of the hoop stress at location B, normalized by the amplitude of the incident P-wave.

σ_{rt1}: Amplitude of the shear stress at location A, normalized by the amplitude of the incident S-wave.

σ_{rt2}: Amplitude of the shear stress at location B, normalized by the amplitude of the incident S-wave.

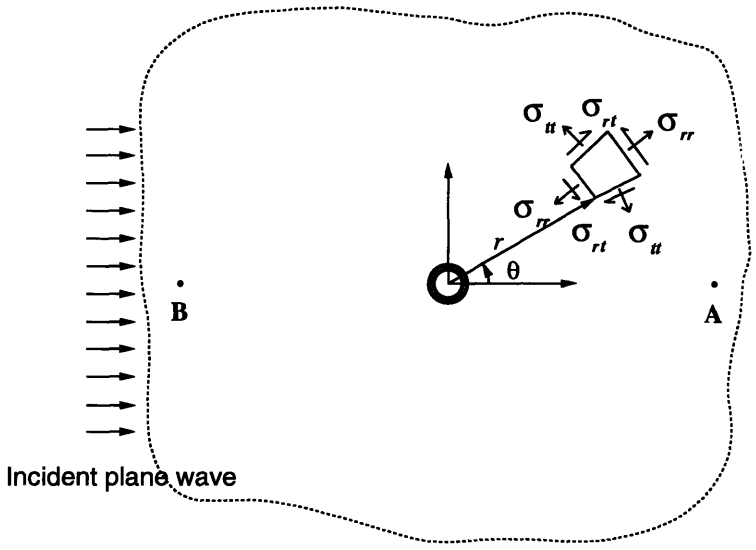

Figure A-3: Wavefield stress components.

The material model is subjected to excitations of P and SV plane waves with different nondimensionalized wavenumbers summarized in Table A.2. Nondimensional wavenumbers resulted from multiplying the wavenumber by the fiber radius a to yield a unitless quantity to represent the scatterer. Such wavenumbers have been determined from spectral plots which have indicated that an interphase is most discernible at the values chosen. Solution wavefields are calculated individually, as if only one excitation is applied at any particular time. The output components from each of these excitation waves is also listed in Table A.2.

The actual data set consists of 18,480 lines of data known as *records*. Each record consists of a particular interphase material state and the accompanying wavefield it creates when excited. The interphase state in each record contains the following interphase

Table A.2: Applied Steady Stress Wave Excitations.

wave type	wavenumber	non-zero outputs	outputs of interest
P	22.50	σ_{rr1}, σ_{tt1}, u_{r1}, σ_{rr2}, σ_{tt2}, u_{r2}	σ_{rr1}, σ_{tt1}, u_{r1}
P	10.35	σ_{rr1}, σ_{tt1}, u_{r1}, σ_{rr2}, σ_{tt2}, u_{r2}	σ_{rr2}, σ_{tt2}, u_{r2}
S	26.35	σ_{rt1}, u_{t1}, σ_{rt2}, u_{t2}	σ_{rt1}, u_{t1}
S	11.10	σ_{rt1}, u_{t1}, σ_{rt2}, u_{t2}	σ_{rt2}, u_{t2}

properties: thickness t, density ρ, Poisson's ratio ν, and modulus of elasticity E. The output wavefield components in each record are: σ_{rr1}, σ_{tt1}, u_{r1}, σ_{rr2}, σ_{tt2}, u_{r2}, σ_{rt1}, u_{t1}, σ_{rt2}, u_{t2}. Both P and SV wave excitations are considered for each interphase material state; however in the data set the wavefields resulting from each excitation are contained in the same record along with the interphase properties which created them.

Appendix B

Correlation

B.1 Purpose

Among the most widely used comparison procedures in statistics are forms of correlation. It is used as an *exploratory* tool, that is one designed for prediction and/or evaluation purposes. There are many variants of correlation, such as simple correlation, canonical correlation, or serial correlation. All have in common the basic idea that two quantities are under consideration, although the two quantities may be aggregate or otherwise processed quantities of data.

In order to determine if there is a linear relationship between two scalar quantities, *simple* or *bivariate correlation* is used. The two variables under consideration may or may not be simply related or may not even possess the same units. Correlation merely intends to describe the change of one with respect to the change of the other, i.e. the degree to which the variables are linearly related. It is concerned with measuring the *relationship* or *strength of association* among variables [31].

B.2 Procedure

The measure of correlation between two variables, say, x and y, is called the *correlation coefficient* r, defined as [14]

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{B.1}$$

where x_i and y_i are the ith observations of x and y in the data set, respectively and $\bar{x}$ and $\bar{y}$ are the average values of x and y, respectively.

The correlation coefficient can range from -1 to $+1$. Thus, a perfect positive correlation between two variables is equivalent to a correlation coefficient of $+1$, while a perfect negative correlation is equivalent to a value of -1. A value of zero indicates no correlation between two variables. In the case of perfect correlation, one variable can be expressed as a function of the other, such as $y = 2x - 3$, or $y = x$. Fig. B-1 [32] depicts sample x, y plots for various correlation coefficients r.

Nonlinear correlation is the application of simple correlation to transformed variables. That is, should one desire to find the correlation between a quantity y and a quantity x^2, just two steps are necessary:

1. The set of data x^2 is created for each value in the domain of x.

2. Simple correlation between y and the transformed variable x^2 is performed.

In a similar way other transformed variables can be considered for correlation.

Multivariate correlation analysis as the name implies, is concerned with the correlations that exist among several variables. "Multivariate" suggests that many (i.e. more than two) variables are involved, and indeed this is the general case of the bivariate situation. The only requirement is that all observations of data are made with respect to each and every variable under consideration. A *correlation matrix* is used as a convenient representation of the correlation coefficients among the variables.

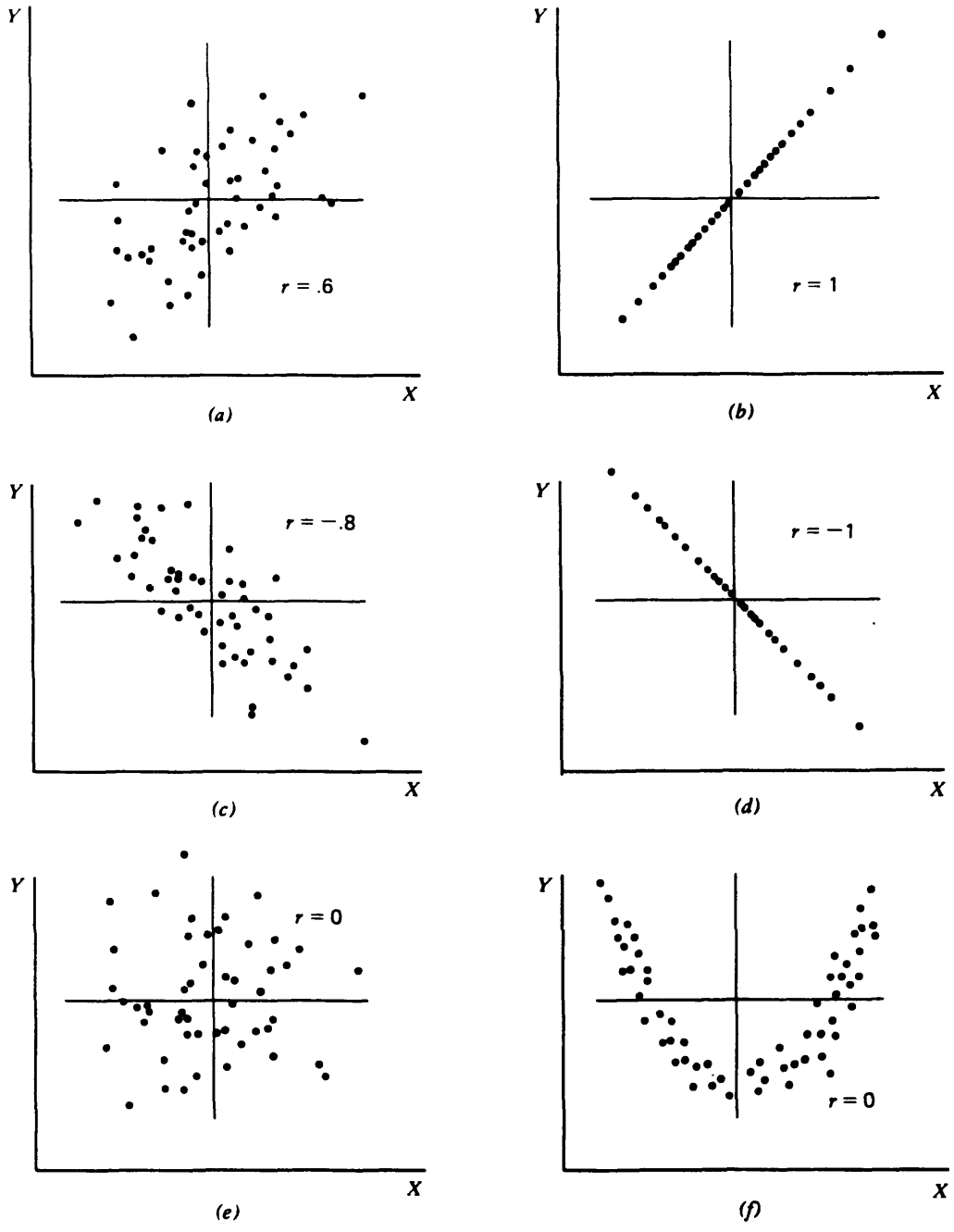

Figure B-1: Some sample scatters to illustrate various values of r.

Appendix C

Multiple Regression

C.1 Purpose

In order to discuss multiple regression, a comment on terminology is necessary. The terms *criterion variable* and *predictor variable* are used in lieu of the more common terms *dependent variable* and *independent variable*, respectively [48]. The reason for this is to avoid confusion with the more widespread definition of the mathematical terms *dependent* and *independent*. Dependent relationships are those where a known value of one variable fixes the value of another *dependent* on it, such as in the function $y = x + 4$, where y is fixed by x and thus x is the independent variable and y is the dependent variable.[1] And when two variables are said to be *independent*, the implication is that they do not covary with one another to any degree.

In correlation and regression analysis, those perceptions are not entirely an accurate description of the underlying relationships. While two variables may in fact generally increase with one another, they may not be mathematically dependent. Similarly if we consider a model with several "independent" variables, then according to our definition the model would be of no use because the variables are independent of the variable of interest. To avoid this confrontation, the name *criterion variable* is applied to the variable

[1] Or equivalently, the expression $x = y - 4$ where x is the dependent variable and y is the independent variable.

we are interested in predicting, and the name *predictor variable* or *predictor* is applied to the variable about which some knowledge is assumed, for prediction. In the case of regression analysis only, the predictor variables are also called *regressor variables* or *regressors*.

Regression is a method of fitting a function to a set of x,y points in order to use the function for subsequent interpolation. The function will depend on the data set considered, but is not sensitive to small changes in the contents of the data set. This fact makes regression techniques ideally suited for scientific prediction. Multiple regression extends this concept to data sets of more than two variables.

Multiple regression accounts for more than one predictor variable x. It allows the investigation of the effects on the criterion variable y of several variables simultaneously. Even if one predictor variable is deemed sufficient for prediction by some analysis, it is wise to include other variables influencing y for several other reasons [32]:

1. To reduce *stochastic error*, or random error due to inherent variability. This in turn reduces the residual variance (that portion of variability unexplained by regression modeling), which has the important effect of making prediction intervals more precise.

2. To eliminate bias due to considering a single variable as affecting y. Since it is assumed that a perfect correlation with the criterion variable is not available (lest one would not be looking to this approach), any single predictor will tend to either overestimate or underestimate the criterion variable, generally. The implication is that with more than one variable, the bias will tend to decrease.

As a consequence of implementing multivariate concepts, a better overall fit to the data is available, i.e. the average fluctuation of the data points about the fitted model is reduced. This allows more precise statistical conclusions about how a given x affects y. It also permits the effect of one predictor alone on y to be studied as well as the combined effect of several predictors. The end goal of all regression analyses is to reduce the error of prediction, or equivalently to account for more of the variance in the criterion variable.

C.2 Formulation

C.2.1 Regression Model

The multiple regression equation has the form [9]

$$\hat{y} = \alpha_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k \tag{C.1}$$

where $\hat{y}$ is the predicted value of the criterion variable, α_o is the intercept or constant in the equation (which is often called *bias*), $\beta_{j,(j=1,2,3,...,k)}$ are the coefficients corresponding to the predictor variables x_j, and $x_{j,(j=1,2,3,...,k)}$ are predictor variables, namely, some parameters which can be measured, known or estimated.

For a given set of data, the regression model utilizes the least squares method to determine the regression coefficients α_o and β_j. Similar to simple regression, the process is to determine the regression coefficients which minimize the squared residual error sum $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, which is expressed as a function of the α_o and β_j. The term $(y_i - \hat{y}_i)$ is called the *residual* at a point, and y_i is the actual value of the point in the data set whereas $\hat{y}_i$ is the value predicted by the regression model. For a regression equation with k predictors on a set of data with n elements, the function to minimize becomes a set of $k+1$ simultaneous linear equations for the unknowns α_o and β_j. The solution to this set of equations is unique, thereby providing one unique regression equation for a particular model and set of data.

Multiple regression models are readily adaptable to the nonlinear case, taking the x_j to be transformed variables of the original predictors. The nonlinear models are nonlinear only in the predictor variables, not in the coefficients of those variables. Therefore, the general expression for a model of this form is:

$$\hat{y} = \alpha_o + \beta_{n1} x_n + \beta_{n2} x_n^p + \beta_{n3} x_n^q + ... + \beta_{m1} x_m + \beta_{m2} x_m^r + \beta_{m3} x_m^s + ... \tag{C.2}$$

where p, q, r and s indicate powers of x_j, which may not be equal.

Here we can have $x_l = x_n^q$, with l and n denoting arbitrary subscripts and q some

power of interest. Then the transformed variable x_l is treated just like another regressor, as the model has no knowledge of the transformation. This technique is often referred to as *polynomial regression.* This is a misnomer of sorts, since q can be any real number, and is for no reason restricted to integers.

Once the model is established, its capability to predict must be evaluated. For this there are two statistical measures of model effectiveness: coefficient of multiple determination R^2 and Mallows' statistic C_p.

C.2.2 Coefficient of Determination R^2

The *coefficient of multiple determination* is defined as [32]

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{C.3}$$

or, more succinctly as

$$R^2 = \frac{SSR}{S_{yy}} \tag{C.4}$$

where SSR is $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ and S_{yy} is defined by $\sum_{i=1}^{n}(y_i - \bar{y})^2$. SSR is known as the *sum of squares due to regression,* while S_{yy} is called *total sum of squares.*

The determination coefficient is a ratio of the variability in y explained by the regression model to the total variability in criterion variable y. For instance, a model which explains y totally has R^2 equal to 1, while a model with no ability to explain y will have a value of zero. Fig. C-1 offers a graphical explanation of the terms in the eqn. C.3 for R^2. The variation of a single point is $y_i - \bar{y}$, and the variation of a single estimated point is $\hat{y}_i - \bar{y}$. Thus, a measure of the entire variabilty in a data set is available from $\sum_{i=1}^{n}(y_i - \bar{y})^2$. The portion of the variability in a data set which is explained by the multiple regression model is $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$. Hence, R^2 measures the fraction of variability in a data set that can be accounted for through regression.

It is important that two models of different k quantities not be compared on the basis

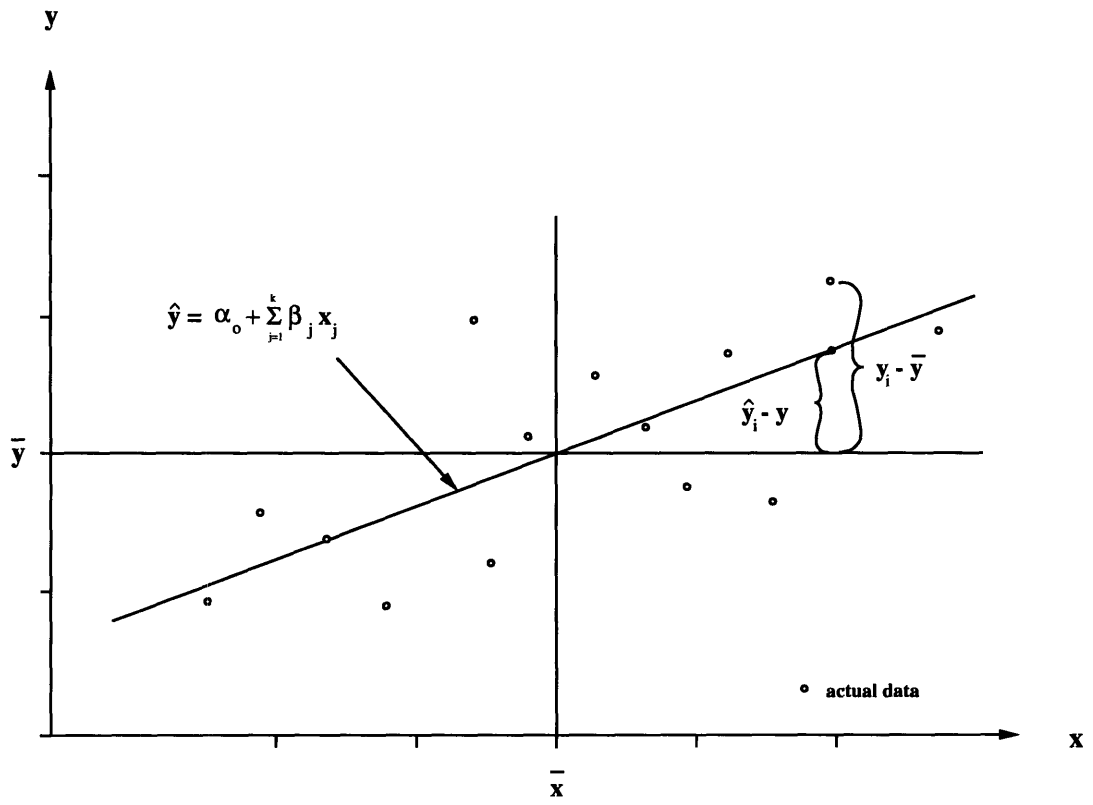

Figure C-1: Measures of variation in a data set.

of R^2 alone; this is because R^2 will increase for a particular model with the addition of any regressor variable, as long as β_j for that regressor is not zero. The adjusted R^2 is needed for comparison between two regression models of interest when the difference in number of parameters becomes large. It is calculated from the standard coefficient of multiple determination according to [9]

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \tag{C.5}$$

Comparing R^2 and R^2_{adj} permits more accurate quantification of the relative usefulness of more than one regression model.

C.2.3 Mallows' Statistic C_p

Mallows' statistic C_p also permits comparison of the effectiveness of several models. It is a more complex summary of the model behavior, and aids in determining what experimental value, if any, is realized from the addition of another predictor variable to the regression equation. Its definition is [9]

$$C_p = \frac{(S_{yy} - SSR)}{MSE_{fullmod}} - n + 2p \tag{C.6}$$

where n is the number of observations in the data set, p is the number of parameters (predictor variables *and* intercept) in the regression equation and $MSE_{fullmod}$, known as the *model variance*:

$$MSE_{fullmod} = \frac{S_{yy}^{fullmod} - SSR^{fullmod}}{n - k - 1} \tag{C.7}$$

where k is the maximum number of predictor variables (or regressors) in the equation, with all possible variables included.

This statistic weighs the increase in variability explained by the addition of a regressor to the total model variance, accounting for the added complexity of the additional term. That is to say, the C_p statistic is a ratio of the total variance unaccounted for in the model to the total model variance possible, and then adjusted for the number of parameters in the equation. It is essentially measuring the "unexplained variance per regressor" and therefore is a good quantity for comparing to models generated from the same data. A desirable C_p is a low value, typically very close to the value of p. The full model always has a C_p of p.

Most important when analyzing C_p values is that any models used have similar C_p values. Although the lowest value represents the best model from this standpoint, the range on C_p can be great; thus, any two similar values indicate acceptable models. A researcher typically uses this information in concert with other information to make a decision about which model to choose; it is a part of a voluminous amount of secondary information which is available for model evaluation.

C.2.4 Prediction Intervals

Multiple regression yields coefficients which are used to create a model of a particular data set. The regression equation will then output one value for any given combination of input predictor values. For predicting the single observed value y_o, the best estimate is the point on the regression line at x_o, $\hat{y}_o = \alpha_o + \sum_{j=1}^{k} \beta_j (x_o)_j$.

Since any solution surface is likely to vary from the actual data points, analytical techniques used for prediction often yield an interval estimate for the solution. This is true in regression, and the interval is the *prediction interval.* The prediction interval for a value predicted by regression is defined as [9]

$$PI_{1-\alpha} = \hat{y} \pm t_{\frac{\alpha}{2},n-k-1} \sqrt{s_{y,x}^2 (1 + \frac{1}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} p_{ij} (x_i^* - \bar{x}_i)(x_j^* - \bar{x}_j))} \qquad \text{(C.8)}$$

where $(1 - \alpha)$ is the percent confidence desired, $t_{\frac{\alpha}{2},n-k-1}$ is the *student's t* statistic [9] which in the limit as $n \to \infty$ approaches the *normal* z statistic [9], x_i^* is the specific known value of the predictor x_i, and p_{ij} are the elements of the solution coefficient matrix from the minimization of residuals [9]. $s_{y,x}^2$ is the variance about the trend line, and is computed as [9]

$$s_{y,x}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - k - 1} \qquad \text{(C.9)}$$

The quantity $s_{y,x}^2$ is the greatest contributor to the prediction interval size, since $s_{y,x}^2$ is dependent not only on the regression performed but also on the variability in the data set.

Concisely stated, the prediction interval is that region of the criterion variable range within which future observations of new data are found, with a specified confidence. It is an expression which accounts for the variability of the data, the regression performed, and the number of observations in the data set.

The prediction interval curve takes the general shape shown in Fig. C-2. Thus it is apparent that the intervals are *not* the same at every point of the domain, and tend to be wider towards the ends of the predictor domain. This is due to the nature of the analysis

in the respect that the interpolation near the endpoints is not as reliable as that closer to $\bar{x}$.

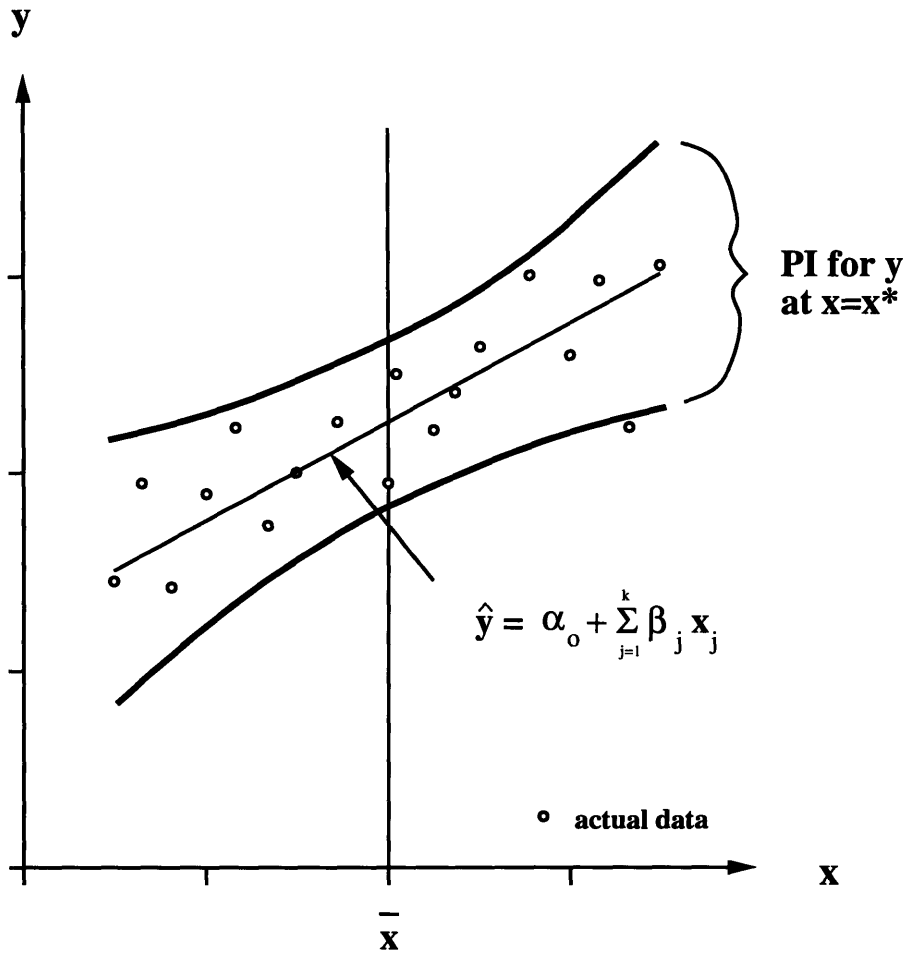

Figure C-2: Prediction intervals.

The limits of the prediction interval are those within which the specified confidence percentage of future observations are expected to lie. This confidence percentage can be increased albeit at the "cost" of wider intervals; conversely, the prediction interval can be decreased in size at the expense of reduced confidence [14].

The type of techniques discussed herein are called *inferential statistics* because they involve the drawing of conclusions regarding population characteristics (parameters) based on information (statistics) obtained from the analysis. They deal with both estimating parameters and testing hypothesized values of the population parameters.

Appendix D

Stepwise Regression

D.1 Purpose

When there are a large number of regressors, it becomes important to determine which ones are important to have in the model for prediction purposes. It also becomes important to protect against "data mining," or using the data to determine what the model should be, instead of choosing a model based on suspected relationships or additional information [32]. *Multicollinearity* can become a problem due to two regressors being highly correlated to one another. *Stepwise regression* is a process which starts with the simplest model and then adds regressors one at a time, in order of their contribution to the model. Additionally it will check all regressors at each step and delete those whose contribution drops due to the addition of another predictor variable. As a result the model has those (and only those) variables which are statistically significant for the prediction of the criterion variable. This procedure and ones like it are often called *search* procedures because of their incremental checking nature. Stepwise regression is perhaps the most widely used search procedure.

The most important benefit of this procedure is that the uncertainty associated with the question of which regressors to choose has essentially been eliminated from the problem under study. Accordingly, other effects can be analyzed without this complication.

D.2 Procedure

The procedure takes place in steps, which are either an addition or deletion of a variable from the model. The steps can be summarized as follows [31]:

1. The stepwise process begins by including in the model the one predictor variable x having the highest simple correlation with the criterion variable y. Recall the simple population correlation coefficient for a group of data was defined by:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}}$$

   A statistical test called an *F test* [14] is used to determine the statistical significance of a variable in a regression equation. If the F statistic for this variable equals or exceeds the critical value needed to become part of the model equation (denoted F_E), the variable is included and the process continues. Should it be less than the critical value F_E, then the process terminates and no independent variables are included in the model, that is no model is created.

2. Once the first independent variable (called x_1) is included in the model, the contribution of each of the remaining $k - 1$ predictor variables is determined for the model that already includes variable x_1. A *partial F test* [14] (an F test which accounts for the already included regressors) is then performed for each of these variables, given that variable x_1 is already included in the regression model. If the largest partial F value (largest statistical significance) from these tests equals or exceeds the critical value of F needed to enter the model (F_E), then a second variable (call it x_2) is included in the model. On the other hand, if the largest partial F is less than the critical value F_E, the process terminates and only one variable is included in the regression equation.

3. A feature of the stepwise procedure is that a variable which has been entered into the model at an earlier stage may subsequently be removed once other predictor variables have been evaluated. Presuming at this point that the model contains at least two predictor variables (x_1 and x_2), it may now be determined whether any of the variables already included are no longer important given that others have been subsequently added. Then these could be deleted from the model. A partial F statistic for each variable already included in the model is computed, taking into account the effect of the other variables in the model. If the smallest of these F values (least significant) is less than or equal to the critical value of F for removal (denoted F_R), then the particular variable corresponding to that F value is removed from the regression model. However, if the smallest partial F value is greater than F_R, then the corresponding variable is not removed from the model.

4. To complete the process for the remaining predictor variables, the last two steps are repeated until there are no other variables to consider as regressors, or no other variable is significant enough to be added to the model. In this fashion an optimized model is attained which contains only variables which are significant to the model.

Appendix E

Data Set

The following is a condensed version of the original data set used. Ellipsis points appear where data have been omitted.

$t,\ \rho,\ \nu,\ E,\ \sigma_{rr1},\ \sigma_{tt1},\ u_{r1},\ \sigma_{rr2},\ \sigma_{tt2},\ u_{r2},\ \sigma_{rt1},\ u_{t1},\ \sigma_{rt2},\ u_{t2}$

0.0050 1.2115 0.1000 1.0284 0.418071 0.188308 0.418120 0.123810 0.055622 0.123812 0.365506 0.366617 0.102449 0.102456

0.0050 1.2115 0.1000 1.1754 0.414292 0.186641 0.414343 0.124015 0.055714 0.124017 0.362124 0.363239 0.102355 0.102361

0.0050 1.2115 0.1000 1.3223 0.411475 0.185393 0.411527 0.124175 0.055786 0.124177 0.359472 0.360592 0.102299 0.102306

0.0050 1.2115 0.1000 1.4692 0.409332 0.184439 0.409384 0.124303 0.055844 0.124305 0.357329 0.358452 0.102271 0.102277

0.0050 1.2115 0.1000 1.6161 0.407678 0.183697 0.407730 0.124406 0.055890 0.124408 0.355553 0.356679 0.102262 0.102268

0.0050 1.2115 0.1000 1.7631 0.406389 0.183115 0.406441 0.124490 0.055928 0.124492 0.354050 0.355178 0.102266 0.102273

0.0050 1.2115 0.1000 1.9100 0.405380 0.182655 0.405432 0.124559 0.055959 0.124560 0.352756 0.353886 0.102282 0.102288

0.0050 1.2115 0.1000 2.0569 0.404590 0.182291 0.404642 0.124615 0.055984 0.124617 0.351625 0.352756 0.102306 0.102312

0.0050 1.2115 0.1000 2.2038 0.403974 0.182003 0.404026 0.124662 0.056005 0.124663 0.350624

0.351756 0.102336 0.102343

0.0050 1.2115 0.1000 2.3507 0.403499 0.181777 0.403551 0.124700 0.056022 0.124702 0.349727 0.350860 0.102372 0.102379

0.0050 1.2115 0.1000 2.4977 0.403140 0.181601 0.403191 0.124732 0.056036 0.124733 0.348916 0.350049 0.102412 0.102419

0.0050 1.2115 0.1000 2.6446 0.402876 0.181469 0.402928 0.124757 0.056048 0.124759 0.348176 0.349309 0.102456 0.102462

0.0050 1.2115 0.1000 2.7915 0.402694 0.181372 0.402745 0.124778 0.056057 0.124779 0.347495 0.348629 0.102503 0.102509

0.0050 1.2115 0.1000 2.9384 0.402579 0.181305 0.402630 0.124794 0.056064 0.124796 0.346864 0.347998 0.102552 0.102558

0.0050 1.2115 0.1600 1.0284 0.417241 0.187885 0.417289 0.123775 0.055607 0.123777 0.366617 0.367727 0.102531 0.102537

.

.

.

0.0450 3.6346 0.4000 2.0569 0.536312 0.239659 0.536300 0.054270 0.024369 0.054270 0.344340 0.345450 0.083829 0.083834

0.0450 3.6346 0.4000 2.2038 0.523452 0.233618 0.523430 0.053294 0.023930 0.053294 0.329093 0.330476 0.085547 0.085551

0.0450 3.6346 0.4000 2.3507 0.510834 0.227974 0.510810 0.052461 0.023555 0.052461 0.324799 0.326398 0.087159 0.087164

0.0450 3.6346 0.4000 2.4977 0.499094 0.223159 0.499084 0.051759 0.023240 0.051759 0.324465 0.326229 0.088689 0.088693

0.0450 3.6346 0.4000 2.6446 0.488985 0.219395 0.489000 0.051174 0.022976 0.051173 0.326018 0.327918 0.090152 0.090156

0.0450 3.6346 0.4000 2.7915 0.480637 0.216385 0.480676 0.050687 0.022757 0.050687 0.329559 0.331582 0.091561 0.091565

0.0450 3.6346 0.4000 2.9384 0.473285 0.213436 0.473336 0.050277 0.022572 0.050277 0.335316 0.337448 0.092924 0.092928

0.0500 1.2115 0.1000 1.0284 0.473023 0.213021 0.473075 0.124388 0.055882 0.124390 0.367424 0.368597 0.090904 0.090908

0.0500 1.2115 0.1000 1.1754 0.475451 0.213972 0.475496 0.121051 0.054383 0.121052 0.331064 0.332354 0.089401 0.089405

0.0500 1.2115 0.1000 1.3223 0.461929 0.207778 0.461968 0.118923 0.053426 0.118925 0.307487

0.308925 0.088643 0.088648

0.0500 1.2115 0.1000 1.4692 0.440289 0.198101 0.440327 0.117577 0.052821 0.117578 0.292242 0.293826 0.088319 0.088323

0.0500 1.2115 0.1000 1.6161 0.416388 0.187590 0.416432 0.116751 0.052450 0.116753 0.282225 0.283937 0.088254 0.088258

0.0500 1.2115 0.1000 1.7631 0.393805 0.177815 0.393860 0.116282 0.052239 0.116284 0.275454 0.277275 0.088346 0.088350

.

.

.

0.0850 2.6654 0.4000 2.4977 0.436211 0.199931 0.436379 0.081366 0.036555 0.081367 0.406915 0.408810 0.054859 0.054861

0.0850 2.6654 0.4000 2.6446 0.436277 0.199677 0.436435 0.082890 0.037241 0.082891 0.394096 0.395812 0.058103 0.058105

0.0850 2.6654 0.4000 2.7915 0.429866 0.195742 0.429990 0.084019 0.037749 0.084021 0.365263 0.366826 0.061020 0.061021

0.0850 2.6654 0.4000 2.9384 0.419520 0.189684 0.419597 0.084796 0.038098 0.084798 0.332529 0.334032 0.063669 0.063671

0.0850 2.9077 0.1000 1.0284 0.352289 0.158351 0.352344 0.017663 0.007936 0.017663 0.465531 0.468302 0.017561 0.017579

0.0850 2.9077 0.1000 1.1754 0.495692 0.225781 0.495860 0.001331 0.000597 0.001331 0.674065 0.677264 0.022551 0.022567

0.0850 2.9077 0.1000 1.3223 0.682652 0.305296 0.682680 0.012183 0.005474 0.012184 0.680439 0.683205 0.026293 0.026307

0.0850 2.9077 0.1000 1.4692 0.758406 0.341312 0.758488 0.017996 0.008086 0.017996 0.625744 0.628272 0.031193 0.031204

0.0850 2.9077 0.1000 1.6161 0.739844 0.334829 0.739978 0.020500 0.009210 0.020500 0.612996 0.615039 0.036990 0.036998

.

.

.

0.1000 3.6346 0.3400 2.9384 0.397353 0.175803 0.397284 0.067522 0.030336 0.067523 0.467890 0.469304 0.060746 0.060760

0.1000 3.6346 0.4000 1.0284 0.480604 0.217389 0.480697 0.113714 0.051091 0.113716 0.572137 0.576394 0.035229 0.035265

0.1000 3.6346 0.4000 1.1754 0.738006 0.334128 0.738162 0.107419 0.048262 0.107420 0.508559 0.511155 0.022377 0.022329

0.1000 3.6346 0.4000 1.3223 0.711451 0.323425 0.711623 0.092469 0.041544 0.092470 0.375567 0.377876 0.047445 0.047418

0.1000 3.6346 0.4000 1.4692 0.643528 0.292150 0.643670 0.079733 0.035824 0.079735 0.518759 0.521661 0.069735 0.069731

0.1000 3.6346 0.4000 1.6161 0.611279 0.276900 0.611392 0.069877 0.031393 0.069878 0.545168 0.547694 0.088809 0.088828

0.1000 3.6346 0.4000 1.7631 0.560440 0.254759 0.560573 0.063328 0.028446 0.063328 0.460664 0.462560 0.100299 0.100333

0.1000 3.6346 0.4000 1.9100 0.526865 0.240149 0.527016 0.062323 0.028000 0.062324 0.381823 0.383688 0.100469 0.100507

0.1000 3.6346 0.4000 2.0569 0.520321 0.235938 0.520433 0.062584 0.028122 0.062585 0.405538 0.407964 0.092149 0.092182

0.1000 3.6346 0.4000 2.2038 0.510972 0.228480 0.510976 0.060258 0.027070 0.060258 0.499995 0.502831 0.081130 0.081155

0.1000 3.6346 0.4000 2.3507 0.473548 0.209047 0.473460 0.058434 0.026242 0.058434 0.582975 0.585938 0.071701 0.071718

0.1000 3.6346 0.4000 2.4977 0.413248 0.182468 0.413168 0.060847 0.027329 0.060847 0.614673 0.617544 0.065613 0.065626

0.1000 3.6346 0.4000 2.6446 0.349046 0.156029 0.349042 0.065365 0.029368 0.065366 0.598306 0.600905 0.062446 0.062458

0.1000 3.6346 0.4000 2.7915 0.305485 0.137560 0.305524 0.068018 0.030562 0.068020 0.559192 0.561431 0.060802 0.060815

0.1000 3.6346 0.4000 2.9384 0.310640 0.139151 0.310667 0.068556 0.030798 0.068557 0.509927 0.511774 0.059575 0.059588

Appendix F

SAS^{tm} Programs

The following are the programs input into the SAS^{tm} package to perform the necessary calculations to complete this research. Although some values have been changed for particular analyses, one of each program utilized is provided. Also, other programs have been utilized to prepare the data for input into the SAS^{tm} architecture; such programs are not included here. Discussion of SAS^{tm} features and usage is found in references [49-51].

Program 1: Tglm

```
title1  'First Iter Analysis:  linear in stresses only';
title2  'Infile is paro.nonu';

data par00;
   infile '/mit/bitbucket/steve/thesis/paro.nonu';
   input t r n e srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2;
   srtsq = srt2*srt2;
   srrsq = srr2*srr2;
run;                                                          10


proc corr data=par00 ;
   var t--srrsq ;
run;
```

```
 title3  'Correlation Analysis';


* Thickness Analysis;
                                                                                    20
* proc stepwise data=par00 ;     title3;
*    model t= srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2  /stepwise sle=.05 sls=.05;
* run;


 proc stepwise data=par00 ;     title3;
   model t= srr1          srr2          srt1     srt2      /stepwise sle=.05 sls=.05;
 run;


                                                                                    30
* proc stepwise data=par00 ;     title3;
*    model t= srr1          srr2          srt1     srt2     /stepwise sle=.05 sls=.05;
* run;


* data interp;
*    input t  r n e srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2;
*    cards;
*     . . . .  .2333 .1567 . .1263 .0010 . .8130 . .1099  .
*     . . . .  .6500 .3700 . .0104 .0600 . .9010 . .0545  .                        40
*     . . . .  .3078 .1934 . .0978 .0123 . .4567 . .0543  .
*     ;
* run;


* data res;
*    set par00 interp ;
* run;


                                                                                    50
```

```
* proc glm data=res;            * title2  'T regressed on stresses';
*    model t= srr1      srr2      srt1 srt2    / cli alpha=.10;
*    output out=guess predicted=predt ;
* run;


 proc stepwise data=par00 ;     title3;
    model t= srr1 srtsq srr2 srrsq srt1 srt2 /stepwise sle=.05 sls=.05;
 run;
                                                                          60

* proc stepwise data=par00 ;     title3;
*    model t= srr1       srr2       srt1 srt2 /stepwise sle=.05 sls=.05;
* run;


 proc glm data=res;              title2  'T regressed on stresses';
    model t=srr1   srr2 srt1 srt2 srtsq srrsq/cli alpha=.10;
*    output out=guess predicted=predt;
 run;                                                                     70


 title1  'Plot analysis :::' ;
 title2  'Infile is paro.005' ;

 proc plot data=par00 ;
    plot t*srr1; plot t*srr2; plot t*srt1; plot t*srt2;plot t*srtsq;plot t*srrsq;
 run;
```

Program 2: Treg

```
title1 'First Iter Analysis:  linear in stresses only';
title2 'Infile is parot.dat';

data par00;
   infile '/mit/bitbucket/steve/thesis/parot.dat';
   input t r n e srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2;
run;

proc corr data=par00 ;                                                    10
   var t--ut2 ;
run;
   title3 'Correlation Analysis';

* Thickness Analysis;

proc stepwise data=par00 ;     title3;
   model t= srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2 /stepwise sle=.05 sls=.05;
run;
                                                                          20

proc stepwise data=par00 ;     title3;
   model t= srr1       srr2       srt1    srt2    /stepwise sle=.05 sls=.05;
run;


* proc stepwise data=par00 ;     title3;
*    model t= srr1       srr2       srt1    srt2    /stepwise sle=.05 sls=.05;
* run;
                                                                          30

* Modulus (E) Analysis;
```

```
* proc stepwise data=par00 ;      title3;
*    model e= srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2 /stepwise sle=.05 sls=.05;
* run;


 data interp;
    input t  r n e srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2;                    40
    cards;
    . . . .  .2333 .1567 . .1263 .0010 . .8130 . .1099  .
    . . . .  .6500 .3700 . .0104 .0600 . .9010 . .0545  .
    . . . .  .3078 .1934 . .0978 .0123 . .4567 . .0543  .
    ;
 run;


 data res;
    set par00 interp;                                                                50
 run;


 proc reg data=res;              title2  'T regressed on stresses';
    model t= srr1      srr2      srt1 srt2    / clm cli;
*    output out=guess predicted=predt;
 run;
```

Program 3: classify

```
options nodate nonumber pagesize=1800 linesize=80;

title1 'First Iter Analysis: linear in stresses only';
title2 'Infile is paro.nonu';

data par00;
   infile '/mit/bitbucket/steve/thesis/paro.nonu';
   input t r n e srr1 stt1 ur1 srr2 stt2 ur2 srt1 ut1 srt2 ut2;
   srtsq=srt2*srt2;                                                    10
   srrsq=srr2*srr2;
run;


* proc corr data=par00 ;
*    var t--srrsq;
* run;
*    title3 'Correlation Analysis';

                                                                       20
* proc stepwise data=par00 ;
*    title1 'Stepwise - Linear';
*    model t= srr1      srr2     srt1 srt2 /stepwise sle=.05 sls=.05;
* run;


* proc stepwise data=par00 ;
*    title1 'Stepwise - Nonlinear';
*    model t= srr1 srtsq srr2 srrsq srt1 srt2 /stepwise sle=.05 sls=.05;
* run;                                                                 30


* proc reg     data=par00 ;
```

```
*    title1  'REG, with stepwise & l95/u95';
*    model t= srr1 srtsq srr2 srrsq srt1 srt2 /selection=stepwise sle=.05 sls=.05;
*    output out= guess1 predicted=  predt1 l95 = low195  u95 = up195;
* run;


 proc glm      data=par00 ;                                                    40
    title1  'GLM linear,  cli alpha';
    model t= srr1       srr2       srt1 srt2 /nouni; * /cli   alpha=.10;
    output out= guess1 predicted = predt1  l95 = low195  u95 = up195;
 run;


 proc glm      data=par00;
    title1  'GLM nonlinear, cli alpha';
    model t= srr1 srtsq srr2 srrsq srt1 srt2  /nouni; * /cli   alpha=.10;
    output out= guess2 predicted = predt2  l95 = low295  u95 = up295;          50
*    press=names rstudent= names;
 run;


 title1 ;
 data   good1  bad1;
    set guess1;
    keep   t   predt1  low195  up195   zz;
      zz = .0001;
                                                                             60
      if (up195-low195) > zz  then  zz = (up195-low195);

      if     t  <=  up195  and      t  =>  low195 then  output good1;

      else   output  bad1;
 run;
```

```
data  good2  bad2;
  set guess2;                                                    70
  keep  t  predt2  low295  up295 zz;
    zz = .0001;

    if (up295−low295) > zz  then  zz = (up295−low295);

    if    t  <=  up295  and     t  =>  low295 then  output good2;

    else  output  bad2;
run;
                                                                 80

proc print data= good1;
run;

proc print data= bad1;
run;

proc print data= good2;
run;
                                                                 90
proc print data= bad2;
run;


* title1  'Plot analysis :::';
* title2  'Infile is paro.005';

* proc plot data=par00 ;
*   plot t*srr1;
*   plot t*srr2;                                                100
*   plot t*srt1;
*   plot t*srt2;
*   plot t*srtsq;
```

```
*     plot t*srrsq;
* run;
```
