# Induction of Multivariate Regression Trees for Design Optimization

**B. Forouraghi[1,2], L.W. Schmerr[1,3] and G.M. Prabhu[2]**
Center for NDE[1], Computer Science Department[2], and Aerospace/Engineering Mechanics Department[3]
Iowa State University
Ames, IA 50011
bforoura@cnde.iastate.edu, lschmerr@cnde.iastate.edu, prabhu@cs.iastate.edu

**Abstract**

In this paper we introduce a methodology within which multiobjective design optimization is approached from an entirely new perspective. Specifically, we demonstrate that multiple-objective optimization through induction of multivariate regression trees is a powerful alternative to the conventional vector optimization techniques. Furthermore, in an attempt to investigate the effect of various types of splitting rules on the overall performance of the optimizing system, we present a tree partitioning algorithm which utilizes a number of techniques derived from diverse fields of statistics and fuzzy logic. These include: three multivariate statistical approaches based on dispersion matrices, two newly-formulated fuzzy splitting rules based on Pearson's parametric and Kendall's nonparametric measures of association, Bellman and Zadeh's fuzzy decision-maximizing approach within an inductive framework, and finally, the multidimensional extension of a widely-used fuzzy entropy measure. In terms of potential application areas, we highlight the advantages of our methodology by presenting the problem of multiobjective design optimization of a beam structure.

## 1. Introduction

Most engineering design problems involve optimization of several noncommensurable objectives in presence of multiple constraints. For example, in a quality control application, the main goal may be to optimize the design of an electric discharge machining (EDM) process in which design variables such as pulse duration and discharge current directly determine multiple, conflicting responses such as electrode wear, surface roughness and metal removal rate (Osyczka 1984). Optimization of such design problems primarily involves the determination of Pareto-optimal solutions where an individual objective can be further improved only at the cost of degrading at least one other objective (Chankong & Haimes 1983). It must be emphasized that most conventional optimization methods available to date, such as response surface methodology (Snee 1985), vector optimization techniques (Chankong & Haimes 1983) and multivariate analysis of variance (Harris 1985), typically generate Pareto-optimal 'point' solutions for which a vector-valued objective function is optimized. In many robust design applications, however, either due to economical factors or processing limitations, it is desirable to provide the designer with a range of values or 'surfaces' for design parameters for which the individual variances among responses are minimal while response means are fixed on their optimal target values (Phadke 1989). Consequently, 'surface' Pareto-optimal solutions enable the designer to examine various design scenarios without gross departures from optimal response regions.

One of the most viable artificial intelligence approaches for generating 'surface' Pareto-optimal solutions which has not been investigated before is the symbolic search technique of regression trees. Traditionally, tree-structured approaches to regression such as classification and regression trees (CART) (Breiman et al. 1984) and inductive partitioning with regression trees (IPRT) (Shien & Joseph 1992) have emphasized univariate regression analysis. These algorithms are powerful in that not only do they perform ordinary regression, but they also represent complex regression surfaces in terms of a number of simpler regression subsurfaces. Detailed examination of these subsurfaces can therefore help identify design regions where a product or process response is optimized. Hence, in addition to pinpointing optimal response regions, tree-structured approaches to optimization offer the advantage of explicating the knowledge that actually constitutes the optimality of the generated solutions. For instance, a sequential examination of various leaves in an induced regression tree can potentially show how deviations from a particular region of interest in the design space affect the overall objective functions. Lack of this type of systematic analysis is obviously one of the shortcomings of most traditional approaches to optimization including the elliptical technique for design centering (Abdel-Malek & Hassan 1991).

To summarize, in this paper we present a new framework within which multiobjective optimization is accomplished through induction of multivariate regression trees. Furthermore, we present a tree partitioning algorithm which utilizes a number of splitting rules based on concepts from statistics and fuzzy logic. Obviously, the choice of using the traditional statistical formulations in this work was instigated by the historic fact that statistics is a firmly established science with many facets which render it a particularly viable tool in many scientific applications. The theory of fuzzy sets (Gui & Georing 1990), on the other hand, is a more recently developed concept, and it too has proven to be an invaluable tool in a wide array of applications ranging from pattern recognition and clustering to design of digital circuits and relational databases (Pal 1991). In fact, within the context of multiobjective optimization, Bellman and Zadeh's fuzzy approach to optimization (Bellman & Zadeh 1970) has been widely implemented in many engineering applications such as structural optimization (Rao 1987). Therefore, in an attempt to examine the effect of various types of tree partitioning rules on the overall learning process, and also, to assess the feasibility of techniques based on fuzzy logic we describe seven splitting rules. Specifically, these include: two statistical decision rules based on dispersion matrices, a statistical measure of covariance complexity which is typically used for obtaining multivariate linear models (Bozdogan 1990), two newly-formulated fuzzy partitioning methods based on Pearson's parametric (Harris 1985) and Kendall's nonparametric (Simon 1977) measures of association, Bellman and Zadeh's decision-maximizing fuzzy approach to optimization in an inductive framework, and finally, the multidimensional extension of a widely-used measure of fuzzy entropy (Kosko 1990).

The remainder of this paper is organized as follows. Section 2 describes our methodology for transforming the problem of multiobjective optimization into induction of multivariate regression trees using various splitting criteria. Section 3 presents key results of applying techniques described in this paper to a multiobjective design problem. And finally, Section 4 summarizes the paper.

## 2. Multivariate Regression Trees

The basic element for inducing a multivariate regression tree is a set of training examples which provides a capsule view into the objective/constraint space. These examples essentially enable the learning algorithm to incrementally construct a complex regression surface from a number of simpler regression subsurfaces. This piecewise model construction is accomplished in a top-down fashion by successive partitioning of the training population at each level of the tree in an attempt to identify compact clusters in the response region. Examination of these clusters in turn can identify location of the optimal solution where an objective can be further improved only by degrading one or more objectives (Chankong & Haimes 1983). The following provides more details regarding the tree induction process.

Basically, given a learning sample $L = (X_1,Y_1), (X_2,Y_2),..., (X_N,Y_N)$, the learning algorithm produces a prediction rule $d$ which is a mapping from the n-dimensional predictor or attribute space ($X_i$'s) to the p-dimensional response (objectives and constraints) region ($Y_i$'s). The learning sample, therefore, contains N examples where each example associates a p-dimensional response vector with an n-dimensional predictor vector. Initially, all N examples reside at the root of an empty tree. Following a divide-and-conquer approach, the root node is split into two left and right nodes such that $n_1$ of the original N examples fall in the left node and the remaining $n_2$ cases in the right node ($N = n_1 + n_2$). This splitting is facilitated by selection of an attribute and a threshold for partitioning the attribute's range into two regions (Fayyad & Irani 1992). Among all possible attribute/threshold pairs, the pair that results in the 'best' split, where the resulting left and right nodes maximize some measure of fitness, is selected and the node is split accordingly. The process of partitioning is then recursively applied to all newly generated nodes until some stopping criterion is met. In our case, a multivariate heuristic which dictates that the number of examples in a node has to be at least as large as the number of responses was used. Furthermore, after a tree is completely grown in the prescribed manner, some type of pruning will prove beneficial should the problem of overspecialization cause detrimental effects on overall efficiency of the learning system (Breiman et al. 1984).

After the learning phase is complete, the induced tree contains a number of paths which start from the root and end in a terminal node or leaf. Each path therefore pinpoints a regression subsurface by the virtue of examples that are contained in its leaf. A leaf's set of examples can be viewed as a cluster in the response region which is characterized by its mean vector $\mu$ and covariance $\Sigma$. The goodness of these clusters is in turn determined by a variety of statistical and fuzzy partitioning techniques which are explained below. In ensuing discussions assume that the response matrix at a given node is R (m by p matrix) which contains $m$ p-dimensional response vectors and that covariance of R is $\Sigma$. For fuzzy splitting criteria further assume that R is converted to the multidimensional fuzzy set M (m by p matrix) by fuzzifying individual responses $r_{ij}$ in R into $\mu_{ij}$ in M using (Sakawa 1983):

$\mu_{ij} = (r_{ij} - r_j^{min}) / (r_j^{max} - r_j^{min})$, for i=1,...,m; j=1,...,p

It must be mentioned that for a given response *j*, $r_j^{min}$ and $r_j^{max}$ are found by scanning rows (i=1,...,m) of R, and they can be interchanged depending upon whether the goal is to maximize or minimize the given response *j* in the fuzzy domain.

The first two nonfuzzy splitting criteria (Methods 1 & 2) used for tree induction are the trace and determinant of the covariance matrix which denote the sum of individual response variances and the generalized variance, respectively (Everitt 1974). Minimization of trace, which totally ignores the interaction among responses, attempts to locate spherically-shaped response clusters where individual variances are minimal. On the other hand, minimization of the generalized variance, $|\Sigma|$, helps identify parallelotopes formed by response vectors which have minimal volume (Tatsuoka 1971).

The third partitioning rule (Method 3) uses Bozdogan's information-theoretic *covariance complexity* measure which is typically used for selection and evaluation of multivariate models (Bozdogan 1990). Essentially, the covariance complexity metric measures how the individual subcomponents of a model or a system interact with one another. In the case of multivariate regression trees, we use a tree as a representative of an underlying model that is to be captured through the induction process. At each level of partitioning, a given node's original population of responses R is divided into two subpopulations such that the covariance complexity of the resulting subpopulations are minimal. The overall task hence is to evaluate the degree of interaction that exists between responses in R and select partitions which result in minimal entropy or disorder. This can be accomplished by assigning the following covariance complexity measure to the covariance matrix of a population R:

$$CC(\Sigma) = 0.5 \cdot p \cdot \log_2 [\mathrm{trace}(\Sigma)/p] - 0.5 \cdot \log_2 |\Sigma|$$

where *p* is the number of responses (objectives and constraints). During the tree splitting process then a parent node is split into two nodes such that the measures of covariance complexity of the newly generated nodes are minimal.

The next two partitioning criteria (Methods 4 and 5) are based on Pearson's parametric (Harris 1985) and Kendall's nonparametric (Simon 1977) measures of association $\rho$ and $\tau$, respectively. The main motivation here is to discover the degrees of relationship between two responses R and S which may involve linear or nonlinear components. It must be emphasized that Pearson's $\rho$ is particularly suitable for situations where responses exhibit linear relationship. However, in many situations, linear approximations may become extremely misleading when the relationships involve nonlinear components. To this end, Kendall developed the correlation measure $\tau$ which is not based on any parametric assumptions and is more likely to discover monotonic behavior between responses.

More formally, given the data $(R_1,S_1),\ldots,(R_N,S_N)$, Pearson's degree of linear relationship $\rho$ between responses R and S is:

$$\rho_{RS} = \Sigma_{i=1,\ldots,N} [(R_i - R^{mean})(S_i - S^{mean})] / \sigma_R \cdot \sigma_S$$

where $\sigma_R$ and $\sigma_S$ are the standard deviations of R and S, respectively. Also, Kendall's degree of monotonic relationship between R and S is:

$$\tau_{RS} = [2/N(N-1)] \, \Sigma_{i<j} \Sigma \, [\mathrm{sign}(R_i - R_j) \cdot \mathrm{sign}(S_i - S_j)]$$

where the sign function takes values +1, 0 or -1 depending upon whether its argument is positive, zero or negative. For the sake of simplicity, the following generically refers to $\rho$ and $\tau$ as $\chi$ since the forthcoming analysis is symmetric with respect to both of these measures.

The measure of association $\chi$ attempts to discover the relationship between any two given responses. For example, if R and S tend to grow in a similar direction, $\chi_{RS}$ approaches 1. Conversely, if $\chi_{RS}$ approaches -1, it is concluded that R and S grow in opposite directions. Furthermore, $\chi_{RS}$ values near 0 imply absence of any relationship (linear in the case of $\rho$ and monotonic in the case of $\tau$) between the two responses. Considering this, we can now incorporate elements from fuzzy logic as follows. Assume that a particular node's set M contains fuzzified responses as explained previously. Now, regardless of whether any individual response is to be maximized or minimized, the chief goal in the fuzzy domain is to locate regions where fuzzy responses approach their maximum values. Hence, given M, we obtain the matrix of correlation coefficients T (p by p matrix) where each $\chi_{ij}$ for responses *i* and *j* (i, j = 1,...,p) is computed using Pearson's or Kendall's measures ($\chi_{ii} = 1$, $\chi_{i>j} = \chi_{i<j}$). Note that since T is symmetric, only its above-diagonal elements, $\chi_{i<j}$, are considered for further calculations. These p(p-1)/2 elements, which are pairwise measures of association between fuzzy responses in M, take values between -1 and 1. However, the desired clusters to be found are those for which as many of these correlation values approach 1 as possible which simply means that all or most of the responses are approaching their expected extrema in a given region. To accomplish this, T's above-diagonal $\chi_{ij}$ correlation coefficients are fuzzified using either linear or exponential membership function transformations. The aspiration levels of -1 and 1 are used in the fuzzification process to indicate that correlation values of 1 are desirable to attain maximum degree of belongingness. The cluster under consideration is then

assigned the degree of trend fitness:

$$TF(M) = [\min_k \{\mu(\chi_k)\}], \text{ for } k=1,\ldots, p(p-1)/2$$

The overall objective, therefore, is to identify splits for which the produced clusters have maximal TF measures.

The sixth splitting criterion (Method 6) is based on Bellman and Zadeh's approach to multiobjective optimization (Bellman & Zadeh 1970). To give a brief overview, consider making a decision D which can be seen as a confluence of *n* objectives and constraints denoted by responses $R_1,\ldots, R_n$. The optimal decision in the fuzzy domain then can simply be viewed as the intersection of fuzzy sets $\mu(R_1),\ldots,\mu(R_n)$ where each $\mu(R_i)$ is calculated using appropriate membership function transformations. More formally, the optimization task can be formulated as finding an optimum predictor vector X* for which the measure $\mu_D(X^*) = \min_i \{\mu(R_i(X))\}$ is maximized. Typically, after proper transformation of the problem at hand into the fuzzy domain, X* is found using nonlinear programming (Rao 1987). In our framework, however, Bellman-Zadeh's approach is used for splitting a node such that the measure:

$$BZ(M) = [\max_i \min_j \{\mu_{ij}\}], \quad \text{for } i=1,\ldots,m \text{ and } j=1,\ldots,p$$

is maximized for a particular multidimensional fuzzy set M under consideration.

And Finally, the last inductive partitioning technique (Method 7) to be discussed relies on fuzzy entropy (Kosko 1990). Basically, given a fuzzy set A with its complement $A^c$, fuzzy entropy of A:

$$FE(A) = C^o(A, A^c) / C^u(A, A^c)$$

measures how fuzzy actually A is, where $C^o$ and $C^u$ denote counts of overlap and underlap between A and $A^c$, respectively. In a top-down inductive approach, the fuzzy entropy measure can be used to identify fuzzy clusters M which exhibit minimal amount of fuzziness at each partitioning level. The basic definition of entropy, however, has to be extended so that fuzziness of the multidimensional fuzzy set M can be calculated. To accomplish this, first, M's complement, $M^c$, is calculated where each $\mu_{ij}^c$ in $M^c$ is complement of $\mu_{ij}$ in M. Then, fuzzy sets I and U (both m by p matrices), which denote the intersection and union of M and $M^c$, are calculated where elements $i_{ij}$ in I and $u_{ij}$ in U are $\min(\mu_{ij}, \mu^c_{ij})$ and $\max(\mu_{ij}, \mu^c_{ij})$, respectively. Consequently, we define the fuzzy entropy measure of a multidimensional fuzzy set M as:

$$FE(M) = [\max_i \min_j \{i_{ij}\}] / [\max_i \min_j \{u_{ij}\}], \quad \text{for } i=1,\ldots,m \text{ and } j = 1,\ldots,p$$

During the course of tree induction, then, the attribute/threshold pair for which the resulting clusters have minimal fuzzy entropy are selected and the node is split accordingly.

## 3. Optimization of a Beam Structure

In this section we present the problem of multiobjective optimum design of a beam structure. Given the design variables $X_1$ and $X_2$ which respectively represent the length of the part 1 of the beam and the interior diameter of the beam, the design task involves minimization of the three objectives of beam volume ($F_1$), static compliance of the beam ($F_2$) and the bending stress of the beam ($\sigma_g$), respectively (for more details see Osyczka 1984).

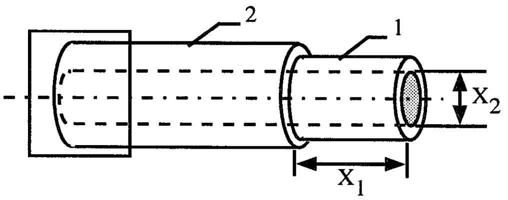


Figure 1. The beam structure

In regards to preparation of learning and testing cases, it was decided to uniformly sample the design region ($X_1$, $X_2$) in 900 distinct points between (10,40) and (300,75.2) in order to ensure that the response surfaces were adequately represented to the learning algorithm. Following a widely-used variance-stabilizing technique, the objective and constraint responses were transformed into log domain (Hahn 1971) which helped the overall learning efficiency for parametric as well as nonparametric induction criteria. The 900 points were then randomly shuffled and divided into two sets of size 450 each, namely, $T_{450}$ and $L_{450}$. The set $T_{450}$ was dedicated entirely to testing purposes while $L_{450}$ was used for the learning process. Samples of size 100, 150 and 200 were then randomly drawn from the overall learning set $L_{450}$. The learning phase then proceeded by inducing a regression tree on each of the randomly drawn samples $L_{100}$, $L_{150}$ and $L_{200}$ for each of the fuzzy and nonfuzzy splitting methods. This entire process of random selection of training samples, learning and testing was repeated a total of five times for each tree-growing technique so that results could be represented with 95% confidence

After completion of the learning phase, relative regression errors which normally vary between 0.0 (perfect regression model) and 1.0 (poor model) were computed for each induced tree. Our error analysis is similar to CART's (Breiman et al. 1984) except that it was extended for multivariate cases by substituting Mahalanobis distances (Everitt 1974) for ordinary Euclidean-based error distances in order to account for covariances that exist

| Method | $L_{100}$ | $L_{150}$ | $L_{200}$ |
|---|---|---|---|
| 1 | 0.0060 ± 0.0015 | 0.0044 ± 0.0014 | 0.0036 ± 0.0018 |
| 2 | 0.0076 ± 0.0031 | 0.0050 ± 0.0019 | 0.0044 ± 0.0022 |
| 3 | 0.0068 ± 0.0010 | 0.0050 ± 0.0019 | 0.0026 ± 0.0006 |
| 4 | 0.0046 ± 0.0020 | 0.0040 ± 0.0008 | 0.0040 ± 0.0042 |
| 5 | 0.0044 ± 0.0011 | 0.0038 ± 0.0018 | 0.0022 ± 0.0005 |
| 6 | 0.0050 ± 0.0012 | 0.0040 ± 0.0029 | 0.0026 ± 0.0014 |
| 7 | 0.0042 ± 0.0016 | 0.0034 ± 0.0018 | 0.0098 ± 0.0209 |

Table 1. Relative regression errors with 95% confidence

among responses. Table 1 summarizes relative regression errors for tree-partitioning methods 1 through 7.

The performance measures shown in Table 1 reveal two important facts. First, the overall inductive generalization power, and consequently, the regression accuracy of all partitioning techniques generally improves as the size of training sets increases from 100 to 200. And second, the accuracy of regression surfaces obtained through the use of fuzzy splitting criteria matches, and in few instances surpasses, the accuracy of solutions generated by well-established statistical techniques.

Relative regression errors are a good indicator of how an induced tree generalizes given a learning sample and a testing sample. The litmus test, however, lies in detailed examination of non-inferior solutions arrived at by a regression tree. These optimal solutions are represented by the terminal nodes of an induced tree and essential indicate 'tight' clusters in the response region. Table 2 summarizes some of the solutions generated by each induction method on training samples of size 200 which produced the most accurate results. Furthermore, in order to verify a tree's predicted range of responses for a specific range of design variables, we employed the following technique. The objective and constraint functions were evaluated for roughly about 5000 points in an induced tree's predicted optimum design region. Means and standard deviations of the generated responses were then computed to verify the tightness of clusters which were formed in the predicted response region. These verified solutions appear in the last column of Table 2.

A detailed examination of Table 2 reveals that in contrast to traditional multiobjective techniques which result in distinct Pareto-optimal point-solutions, our technique identifies Pareto-optimal regions. For example, in contrast to the two optimal solutions reported by Osyczka (marked "O") which were obtained using an ordinary vector optimization technique consider the solution which was jointly obtained by methods 2 and 3. This solution recommends the design region $(X_1,X_2)$ = ([44.7, 64.8], [65.3, 75.2]) where the designer can safely choose any values for design parameters $X_1$ and $X_2$ within the proposed bounds. The corresponding objective and constraint functions, as predicted by the learning algorithm,

| Method | Optimum Values of Design Variables $(X_1, X_2)$ | | Predicted Pareto-Optimal $(F_1$ $10^6$ | $F_2$ $10^{-3}$ | $\sigma_g)$ | Verified Pareto-Optimal $(F_1$ $10^6$ | $F_2$ $10^{-3}$ | $\sigma_g)$ |
|---|---|---|---|---|---|---|---|---|
| "O" | (237.0, 66.4) | | (3.7 | 0.425 | 107.7) | (3.7 | 0.425 | 107.7) |
| | (224.7, 58.6) | | (4.5 | 0.382 | 75.3) | (4.5 | 0.382 | 75.3) |
| 1 | [149.6, 176.0] | [68.3, 71.0] | (3.5 | 0.438 | 90.1) | (3.5 ± 0.09 | 0.438 ± 0.007 | 92.2 ± 7.2) |
| | [118.3, 149.6] | [68.3, 71.0] | (3.6 | 0.436 | 78.8) | (3.6 ± 0.09 | 0.435 ± 0.007 | 75.8 ± 7.0) |
| 2, 3 | [44.7, 64.80] | [65.3, 75.2] | (3.7 | 0.442 | 39.1) | (3.7 ± 0.30 | 0.441 ± 0.023 | 35.6 ± 10.5) |
| 4 | [10.0, 54.70] | [71.0, 73.7] | (3.6 | 0.455 | 18.6) | (3.6 ± 0.09 | 0.455 ± 0.007 | 23.6 ± 9.7) |
| | [118.3, 211.7] | [68.3, 71.0] | (3.5 | 0.439 | 93.9) | (3.5 ± 0.11 | 0.439 ± 0.008 | 93.4 ± 16.4) |
| 5, 6 | [124.8, 152.4] | [68.3, 71.0] | (3.6 | 0.436 | 78.0) | (3.6 ± 0.09 | 0.436 ± 0.007 | 78.5 ± 6.7) |
| | [47.4, 102.4] | [71.0, 73.7] | (3.4 | 0.456 | 60.7) | (3.5 ± 0.09 | 0.456 ± 0.007 | 54.8 ± 12.5) |
| 7 | [132.4, 159.9] | [68.3, 71.0] | (3.6 | 0.436 | 82.7) | (3.6 ± 0.09 | 0.437 ± 0.007 | 82.7 ± 6.9) |
| | [72.4, 132.4] | [68.3, 71.0] | (3.7 | 0.434 | 66.8) | (3.7 ± 0.10 | 0.434 ± 0.006 | 58.0 ± 10.4) |

Table 2. Optimal solutions produced by various fuzzy and nonfuzzy splitting criteria

take the values ($3.7\times10^6$, $0.442\times10^{-3}$, 39.1) while thorough examination of roughly 5000 points in this particular design region verifies that actual responses center around the values ($3.7\times10^6 \pm 0.30\times10^6$, $0.441\times10^{-3} \pm 0.023\times10^{-3}$, $35.6 \pm 10.5$) which are still well within optimal bounds. Osyczka's solutions, on the other hand, merely indicate that (237.0, 66.4) or (224.7, 58.6) are optimal design parameter values, and they are rigid in the sense that they fail to provide the designer with a range within which different design scenarios can be examined and subsequently realized without gross departures from optimal response regions.

## 4. Conclusions

In this paper we introduced a new methodology within which the problem of multiobjective optimization is transformed into induction of multivariate regression trees. Moreover, we demonstrated how the tree growing process can be accomplished by utilizing a number of concepts from diverse fields of statistics and fuzzy logic. In particular, seven splitting criteria were devised and implemented which include: three statistical methods based on dispersion matrices, two newly formulated fuzzy approaches based on Pearson's parametric and Kendall's nonparametric measures of association, Bellman-Zadeh's fuzzy approach to optimization in an inductive framework, and finally, the multidimensional extension of a fuzzy measure of entropy.

We also compared the overall performance of the learning system for the fuzzy and nonfuzzy methods. Our empirical results indicate that utilization of fuzzy splitting criteria offers a degree of flexibility in terms of the learning system's efficiency which traditional multivariate statistical methods lack. To illustrate this point, we presented the problem of multiobjective design of of a beam structure.

## Acknowledgments

This work was supported by the Center for NDE, Iowa State University, Ames, Iowa.

## References

Abdel-Malek, H.L., and Hassan, A.S. 1991. The Elliptical Technique for Design Centering and Region Approximation. *IEEE Transactions on Computer-Aided Design* (10)8:1006-1013.

Bellman, R.E., and Zadeh, L.A. 1970. Decision Making in a Fuzzy Environment. *Management Science* (17)4:141-163.

Bozdogan, H. 1990. On the Information-Based Measure of Covariance Complexity and Its Application to the Evaluation of Multivariate Linear Models. *Communication in Statistics--Theory & Methodology* (19)1:221-279.

Breiman, L.; Friedman, J.H.; Olshen, R.A.; and Stone, J.S. 1984. *Classification and Regression Trees*, The Wadsworth & Brooks/Cole Advanced Books & Software.

Chankong, V., and Haimes, Y.Y. 1983. Optimization-Based Methods for Multiobjective Decision-Making: An Overview. *Large Scale Systems: Theory and Applications* 5:1-33.

Everitt, B. 1974. *Cluster Analysis*. Wiley & Sons.

Fayyad, U.M., and Irani, K.B. 1992. On the handling of Continuous-Valued Attributes in Decision Tree Generation. *Machine Learning* (81)1: 87-102.

Gui, X.Q., and Georing, C.E. 1990. Introduction to Fuzzy Set Theory and Applications. *Transactions of ASAE* (33)1:306-313.

Hahn, G.J. 1971). How Abnormal Is Normality? *Journal of Quality Technology* (3)1:18-22.

Harris, R.J. 1985. *A Primer of Multivariate Statistics.* Academic Press, Inc.

Kosko, B. 1990. Fuzziness vs. Probability. *International Journal of General Systems* (17):211-240.

Osyczka, A. 1984. *Multicriterion Optimization in Engineering*. Ellis Harwood Limited.

Pal, S.K. 1991. Fuzzy Tools For the Management of Uncertainty in Pattern Recognition Image Analysis, Vision and Expert Systems. *International Journal of Systems Sciences* (22)3:511-549.

Phadke, M. 1989. *Quality Engineering Using Robust Design*. Prentice Hall.

Rao, S.S. 1987. Multi-Objective Optimization of Fuzzy Structural systems. *International Journal for Numerical Methods in Engineering* (24):1157-1171.

Sakawa, M. 1983. Interactive Fuzzy Decision Making for Multiobjective Linear Programming Problems and Its Applications. *IFAC Fuzzy Information* 295-300.

Shien, D.S., and Joseph, B. 1992. Exploratory Data Analysis Using Inductive Partitioning and Regression Trees. *Industrial Chemical Engineering Research* (31):1989-1998.

Simon, G. 1977. Multivariate Generalization of Kendall's Tau with Application to Data Reduction. *Journal of the American Statistical Association*, (72)358: 367-376.

Snee, R.D. 1985. Computer Aided Design of Experiments. *Journal of Quality Technology,* (17)4:222-235.

Tatsuoka, M.M. 1971. *Multivariate Analysis: Techniques for Educational and Psychological Research.* Wiley &Sons.