

The Pennsylvania State University
The Graduate School

ANALYSIS AND OPTIMIZATION OF PROFILE AND SHAPE
RESPONSE EXPERIMENTS

A Dissertation in
Industrial Engineering and Operations Research
by
Hussam Alshraideh

© 2011 Hussam Alshraideh

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2011

The dissertation of Hussam Alshraideh was reviewed and approved* by the following:

Enrique del Castillo
Distinguished Professor of Industrial Engineering and Statistics
Dissertation Advisor, Chair of Committee

M. Jeya Chandra
Professor of Industrial and Manufacturing Engineering

Paul Griffin
Head, Industrial and Manufacturing Engineering Department

Murali Haran
Professor of Statistics

*Signatures are on file in the Graduate School.

Abstract

An engineering process that exhibits a response in the form of a univariate (or one-dimensional) curve whenever new experimental conditions are tried is said to have a *profile*, or *functional* response. Likewise, a manufacturing process or engineering system where the response of interest is the geometry of a product or part is said to have a *shape* response. A shape response can relate to a planar (two-dimensional) geometrical feature, or to a three-dimensional one. The overall theme of this dissertation is the modeling and optimization of engineering processes that have either a profile or a shape response. The models and methods described in this dissertation have application mainly in manufacturing, engineering design, and computer experiments.

Statistical Shape Analysis (SSA) is a relatively new area within Statistics. Traditionally the realm of biological applications, it has been recently applied to manufacturing problems. D.G. Kendall, in pioneering work conducted in the 1980's, defined the shape of an object as the geometrical information that remains once certain similarity transformations, namely, rotations excluding reflections, translations, and dilatations (or dilations) are filtered out. His work is based on a landmark representation of an object, where a landmark consists of the coordinates of a point measured on the object together with a label, with labels that correspond from object to object. This representation turns out to be relevant in manufacturing, since data obtained using a coordinate measuring machine will typically have this appearance. Over the last 20 years, several SSA tests have been proposed to detect differences in the mean shape between objects, but little work exists on the relative merits of these methods. The first part of this dissertation consists of a comprehensive performance analysis of landmark-based tests for mean shape differences. Since the performance of these tests depends on the types of shapes being tested, we consider both shapes that have been studied in the

scarce extant literature on the subject, namely triangles and arbitrary polygons with few landmarks, and also consider shapes of specific interest in manufacturing applications, such as circular and cylindrical geometries with tens to hundreds of landmarks.

An additional problem studied in this dissertation is that of shape optimization, that is, find the best operating conditions that lead to the most desirable shape of the product under fabrication. Previous tests for shape differences are based on Kendall’s definition of shape, which neglects differences in size between objects since it removes dilation (scale) effects, and make up for this deficiency by testing separately for differences in size. As an alternative, we present statistical tests for differences in *form* between the objects, where we define the form of an object as the geometrical information that remains once the effect of rotations and translations, but not dilations, is filtered out. We further develop a form optimization method when noise factors are present, proposing in effect a method for the Robust Parameter Design problem for shape (form) responses. Noise factors are factors that for the purpose of a carefully designed experiment are controllable, but that during normal operation of a production process or during use of a product vary randomly. The goal is to find the controllable factor conditions of the process that achieve a desired part form in the presence of noise factor variability.

The second part of this dissertation deals with profile response processes, their modeling, and subsequent optimization. Methods for this type of processes are mainly based on frequentist model estimation techniques, where the uncertainty in the parameter estimates is not considered during the optimization phase. As shown by J. Peterson, neglecting the uncertainty in the parameter estimates may lead to solutions that will very unlikely achieve the desired process performance. While there exist recent work in profile response systems where a Bayesian point of view is taken for model fitting that does incorporate the uncertainty of the model parameters into the subsequent optimization phase, those models are not flexible enough as they depend on a parametric regression model that is required to fit the mean profile well. As a more flexible alternative to these prior approaches, we present new modeling and optimization methodology for profile response processes based on a spatio-temporal Gaussian Random Function (GRF) model. In this model, the space of the controllable factors corresponds to the “space” dimension, and the space of the locations over which the profile responses are observed corresponds to the “temporal” dimension. The temporal dimension may or may not be actually time in some applications since in general it is equivalent to the “signal” in Taguchi’s signal response models. Similarly as in the first part of this dissertation, the goal here is to find controllable factor operating conditions that will lead to a desirable profile response with highest probability in the presence of noise factor

variability. The approach is fully Bayesian, incorporating the uncertainty of all process parameters present in the model which leads to more reliable predictive posterior probabilities to conformance to specifications for a given optimal solution. A discussion of robustness to the underlying assumptions and tools for checking model assumptions are provided. An adaptive Markov Chain Monte Carlo method for the fitting of the GRF model is presented that shows good convergence behavior.

Contents

List of Figures	ix
List of Tables	xvii
List of Acronyms	xix
Acknowledgments	xx
Chapter 1	
Introduction	1
1.1 Robust Parameter Design	5
1.2 Spatio-Temporal Models	7
1.3 Statistical Shape Analysis of Manufacturing Data	10
1.4 Bayesian Analysis	12
1.5 Dissertation Objectives	15
1.6 Dissertation Outline	16
Chapter 2	
Overview of Previous and Related Research	18
2.1 Statistical Shape Analysis of Manufacturing Data	18
2.1.1 Pre-shape and Shape Space	21
2.1.2 Distances Used in Shape Analysis	23
2.1.3 Principle Component Analysis of Shapes	27
2.1.4 Statistical Inference on Shapes	27
2.2 Modeling and Robust Parameter Optimization of Profile Response Experiments	34

Chapter 3	
Statistical Performance of Tests for Mean Shape Difference with Application in Manufacturing	38
3.1 Performance Analysis of Tests for Non-manufacturing Shapes of Arbitrary Geometry, Isotropic Errors	39
3.1.1 Triangles	39
3.1.2 Performance of Tests for Arbitrary Polygons (Rohlf, 2000; Coward and McConathy, 1996)	44
3.1.3 The Assumption of a Small Error Variance (σ^2) in the F-test	47
3.2 Performance Analysis for Shapes with a Geometry of Interest in Manufacturing, Isotropic Errors	50
3.2.1 Detection of Differences in Mean for 2D Circular Parts . . .	50
3.2.2 Detection of Differences in Mean Shape for 3D Cylindrical Parts	52
3.2.3 Two Way ANOVA Tests for Shapes; Test for Interaction Effect	54
3.3 Performance of Tests for Difference in Shape Under Non-isotropic Variance (2D and 3D)	55
3.4 Performance of Tests Under Non-normal Errors	62
3.5 Conclusions and Summary of Results of This Chapter	64
Chapter 4	
Robust Parameter Optimization of Shape Responses	70
4.1 Shape Optimization	70
4.2 Real Manufacturing Process Examples	77
4.2.1 Titanium Lathe-turning	77
4.2.2 Aluminum Cylinders Lathe-turning	80
4.3 Summary and Conclusions to This Chapter	81
Chapter 5	
Spatio-Temporal Modeling and Optimization of Profile Response Experiments	85
5.1 A Spatio-Temporal Gaussian Random Function Process Model . . .	85
5.2 Model Robustness and Validation	97
5.2.1 Robustness to Normality Assumption	99
5.2.2 Robustness with Respect to Separability and Mean Structure Misspecification	101
5.2.3 Checking the Covariance Separability Assumption	105
5.3 Conclusions and Future Work of This Chapter	112

Chapter 6	
Research Contributions and Areas for Further Research	115
6.1 Research Contributions	115
6.1.1 Statistical Performance of Shape Analysis Tests	116
6.1.2 Robust Parameter Optimization of Shape Responses	118
6.1.3 Robust Parameter Optimization of Profile Responses	118
6.2 Areas for Further Research	119
Appendix A	
Mathematical Notions Used in the Theory of Statistical Shape	
Analysis	123
A.1 Relations, equivalence relations and equivalence classes	123
A.2 Groups and transformations	126
A.3 Projective geometry and Complex projective space	127
A.3.1 Manifolds, tangent space, submersions and immersions	129
A.4 Intrinsic and extrinsic geometry and geodesics	133
A.5 Kendall's Preshape and Shape Spaces	134
Appendix B	
The Landmark Matching Problem	138
Appendix C	
Kronecker Product Decomposition of Matrices	143
Appendix D	
Adaptive MCMC Sampling Algorithm	147
Appendix E	
Computer implementation	151
E.1 Robust parameter optimization of shape responses	151
E.2 Robust parameter optimization of profile responses	152
Bibliography	155

List of Figures

1.1	Observed profiles at each of the 18 considered control factor combinations (H=high, M=medium, L=low). The specification limits L and U are shown in dashed lines.	3
1.2	An example of a 2D shape response experiment.	3
1.3	Prototypical RPD problem showing the inputs (controllable and noise factors) and the corresponding outputs of a process that acts on certain experimental units (i.e., parts in a manufacturing application). This figure is adapted from Del Castillo (2007).	7
1.4	Illustration of three realizations of a “temporal” process (e.g., observed elastic modulus profiles) at four different controllable factor combinations, i.e., four locations in the “ x -space”.	9
1.5	Circularity form error calculation.	11
2.1	A three landmark object (i.e., a triangle) represented in the Cartesian coordinate system (left plot) and its preserved shape as presented in the Bookstein coordinate system (right plot).	19
2.2	Distances between two shapes in preshape space. ρ is the Procrustes distance along the pre-shape surface, d_F is the full Procrustes distance along the tangent plane, and d_p is the partial Procrustes distance (adopted from Del Castillo and Colosimo (2011)).	24
3.1	Kendall’s shape space for triangles in $m = 2$ dimensions. This is $S^2(\frac{1}{2})$, a 2-sphere with radius $\frac{1}{2}$. The shape coordinates are the latitude θ and the longitude ϕ in equations (3.4-3.6). Adapted from Dryden and Mardia (1998).	41
3.2	A representation of different equilateral triangles projected on the tangent plane shown in Figure 3.1 as a function of the rotational angles ξ and ψ , given by equations (3.7 and 3.8).	42

3.3	Power surface for detecting differences in shape between triangles using the ANOVA Permutation test and Goodall’s ANOVA F-test, isotropic variance. The simulation parameters were: error standard deviation σ was set at 0.15, the test size α was set at 0.1, twenty ($= n$) configurations per sample were used, each test was replicated one hundred times, and one hundred permutations were performed. θ is restricted to $[-\pi/2, \pi/2]$ to exclude reflections. On the right, the marginal power plots for fixed ϕ	43
3.4	(a) The two mean shapes of the 2D data set used by Rohlf (2000). (b) The two mean shapes of the 3D data set used by Rohlf (2000). (c) The two mean shapes of the 2D data set used by Coward and McConathy (1996).	45
3.5	Simulated type-I error rate of the ANOVA Permutation test using the data set in Coward and McConathy (1996).	47
3.6	The distribution of the scaled G statistic as σ changes.	48
3.7	The two mean shapes used in the ANOVA identity calculations simulation study.	49
3.8	Circular and bilobed profiles with parameters $\sigma = 0.05$ and $r = 5$	51
3.9	Power curves using 2D Circular profiles.	52
3.10	Sketch of two different mean shapes of cylindrical parts with $r = 5$ and $h = 10$. Left: perfect cylinder ($\delta = 0$), right: “barrel” ($\delta = 1$).	53
3.11	Statistical power for detecting mean shape difference in cylinders for the F-test, the Permutation ANOVA test, the ANOVA form error method, the EDMA-I and the EDMA-II tests. Simulation parameters were: 320 landmarks ($= k$) per configuration, test size α of 0.05, error standard deviation σ of 0.05, each test was replicated one thousand times, and 150 permutations for each test requires permutation.	54
3.12	Power results for detecting an interaction effect in a two-way shape ANOVA for circular data.	56
3.13	Mean shape estimated by the iterative GPA(Σ) algorithm applied to 2D circular shapes with non-isotropic error structure (model 1 in equation (3.12) with $g = 2$ groups were used). Left: before GPA(Σ) alignment, right: after GPA(Σ) alignment.	58

- 3.14 Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under non-isotropic error structure described by equation (3.12). Simulation parameters were: 64 landmarks ($= k$) per configuration, test size α of 0.05, parameters $\{C_{low}, C_{med}, C_{high}\}$ were set to 0.001, 0.005 and 0.01 respectively, each test was replicated one hundred times, and 100 permutations for each test that requires permutations were performed. 59
- 3.15 Statistical power for detecting changes in 3D cylindrical shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under non-isotropic error structure described by equation (3.12). Simulation parameters were: 320 landmarks ($= k$) per configuration, test size α of 0.05, parameters $\{C_{low}, C_{med}, C_{high}\}$ were set to 0.001, 0.005 and 0.01 respectively, each test was replicated one hundred times, and 100 permutations for each test that requires permutations were performed. 60
- 3.16 Statistical power for detecting changes in 2D circular shapes for the F-test, the Permutation ANOVA test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential isotropic errors as described by equation (3.13). Simulation parameters were: 64 landmarks ($= k$) per configuration, test size α of 0.05, 10 configurations ($= n$) per sample, the parameters $\{\tau, \sigma, \phi, r\}$ were set to 0.05, 0.05, 0.05 and 5 respectively, each test was replicated one hundred times, and 100 permutations for tests that require them were performed. 61
- 3.17 Statistical power for detecting changes in 3D cylindrical shapes for the F-test, the Permutation ANOVA test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential isotropic errors as described by equation (3.13). Simulation parameters were: 320 landmarks ($= k$) per configuration, test size α of 0.05, 10 configurations ($= n$) per sample, the parameters $\{\tau, \sigma, \phi, r, h\}$ were set to 0.05, 0.05, 0.05, 5 and 10 respectively, each test was replicated one hundred times, and 100 permutations for tests that require them were performed. 62

3.18	Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential non-isotropic errors as described by equation (3.14). Simulation parameters were: 64 landmarks ($= k$) per configuration, 10 configurations ($= n$) per sample, test size α of 0.05, the parameter $\{\tau, \sigma, \phi, r\}$ were set to 0.05, 0.05, 0.05 and 5 respectively, the matrix \mathbf{B} was set to (1 0.5; 0.5 1.5), each test was replicated one hundred times, and 100 permutations were performed for each test that requires permutation.	63
3.19	Statistical power for detecting changes in 3D cylindrical shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential non-isotropic errors as described by equation (3.14). Simulation parameters were: 320 landmarks ($= k$) per configuration, 10 configurations ($= n$) per sample, test size α of 0.05, the parameter $\{\tau, \sigma, \phi, r, h\}$ were set to 0.05, 0.05, 0.05, 5 and 10 respectively, the matrix \mathbf{B} was set to (1 0.5 0.5; 0.5 1.5 0.5; 0.5 0.5 2), each test was replicated one hundred times, and 100 permutations were performed for each test that requires permutation.	64
3.20	Effect of increasing the number of landmarks: Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test and the ANOVA on form errors under exponential isotropic errors as described by equation (3.13). Simulation parameters were: 320 landmarks ($= k$) per configuration, 10 configurations ($= n$) per sample, test size α of 0.05, the parameter $\{\tau, \sigma, \phi, r\}$ were set to 0.05, 0.05, 0.05 and 5 respectively, each test was replicated one hundred times, and 100 permutations were performed for each test that requires permutation (compare to Figure 3.16).	65
3.21	The fitted kernel density functions from 100,000 simulated circularity form errors ($k = 64$, isotropic errors).	65
3.22	Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test and the ANOVA form error method under Uniform errors. Simulation parameters were: 320 landmarks ($= k$) per configuration, 20 configurations ($= n$) per sample, test size α of 0.05, the parameters $\{a, b\}$ were set to -0.0866 and 0.0866 respectively, each test was replicated 100 times, and 100 permutations were performed for each test that requires them. . . .	66

3.23	Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test and the ANOVA form error method under error terms simulated from $t_5(0, 0.05^2)$. Simulation parameters were: 320 landmarks ($= k$) per configuration, 20 configurations ($= n$) per sample, test size α of 0.05, each test was replicated 100 times, and 100 permutations were performed for each test that requires them.	66
4.1	Target part geometry of example 2.	74
4.2	Observed configurations of example 2 (low=-1 and high=1).	74
4.3	Mean part geometry at each treatment of control and noise factors of example 2 (low=-1 and high=1).	75
4.4	Mean part geometry (blue) overlayed by the target geometry (red) at all control factor combinations.	76
4.5	Observed configurations by ignoring noise factors ($low = -1, high = 1$).	76
4.6	The desired geometry of the final Titanium specimen obtained by lathe-turning.	78
4.7	Observed shape data of the Titanium lathe-turning experiment.	79
4.8	Observed shape data of the Titanium lathe-turning experiment assuming factor B is noise factor.	80
4.9	Observed shape data of the Aluminum lathe-turning experiment.	82
4.10	Observed shape data of the Aluminum lathe-turning experiment considering control factors only.	83
5.1	A schematic representation of the spatio-temporal data structure. Points \mathbf{x}_1 and \mathbf{x}_2 are close in the x -space and hence they are expected to generate similar profiles, \mathbf{y}_1 and \mathbf{y}_2 . In contrast, a point \mathbf{x}_3 in x -space that is farther away from \mathbf{x}_1 and \mathbf{x}_2 is expected to generate a profile response \mathbf{y}_3 that is not that correlated with \mathbf{y}_1 and \mathbf{y}_2	89
5.2	Observed elastic modulus profiles at each of the 14 control factor treatments in the metal injection moulding process. The lower and upper specification limits are shown by dashed lines.	92

5.3	MCMC convergence plots, metal injection example. From top to bottom: posterior densities, trace plots, autocorrelation functions (ACF), expected value plots, MCMC standard error of the estimated posterior mean, and posterior variance plots of the generated covariance parameters $(\phi_x, \psi_x/\kappa, \phi_x)$ (one column of plots per parameter). In here, we show the plots of the ratio ψ_x/κ since it is known in the Spatial Statistics literature (see e.g., Banerjee et al., 2004) that the ratio of these two parameters converges to its true distribution but not each of the two parameters separately.	94
5.4	Metal injection example: (a) Probability of conformance to specifications as a function of design factors. (b) Predicted profiles at the optimal design factor settings. The mean of the posterior predictive density is shown with thick solid line. The other two solid lines are the 10 th and 90 th percentiles of the posterior predictive density. Upper and lower specification limits are shown in dashed lines.	95
5.5	MCMC convergence plots, electric alternator example. From top to bottom: posterior densities, trace plots, autocorrelation functions (ACF), expected value plots, MCMC standard error plots and variance plots of the generated covariance parameters $(\phi_x, \psi_x/\kappa, \phi_x)$ (one column of plots per parameter).	95
5.6	Predicted profiles at the optimal controllable factor settings for the electric alternator example. The mean of the posterior predictive density is shown in solid thick line. The other two solid lines are the 10 th and the 90 th percentiles of the posterior predictive distribution. The upper and lower specification limits are shown by dashed lines.	96
5.7	Observed profiles at each of the 16 design factor treatments in the plastic injection process example. The specifications $L(s_j)$ and $U(s_j)$ are shown by dashed lines.	98
5.8	Predicted profiles at the optimal controllable factor settings. The mean of the posterior predictive density is shown in solid thick line. The other two solid lines are the 10 th and 90 th percentiles of the posterior predictive density. The upper and lower specification limits are shown in dashed lines.	99
5.9	Observed vs predicted profiles at (a) $\sigma_e = 1.5$ and $\epsilon \sim N(0, 1)$, (b) $\sigma_e = 1.5$ and $\epsilon \sim t_5$ used in the normality robustness analysis. Solid lines represent the observed profiles used for model fitting, the mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines.	102
5.10	Normal probability plots for the different error distributions considered in the normality robustness analysis.	103

5.11	Checking the separability assumption, metal injection example: plots of \mathbf{Y}^* and \mathbf{Y}_{approx}^* for (a) intercept mean structure (b) additive mean structure and (c) interaction mean structure. The first row of the observed and approximated data matrix is shown with thick line.	108
5.12	Leave-one-out cross validation results for the metal injection example. Solid lines are the observed profiles used for testing. The mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines. Typically, there would be 24 plots since there are 24 observed profiles, but we only show 12 of them here.	109
5.13	An instance of leave-10%-out cross validation results (10% ≈ 3 out of 24 observed profiles) for the metal injection example. Solid lines are the observed profiles used for testing. The mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines.	109
5.14	Leave-one-out cross validation results for the electric alternator example. Solid lines are the observed profiles used for testing. The mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines.	110
5.15	An instance of leave-10%-out cross validation results for the electric alternator example. Solid lines are the observed profiles used for testing. The mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines.	111
5.16	Leave-one-out cross validation results for the plastic injection process example. Solid lines are the observed profiles used for testing. The mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines.	112
5.17	An instance of leave-10%-out cross validation results for the plastic injection process example. Solid lines are the observed profiles used for testing. The mean, 10 th and 90 th percentiles of the posterior predictive density are shown in dashed lines.	113

A.1	One of the simplest illustrations of preshape and shape space. A) two lines in the original 2-dimensional space; B), preshapes on 2-dimensional Euclidean space, after centering and scaling; C) the corresponding pre-shape space is the (one-dimensional) circumference of a unit circle. The two pre-shapes lie on the single fiber or orbit generated as the preshapes are rotated, hence there is a single shape; D) the shape space for the two lines (Σ_2^2) is zero dimensional (a single point) and corresponds to the only shape that exists in this example.	137
B.1	Two handwritten digit 3's, each with $k = 13$ landmarks. Note that the labels of the landmarks do not correspond between the two shapes.	141
B.2	Matched shapes of Figure B.1 using the proposed matching algorithm. Matched landmark labels are shown in Table B.1.	142
E.1	(a) The programmed SSA graphical user interface. (b) Mean shape estimation graphical user interface. (c) ANOVA of shape responses graphical user interface. (d) Geometric shape RPD graphical user interface.	153
E.2	Model building and optimization of profile response systems MATLAB interface.	154

List of Tables

2.1	One-way ANOVA F test for difference in mean shape.	29
2.2	Two-way ANOVA table for differences in mean shape (Del Castillo and Colosimo, 2011).	30
3.1	Performance of tests for shape difference for the data set provided by Rohlf (Figure 3.4(a)). Results were obtained using a test size α of 0.05, one hundred configurations ($= n$) per sample. Each test was replicated one thousand times, and 150 permutations for the ANOVA Permutation and EDMA tests. Numbers in parentheses are the standard errors of the corresponding simulation estimates.	44
3.2	Performance of tests for shape difference for the data set of Table 3 in Rohlf (2000) (3-dimensional shape). Results were obtained using: the test size α was set at 0.05, one hundred configurations per sample ($= n$) were used. Each test was replicated one thousand times, and 150 permutations were performed for the ANOVA permutation test and the EDMA tests. Numbers in parentheses are the standard errors of the corresponding simulation estimates.	46
3.3	Sum of squares for the two-sample simulation study of the ANOVA identity. As σ increases, the identity breaks down.	49
3.4	Circularity parameter δ at the four possible combinations of A and B.	55
3.5	Means and standard deviations of the simulated form error density functions used in Figure 3.21.	64
4.1	The G_{Mean} values for data in example 2.	75
4.2	G_{Total} values obtained using (4.6) for example 2 data.	77
4.3	Factors and levels in the Titanium machining experiment.	78
4.4	Calculated G_{Total} values for the Titanium lathe-turning data.	79
4.5	Calculated G_{Total} values for the Titanium lathe-turning data assuming factor B is noise factor.	80
4.6	Factors and levels in the Aluminum machining experiment.	81

4.7	Calculated G_{Total} values for the Aluminum lathe-turning data with control factors only.	82
5.1	Specification limits at each of the 7 RPM locations for the electric alternator example.	93
5.2	Optimization results for the electric alternator example. Numbers in parentheses are the standard errors associated with probability estimation.	96
5.3	Control factors in the plastic injection moulding experiment by Wu and Hamada (2000).	97
5.4	Noise factors in the plastic injection moulding experiment by Wu and Hamada (2000).	98
5.5	Optimal solution and point wise probabilities of conformance for the plastic injection moulding data. Numbers in parentheses are the standard errors associated with probability estimation.	99
5.6	True vs. estimated optimal solutions for the 9 model combinations. Numbers in parentheses are the standard errors associated with the conformance probability estimation. The relative error is calculated as $e_x = \ \mathbf{x}_{true}^* - \mathbf{x}_{estimated}^*\ _F / \ \mathbf{x}_{true}^*\ _F$. In case the optimizer returns multiple optimal solutions, only one of them is listed.	104
5.7	True vs. estimated optimal solutions for the 9 model combinations each at 3 levels of the non-separability parameter γ . Numbers in parentheses are the standard errors associated with the conformance probability estimation. The relative error is calculated as $e_x = \ \mathbf{x}_{true}^* - \mathbf{x}_{estimated}^*\ _F / \ \mathbf{x}_{true}^*\ _F$. In case the optimizer returns multiple optimal solutions, only one of them is listed.	106
B.1	Matched landmark labels of Figure B.1	141

List of Acronyms

1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
BIC	Bayesian Information Criterion
CMM	Coordinate Measuring Machine
CNC	Computer Numerical Control
DOE	Design of Experiments
EDMA	Euclidean Distance Matrix Analysis
GP	Gaussian Process
GPA	Generalized Procrustes Analysis
GRF	Gaussian Random Function
MCMC	Markov Chain Monte Carlo
RPD	Robust Parameter Design
SSA	Statistical Shape Analysis
SPC	Statistical Process Control
WLS	Weighted Least Squares

Acknowledgments

“O my Lord! so order me that I may be grateful for Thy favors, which thou hast bestowed on me and on my parents, and that I may work the righteousness that will please Thee: And admit me, by Thy Grace, to the ranks of Thy righteous Servants.” Quraan [27:19]

The successful completion of this dissertation was made possible through the invaluable contribution of a number of people. To say “thank you” to all of you is not even enough to express my gratitude.

First of all, I thank ALLAH for giving me the intellect to understand the complexity of numbers, and giving me strength to complete this research. Most specially, thank you for blessing me with many great people who have been my greatest support in both my personal and professional life.

I would like to express my deep gratitude to my advisor Professor Enrique Del Castillo for his great patience, detailed instructions and insightful comments at every stage of this dissertation. What I have learned from him goes beyond the knowledge he taught me; his attitude toward research, carefulness and commitment has and will have a profound impact on my future academic activities. I would also like to thank Professors M. Jay Chandra, Paul Griffin and Murali Haran for serving as committee members.

I would like to thank my dad and my mom for all that they have done for me. Words cannot express my gratitude to them but time will prove the return of their immense love.

I would like to thank all my friends for the enthusiasm and inspiration, which was always there when I needed it. My special thanks go to my friends from the Islamic Society of Central Pennsylvania for their friendship which was more motivating in my work than any technical assistance they have contributed. Also, I would like to thank my colleagues from the Engineering Statistics Laboratory (ESL) at the Pennsylvania State University, particularly, Dr. Eduardo Santiago, Dr. Byran Smucker, Sam Tajbakhsh and Cedric de Toffol. I would like to thank many others whom I crossed paths with in the past three years I spent in State College.

I would like to thank Professors Bernadette Govaerts and Winson Taam for providing me with the data from their published papers. I would like to thank Mr. Daniel Supko for his help in running the experiments in the FAME lab and Mr. Terry Crust for training me to run the Coordinate Measuring Machine at the Precision and Quality Measurements Laboratory.

Last but not least, my special thanks to the department of Industrial Engineering at the Jordan University of Science and Technology for sponsoring my Ph.D study here at The Pennsylvania State University.

To all people that I know, thank you and God bless you all.

Dedication

To my parents...

Chapter 1

Introduction

In many areas of manufacturing or during the operation of numerous industrial processes, the response of interest in an experiment does not consist simply of a single observation measured at each of the experimental conditions. Instead, a continuous response variable is observed over a given space, and the goal is to model this response. The dimension of the space where the response lies can be one-, two-, or three-dimensional. The case of a one-dimensional (1D) response variable is known as a “profile response” in the Statistical Process Control (SPC) literature (Kim et al., 2003) or a “functional response” in Statistics (Ramsay and Silverman, 2005). Two- and three-dimensional (2D/3D) responses typically refer to the geometry or shape of manufactured parts. The overall goal of this dissertation is to study the analysis and optimization methods for profile and shape responses in manufacturing and engineering design experiments.

An example of a one-dimensional (1D) profile response experiment is given by Nair et al. (2002) who studied the design of an electric alternator. The response of interest in the experiment was the electric current generated at different rotational speeds (in RPMs) at which the alternator operates. The electric current was measured at seven RPM values $\{1375, 1500, 1750, 2000, 2500, 3500, 5000\}$. A designed experiment was run that consisted of 8 controllable factors ($x_1 =$ turns per coil, $x_2 =$ material thickness, $x_3 =$ power, $x_4 =$ width pitch ratio 1, $x_5 =$ width pitch ratio 2, $x_6 =$ diameter ratio, $x_7 =$ diameter length ratio, and $x_8 =$ outer diameter), all varied at three levels, except x_1 , which was only varied at two levels. The experiment also considered two noise factors: $x_9 =$ air gap variation

(two levels), and x_{10} = temperature (three levels). The experimental design used was a Taguchi L_{18} orthogonal array replicated six times (108 factor combinations in total). At each of the 108 factor combinations seven measurements were obtained, each at a given RPM value. Figure 1.1 shows the observed profiles at each of the 18 control factor combinations in the experiment. Based on the data from this experiment, the goal was to find the best controllable factor settings (and therefore, the best alternator design) that lead to a specified shape of the electric current profile with maximum probability. We note here that the data structure of this type of response function is similar to the spatio-temporal data structure found in the Spatial Statistics literature (Cressie, 1993). In a spatio-temporal model applied to the alternator design problem, the design factor space represents the “spatial” space and the measurement locations (RPM values in this example) represent the “time” dimension.

An example of a two-dimensional (2D) shape response was considered by Del Castillo and Colosimo (2011). They reported an experiment in lathe turning of Titanium alloy parts in which the depth of cut and cutting speed were varied according to a 3^2 factorial design. A set of 90 Titanium alloy (Ti-6Al-4V) specimens were machined by lathe-turning. Lathe-turning of the external surface of these specimens was performed for each of the 9 treatments each replicated 10 times. The 2D response of interest is made of the (x, y) -coordinates of the machined Titanium shaft cross-section measured over a series of locations on the surface of the manufactured part. Figure 1.2(a) shows the cross-section of interest in the experimental study, whereas Figure 1.2(b) shows an instance of the response of interest in this lathe-turning experiment measured over a set of 64 points along that cross-section. The goal of the experiment was to determine the effect of depth and cutting speed on the circularity of the parts and to determine the best settings of these factors to achieve the most circular parts.

Finally, a typical three-dimensional (3D) response experiment occurs when machining an engine piston to its final diameter after the forging process. Since pistons have to fit tightly in their cylinders their geometry is crucial. Piston geometry can be expressed by a set of three dimensional points (x, y, z) taken along the surface of the piston. Cutting speed, feed rate, material characteristics, etc., might have an effect on the geometry of the piston. An experiment can be conducted to test

the effect of these factors on the 3D geometry of the piston.

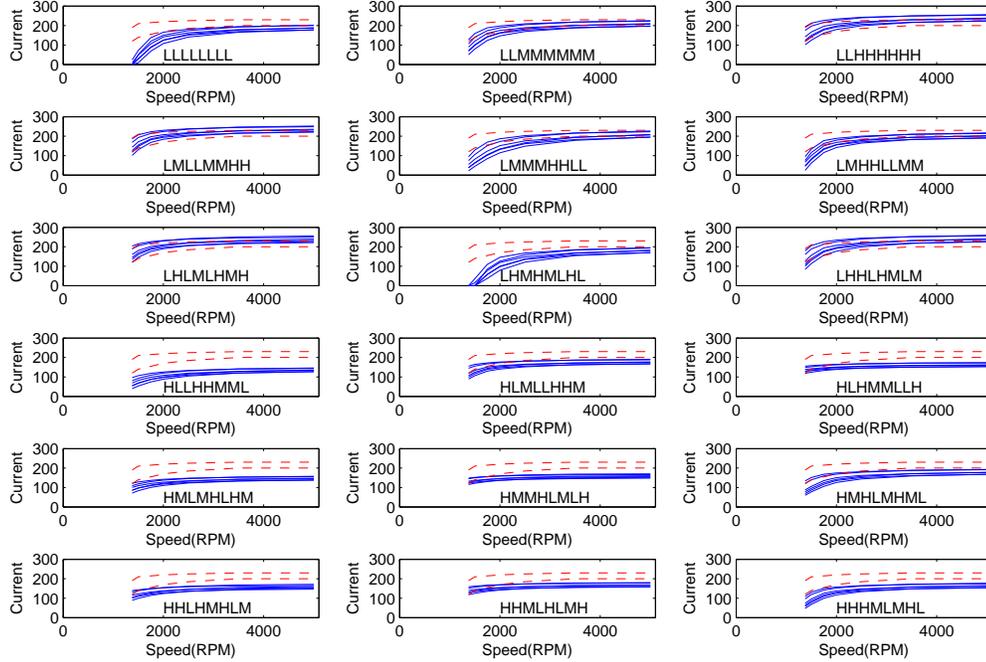
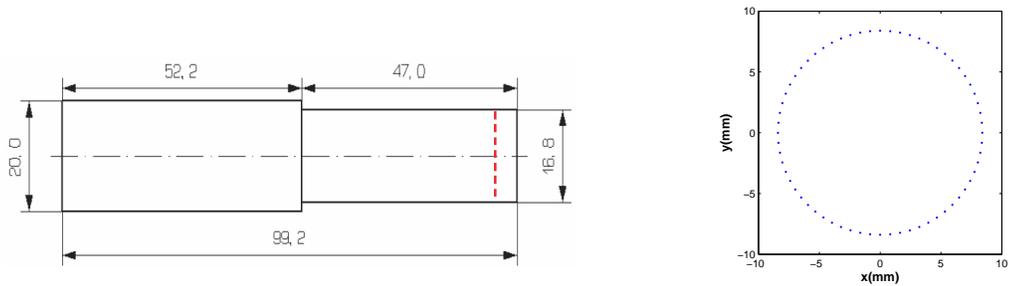


Figure 1.1. Observed profiles at each of the 18 considered control factor combinations (H=high, M=medium, L=low). The specification limits L and U are shown in dashed lines.



(a) A sketch of the machined part with the cross-section of interest shown as a dashed line.

(b) An instance of the response of interest measured over a set of 64 points along the circular cross-section.

Figure 1.2. An example of a 2D shape response experiment.

The overall problem under consideration in this dissertation can be stated as follows: a response of interest in a manufacturing process or in an engineering

design problem is either a profile or a geometric shape. The profile or shape of this response over the underlying m -dimensional space ($m \in \{1, 2, 3\}$) defines the performance or quality of a given product or process under study. The profile or shape of the response is assumed to be modifiable through a set of controllable factors (i.e., factors that are modifiable while the process is running) and is also affected by some noise factors (i.e., factors that are hard or costly to control). The goal is to find the best control factor settings that lead to the desired target response shape or profile in the presence of variation or uncertainty in the noise factors. Problems when noise factors are present and the response is univariate are well-known and receive the name Robust Parameter Design (or RPD) in the Design of Experiments (DOE) literature (Myers and Montgomery, 1995).

Two problems are usually of interest when analyzing designed experiments in manufacturing. The first is to identify the significant factors and the nature of their effects. The second is to optimize the process under study. Designed experiments where the response of interest is the shape of an object have been studied in fields other than manufacturing (Snee and Andrews, 1971). Statistical tests to analyze designed experiments with shape responses for manufacturing applications have been recently introduced by Del Castillo and Colosimo (2011). There has been no study thus far that addresses the performance of these tests for manufacturing applications. Thus, a part of this dissertation is devoted to study the performance of these tests.

Furthermore, Del Castillo and Colosimo (2011) provided an optimization algorithm that finds the best factor settings with the closest mean shape to the target. However, their method does not consider the existence of noise factors and filters size information from that of the shape of the object, testing separately for differences in shape and size. A methodology to solve the RPD problem for shape responses that does not filters scale or size information is therefore needed.

The second theme of this dissertation refers to 1D (profile) response modeling and optimization. In engineering, profile responses are usually sampled at high frequency. Thus, it is natural that response values sampled at nearby locations will tend to be correlated. If the within-profile correlation is strong and neglected, model parameter estimates will be inefficient and will yield less precise predictions (Del Castillo et al., 2011). Inefficient estimators will provide misleading “optimal”

solutions when used to solve an optimization or RPD problem. Hence, the within-profile correlation needs to be modeled.

Current approaches to solve the RPD problem of profile responses include that of Nair et al. (2002). In their model, profiles are modeled using a two-stage parametric model, with the first stage parameters assumed to be modifiable through the control factors. This allows to change the form of the profile response. The model parameters are estimated through frequentist methods, and thus the uncertainties in their estimates are not considered, and this may lead to non-optimal solutions. Furthermore, their method assumes a small number of measurements per profile to be able to estimate the model parameters. A more recent approach is that of Del Castillo et al. (2011). They use a full Bayesian two-stage mixed effects regression model. This approach considers all sources of uncertainty present and provides a probability measure of how well the process will perform, but has some limitations. First, it assumes no profile-to-profile correlation. This assumption might not always be true, specially if profiles are observed at nearby design factor settings. Second, it requires a parametric model for the mean structure. This parametric model must provide an excellent fit for the mean profile shape, otherwise, model predictions will not be accurate. Hence, this approach is not flexible enough to handle arbitrarily shaped profile responses which can not be modeled adequately with a linear statistical model. Therefore, a more flexible modeling approach to model such profile response systems is needed.

Three areas in Engineering and Statistics are impacted by this research, namely: (a) Robust Parameter Design, (b) Spatio-Temporal modeling, and (c) Statistical Shape Analysis of manufacturing data. In the reminder of this chapter, sections 1.1 through 1.3 provide a brief background for each of these three topics. Section 1.4 provides a brief introduction to Bayesian analysis that is required for some of the work considered in this dissertation. Finally, sections 1.5 and 1.6 state the objectives of this dissertation and provide an outline respectively.

1.1 Robust Parameter Design

The origin of Robust Parameter Design (RPD) dates back to the 1980's when Taguchi and Wu (1980) introduced the concept of robustness to “noise factors” in

experimental design. Although the idea of robustness with respect to variations in the model assumptions existed in the Statistics literature for a long time (Box and Wilson, 1951), it was Taguchi who introduced and popularized the idea of finding process settings that are robust or insensitive with respect to uncontrollable sources of variation.

Taguchi divided the factors affecting the process into two types, *control factors* and *noise factors*. Control factors are variables that can always be manipulated during the experiment or while the process is running. For example, the cutting speed or the feed rate can be changed easily to get the desired surface finish during a metal cutting operation. Noise factors are factors affecting the process that are difficult or costly to control once the experiment is completed and the process or product needs to be kept operating or needs to be used, respectively. For example, coolant temperature may also have an effect on the surface finish of a machined part, but coolant temperature is usually beyond the control of the operator. Thus, changes in coolant temperature will lead to variations in the surface finish, but such variations are undesirable. However, for the purpose of a carefully designed experiment, the coolant temperature could be regulated with adequate equipment, not typically available when the machine tool is normally used. In such an experimental setting, coolant temperature is therefore an instance of a noise factor.

Figure 1.3 illustrates a prototypical RPD problem, which can be stated as follows: a response of interest is affected by a set of control factors $\mathbf{x}_c = \{x_{c1}, x_{c2}, \dots, x_{ck}\}$ and a set of noise factors $\mathbf{x}_n = \{x_{n1}, x_{n2}, \dots, x_{nl}\}$. The goal is to find the best settings of \mathbf{x}_c that make the process achieve a desired target response that is insensitive to random variation in \mathbf{x}_n . Following Taguchi, significant control-noise interactions, $x_{ci} \times x_{nj}$, are assumed present, and hence process variation can be controlled by changing control factor settings.

Although the original RPD problem assumes a single (univariate) response, it can be applied to multiple response systems such as profile or geometric shape responses. For example, consider the electric alternator design example described in the introduction to this chapter. In that example we have $\mathbf{x}_c = (x_1, x_2, \dots, x_8)'$ and $\mathbf{x}_n = (x_9, x_{10})'$. The experimental goal is to find the best settings of \mathbf{x}_c that lead to a desired profile response with maximum probability regardless of the

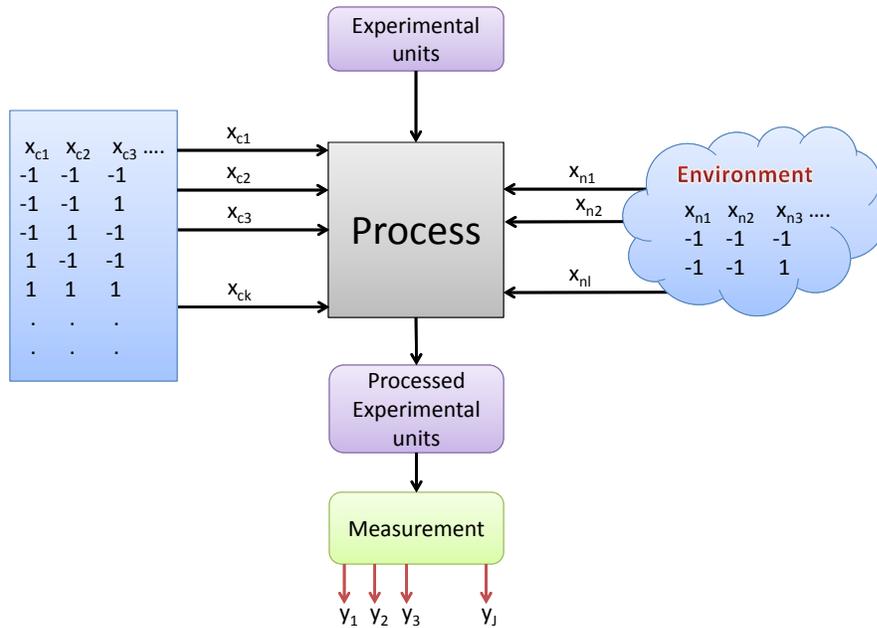


Figure 1.3. Prototypical RPD problem showing the inputs (controllable and noise factors) and the corresponding outputs of a process that acts on certain experimental units (i.e., parts in a manufacturing application). This figure is adapted from Del Castillo (2007).

variability due to the noise factors.

1.2 Spatio-Temporal Models

Spatio-temporal models arise when data is collected across both time and space. An example is a monitoring network of an atmospheric pollutant where data is collected at regular time intervals across certain geographical area. Since the observations at each monitoring site typically form a dependent time series, data analysis has to take into account not only the spatial dependence among the monitoring sites, but also the temporal correlation. Many data sets have a spatio-temporal structure, especially in Geographical applications, and spatio-temporal modeling and analysis have received increased attention in recent years (see Cressie, 1993; Banerjee et al., 2004).

Let \mathbf{d} be an actual physical location and t be a time location (instant or epoch). Then a spatio-temporal response $y(\mathbf{d}, t)$ can be seen in two different ways:

- $y(\mathbf{d}, t) = y_d(t)$: spatially varying time series data.
- $y(\mathbf{d}, t) = y_t(\mathbf{d})$: a dynamic spatial model that evolves over time.

It is necessary to distinguish between the two models, as one may be interested in the time dependence structure, allowing for spatial dependence between time series (first model), or in the space dependence structure, allowing for evolution over time (second model).

Spatio-temporal data are not found only in Geographical applications; they also appear in manufacturing applications. Here, the “time” dimension is not necessarily a temporal one but refers simply to the locations of another variable over which a functional response is observed. For example, Govaerts and Noel (2005) reported an experiment where the elastic modulus of “green” parts (products before the sintering operation) in a metal injection moulding process was studied. The elastic modulus was measured for 25 parts at 701 values of the debinding temperature ranging from 10 to 80° C, and this constitutes the “temporal” dimension in this example. The experiment consisted of two controllable factors in the ingredients of the binder, namely, x_1 =Xanthan concentration (varied at 5 levels from 1 to 5) and x_2 =Chromium/Xanthan concentration ratio (varied at 4 levels from 1:1 to 4:1). If we let the design factor space be the “spatial” space and the temperature locations be the “temporal” space, then this is an instance of a spatio-temporal process. Figure 1.4 shows a plot of the temporal process at three different locations in the “spatial” space constructed by the design factors space, the “ x -space”. Points at nearby temperature values along the same profile may be correlated, in what constitutes temporal correlation. Points at the same temperature value but at different design factor settings might be correlated, and this constitutes spatial correlation.

The most common approach for modeling spatio-temporal data is the Gaussian Process (or Gaussian Random Function) model (Banerjee et al., 2004; Fang et al., 2005). In Gaussian Process modeling, the process that generates the observed data is assumed to follow an infinite dimensional normal distribution. Hence, any finite

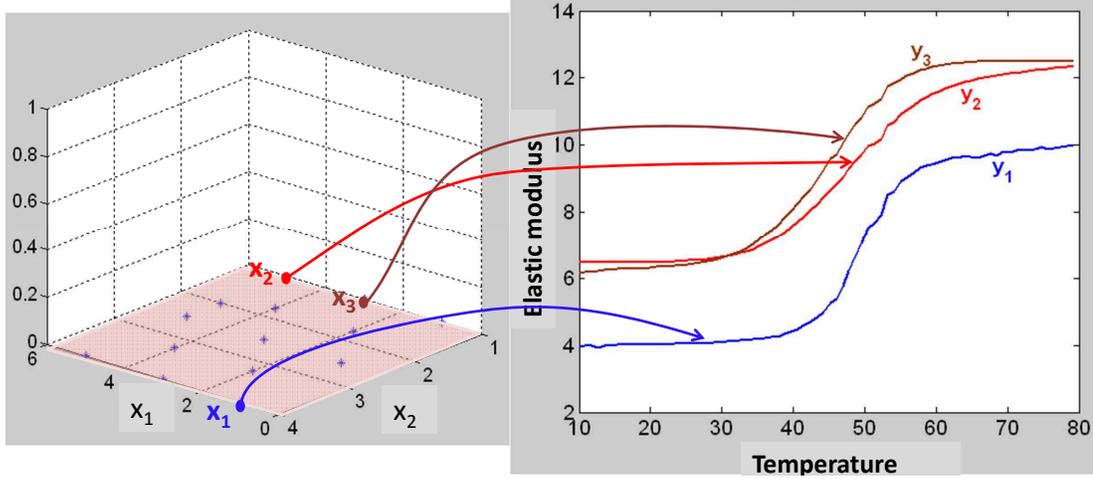


Figure 1.4. Illustration of three realizations of a “temporal” process (e.g., observed elastic modulus profiles) at four different controllable factor combinations, i.e., four locations in the “ x -space”.

set of the observed data will also be normally distributed, that is

$$\mathbf{y}(\mathbf{d}, t) = N_p(\boldsymbol{\mu}(\mathbf{d}, t), \boldsymbol{\Sigma}(\mathbf{d}, t))$$

For an inference based on a single sample of this process to be a valid generalization of the observed process, a regularity condition must hold. This regularity condition is the process *ergodicity*, which broadly means that statistics computed over space and time tend to their “ensemble” quantities if the sample size increases. The ergodicity conditions are very technical, but it turns out they are achieved for a Gaussian stationary process with a spatio-temporal correlation function that decays over space and time (Cressie, 1993). This is the main reason behind the wide use of stationary Gaussian Processes in Spatial Statistics. Chapter 5 of this dissertation presents new methodology based on Gaussian Processes for the solution of the RPD problem for profile responses.

1.3 Statistical Shape Analysis of Manufacturing Data

The statistical analysis of shapes is a relatively new field within the history of Statistics. Only in the 1980s did important work appear on “shape theory” due to Kendall (1984) and Bookstein (1986). The *shape* of an object is defined by Kendall as all the information of the object that is invariant with respect to similarity transformations (rotations -excluding reflections-, translations, and dilations or changes of scale) on the Euclidean space (Kendall, 1984). In manufacturing experiments where the response of interest is the shape of a part, the usual goal is to determine if any of the factors varied in the experiment affect the resulting shape of the parts. This is a similar goal as in classic experimental design, with the additional feature that the response of the process is the complete part geometry. For example, consider the 2D circular parts example by Del Castillo and Colosimo (2011) mentioned at the introduction of this chapter. The 2D shape was measured using a Coordinate Measuring Machine (CMM) that acquired a set of 64 equally spaced points on each part. The goal of the experiment was to determine the effect of depth and cutting speed on the circularity of the parts and to determine the best settings of these factors to achieve the most circular parts. The effect of these factors on the circularity (or cylindricity) of the parts can be studied by conducting an analysis of variance (ANOVA) of the form error for roundness (or cylindricity) of each part. The circularity form error is frequently calculated in tolerancing practice as the smallest difference between the radii of two coaxial circles (or cylinders) that enclose all the measurements in a part (Henzold, 2006; Krulikowski, 1996; see also Figure 1.5). A standard ANOVA (Montgomery, 2009) can then be conducted on the observed form errors to analyze the impact the two factors have on these deviations from circularity.

As an alternative to the use of the form errors and a standard ANOVA test, or for parts with complex geometry for which no standard form error definition exists (i.e., “free-form” manufactured parts), we can use SSA techniques to perform the analysis of the experiment. This was suggested by Del Castillo and Colosimo (2011).

SSA techniques have been widely used over the past two decades in applica-

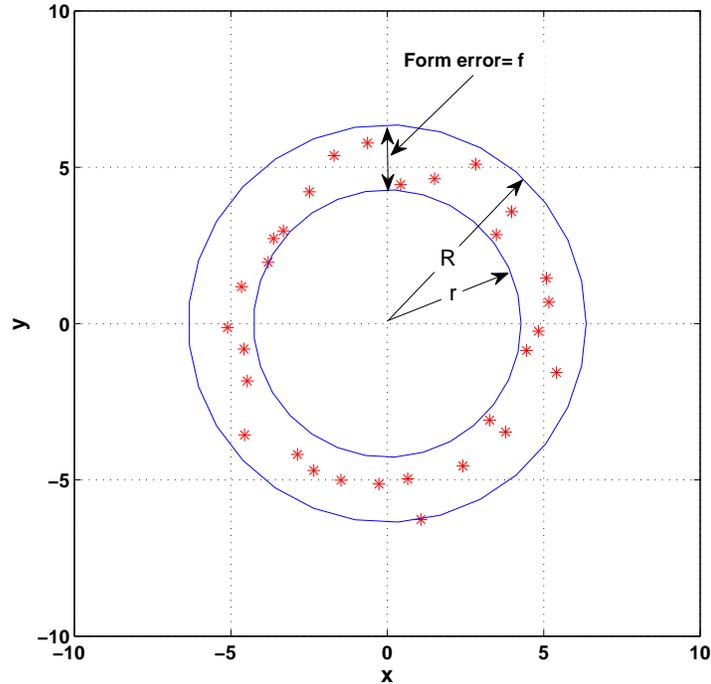


Figure 1.5. Circularity form error calculation.

tions other than manufacturing, for example, in the analysis of shapes of objects of palaeontological, biological, or geological interest, or for text recognition in computer vision. Dryden and Mardia (1998) give an introduction to SSA with emphasis on these non-manufacturing applications.

The general problem in SSA can be stated as follows. Suppose we wish to compare the mean shapes of a groups of n parts each. The groups of parts can correspond to different levels of a factor varied in a manufacturing experiment, e.g., different cut depths. Let \mathbf{X}_{ij} be a $k \times m$ matrix of coordinate measurements ($m=2$ or 3) representing each part j in group i (in the SSA literature, matrix \mathbf{X} is referred to as the configuration matrix -in analogy to mechanics-, and sometimes simply as “the object”). In practice, these coordinates are usually gathered via a CMM or laser scanners (Barcenas and Griffin, 2001) and are assumed to correspond from part to part. In SSA these corresponding points receive the name “landmarks”, which we adopt herein. Landmarks are points of special interest or

unique characteristics, given by the 2- or 3-dimensional Cartesian coordinates of a point on the object surface and a given label for the point, usually a sequential number $1, 2, \dots, k$ which corresponds from object to object.

Suppose each observed configuration matrix can be modeled as a function of the levels of a single controllable factor as follows:

$$\mathbf{X}_{ij} = \boldsymbol{\mu}_i + \mathbf{E}_{ij} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, n \quad (1.1)$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\tau}_i$, with dimension $k \times m$, represents the mean shape of the object obtained under the i^{th} level or treatment and \mathbf{E}_{ij} is a $k \times m$ matrix of errors such that $\text{vec}(\mathbf{E}_{ij}) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is a $km \times km$ covariance matrix and $\text{vec}(\cdot)$ is the operator that concatenates the columns of a matrix into one vector. This is just a one-way ANOVA on a matrix response. We wish to test the significance of the factor levels effect, i.e.:

$$H_o : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \dots = \boldsymbol{\tau}_a = \mathbf{0}$$

versus the alternative

$$H_1 : \text{at least a } \boldsymbol{\tau}_i \neq \mathbf{0}.$$

If H_o is true, all the a groups have the same mean shape $\boldsymbol{\mu}$. We point out that a MANOVA test cannot be applied, given that in manufacturing applications the number of measurements k is typically very large compared to the sample size n . A test for the hypothesis above was given by Goodall (1991), and later extended to the two-way (interaction) case by Del Castillo and Colosimo (2011). The power of these tests to detect deviations with respect to a target shape, particularly those of interest in manufacturing, is not well understood and its study is the subject of chapter 3 in this dissertation.

1.4 Bayesian Analysis

Since the work on profile response optimization presented in chapter 5 in this dissertation follows a Bayesian approach, we provide here few basic notation and definitions from Bayesian statistics that are used in the sequel. Statistical methods

known as frequentist (or classical) assume that unknown parameters are fixed constants, and define probabilities by using limits of event frequencies. Bayesian methods on the other side treat parameters (and any other unknown quantity) as random variables and define probability as the degree to which it is believed an event to be true. The “Bayesian” term comes from the usage of Bayes’ theorem, named after Reverend Thomas Bayes (1702-1761).

Suppose we observe a random variable y and wish to make inferences about another random variable θ , where θ is drawn from some distribution $\pi(\theta)$. From the definition of conditional probability,

$$\pi(\theta | y) = \frac{\pi(y, \theta)}{\pi(y)}$$

but also

$$\pi(y | \theta) = \frac{\pi(y, \theta)}{\pi(\theta)}$$

which implies

$$\pi(y, \theta) = \pi(y | \theta)\pi(\theta)$$

putting these together gives the Bayes theorem

$$\pi(\theta | y) = \frac{\pi(y | \theta)\pi(\theta)}{\pi(y)} \tag{1.2}$$

with p possible outcomes $\Theta = (\theta_1, \dots, \theta_p)$,

$$\pi(\theta_j | y) = \frac{\pi(y | \theta_j)\pi(\theta_j)}{\pi(y)} = \frac{\pi(y | \theta_j)\pi(\theta_j)}{\sum_{i=1}^p \pi(y | \theta_i)\pi(\theta_i)} \tag{1.3}$$

The term $\pi(\theta)$ is called the prior distribution of the possible θ values, while $\pi(\theta | y)$ is called the posterior distribution of θ given the observed data y (see, e.g., Carlin and Louis, 2008).

In Bayesian analysis we start with some initial knowledge/guess about the

distribution of the unknown parameter(s), the prior distribution of Θ , $\pi(\Theta)$. From Bayes theorem, the data (likelihood) is combined with the prior distribution to produce a posterior distribution,

$$\begin{aligned}\pi(\Theta | \mathbf{y}) &= \frac{1}{\pi(\mathbf{y})}\pi(\mathbf{y} | \Theta)\pi(\Theta) \\ &= \text{constant} \cdot \pi(\mathbf{y} | \Theta)\pi(\Theta)\end{aligned}$$

where $\mathbf{y} = (y_1, \dots, y_n)$ is the set of n observations of the random variable y , $\pi(\mathbf{y} | \Theta) = l(\Theta | \mathbf{y})$ is just the likelihood function, and $1/\pi(\mathbf{y})$ is a constant (with respect to Θ). Because of this, the posterior distribution is often written as

$$\pi(\Theta | \mathbf{y}) \propto \pi(\mathbf{y} | \Theta)\pi(\Theta) \quad (1.4)$$

where the symbol “ \propto ” means “proportional to” (equal up to a constant). Note that the constant $\pi(\mathbf{y})$ normalizes $\pi(\mathbf{y} | \Theta) \cdot \pi(\Theta)$ to one, and hence can be obtained by integration,

$$\pi(\mathbf{y}) = \int_{\Theta} \pi(\mathbf{y} | \Theta) \cdot \pi(\Theta) d\Theta$$

The dependence of the posterior on the prior provides an indication of how much information on the unknown parameter values is contained in the data. If the posterior is highly dependent on the prior, then the data likely has little signal, while if the posterior is largely unaffected under different priors, the data are likely highly informative.

In Engineering problems and especially in process optimization applications, it is usually of interest to make predictions of new observations based on the observed data. The probability distribution of a new observation \tilde{y} is defined as

$$\pi(\tilde{y} | \mathbf{y}) = \int_{\Theta} \pi(\tilde{y} | \mathbf{y}, \Theta)\pi(\Theta | \mathbf{y})d\Theta \quad (1.5)$$

where $\pi(\tilde{y} | \mathbf{y})$ is the *posterior predictive density* (see, e.g., Gelman et al., 2004). The integration in (1.5) needs to be done by first generating samples from the posterior distribution $\pi(\Theta | \mathbf{y})$. For complicated probability models the posterior

distribution $\pi(\Theta | \mathbf{y}) \propto \pi(\mathbf{y} | \Theta) \cdot \pi(\Theta)$ is usually not identified with a standard probability distribution. Hence, to sample from it Monte Carlo simulations, especially, Markov Chain Monte Carlo methods are used (Gelman et al., 2004).

1.5 Dissertation Objectives

This research focuses on developing new and efficient techniques to solve the RPD problem for profile and shape response experiments. Specific goals of this research include:

A. 1D profile response systems

1. Develop and study a spatio-temporal Gaussian Process model for profile response experiments where noise factors are present. In this model, a full Bayesian analysis will be used to consider all the uncertainties in the model parameters and in the noise factors.
2. Implement an adaptive, robust and efficient Markov Chain Monte Carlo (MCMC) algorithm that estimates *all* the model parameters of the spatio-temporal Gaussian Process model. This will allow the Bayesian optimization of the profile response of goal number 1. We will develop Bayesian methodology for finding operating conditions of a process \mathbf{x}_c that maximizes the posterior probability a profile response lies between certain tolerance “bands”.

B. 2D and 3D shape responses

There is considerable debate about the merits of the different SSA techniques used to analyze designed experiments, and it is not clear which one to use in a given situation (Dryden and Mardia, 1998). It is the objective of this research to extend the preliminary results of the SSA methods developed by Del Castillo and Colosimo (2011) into several directions:

1. We will investigate the power performance of four different SSA tests (the F-test in Goodall (1991), the ANOVA Permutation test in Del Castillo and Colosimo (2011) and the EDMA-I and the EDMA-II methods in Lele and Richtsmeier (1991) and Lele and Cole (1996) for detecting shape differences in

2D and 3D objects. For cases of interest in manufacturing where a definition of form error exists (such as circularity and cylindricity), the SSA tests will be compared also to the standard approach (ANOVA on the form errors). The performance of the SSA tests will also be investigated under different variations of the error structure.

2. We will extend Del Castillo and Colosimo's (2011) method for shape optimization to allow for noise factors and to account for size separately from shape information. An RPD technique will be developed that finds the optimal mean shape with the minimum variability.

1.6 Dissertation Outline

This dissertation is organized as follows. In Chapter 2, we review the literature on the two main themes this dissertation touches: analysis and optimization of profile response experiments and Statistical Shape Analysis of manufacturing data.

There are several statistical tests found in the literature to analyze designed experiments when the geometry of the parts is of interest. In chapter 3, we study the statistical performance of these tests under several variations of their underlying assumptions. The performance studies are based on simulated data sets as well as real shape data examples found in the literature.

Existing optimization tools for shape responses do not consider variability induced by noise factors. These methods do not separate the size information from the shape information, since the classical definition of shape, according to Kendall (1984), filters dilation effects. In chapter 4, the shape optimization technique proposed by Del Castillo and Colosimo (2011) is extended to the optimization of the *forms* of the parts, where by form we define all the geometric information that remains once location and rotation effects (but not dilations) are filtered out. We also consider the case in which noise factors are present and solve the RPD problem for form responses. The extended method is illustrated through two examples in manufacturing.

In chapter 5, we deal with profile response modeling and optimization. A spatio-temporal Gaussian Process (GP) model is utilized for modeling a profile

response in an experiment. The proposed approach is Bayesian and allows solving RPD scenarios where noise factors exist. The robustness of the methodology with respect to its underlying assumptions is presented. The GP model is illustrated with three real manufacturing profile data examples taken from the literature.

Chapter 6 concludes the dissertation with a discussion about the contributions of this work and about areas for further research.

Chapter 2

Overview of Previous and Related Research

Two main themes are explored in this dissertation. The first is the statistical analysis of 2D and 3D geometries (shapes). The second is the modeling and analysis of 1D profile responses. This chapter presents an overview of existing work in these two topics. Section 2.1 presents related work that has been done in the Statistical Shape Analysis field. Section 2.2 reviews the work that has been done towards modeling and Robust Parameter Design of profile responses.

2.1 Statistical Shape Analysis of Manufacturing Data

Shape analysis dates back to 1917 when D'Arcy Thompson published his famous book *On Growth and Form* in which he studied shape transformations of organisms. This was the beginning of the Morphometrics field in Biology, which evolved through the XX century. However, no work was conducted in the design and analysis of experiments for the shape of an object, until about 50 years after Thompson when Snee and Andrew (1971) studied shape analysis in a designed experimental framework. They developed statistical techniques to analyze agricultural experiments where the shape of carrots and sweet potatoes were of interest. Their work focuses on testing the effect of the experimental factors on the shape of 2D or

3D objects through several diameter to length ratios measured on each object. Statistical shape analysis techniques have been developed and applied in many areas of the natural sciences over the past 30 years, e.g., biology, paleontology and geology. In particular, SSA is known as *geometric morphometrics* in biology, a field in which some authors refer to a “morphometrics revolution” (Adams et al., 2004) given the success SSA had over previous techniques used to analyzed shapes.

Statistical Shape Analysis (SSA) has three origins. The first is that of Kendall (1984) who developed the concepts of pre-shape and shape space, which we will review in section 2.1.1 (see also Appendix A for a review of mathematical notations used in Kendall’s theory). At about the same time Bookstein (1986) began to study shape theoretical problems in zoology. He developed one of the most used shape coordinate systems for planar data, where the shape of the object is preserved by removing translation, rotation and scale effects. Figure 2.1 shows an example of a three-landmark object and its representation in the Bookstein coordinate system. A third early contribution to SSA is the work of Ziezold (1994) who studied the equality of shape distributions. In 1998, Dryden and Mardia summarized developments up to that date in their book *Statistical Shape Analysis*.

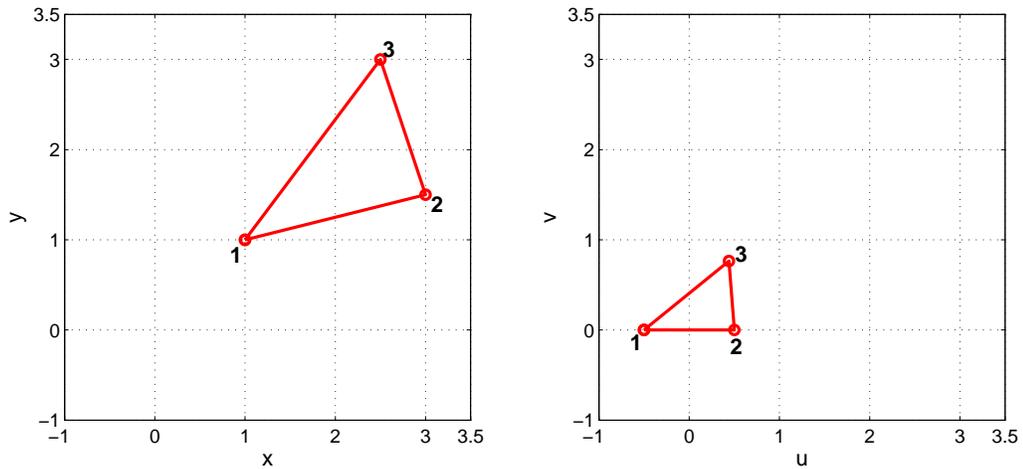


Figure 2.1. A three landmark object (i.e, a triangle) represented in the Cartesian coordinate system (left plot) and its preserved shape as presented in the Bookstein coordinate system (right plot).

An area where shape analysis has received increased attention in recent years is computer vision. In computer vision, the most common problem of interest regarding shape analysis is the classification problem, where a predefined set of shape classes is available and the goal is to assign a newly observed object to one of these classes. Another problem considered in computer vision is shape retrieval, where a certain object is of interest (the query shape) and the goal is to find the set of objects that have similar shapes in a shape database. We point out here that the kind of shape analysis we consider in this work is different from what is being done in computer vision. In this dissertation, a designed experimental framework is assumed where the object shape is of interest and the goal is to study the effect of the experimental factors on the object shape (hence the shapes of objects under study are assumed modifiable). Furthermore, we study shape optimization to improve manufacturing processes performance. For example, recall the 2D circularity designed experiment example by Del Castillo and Colosimo (2011) mentioned in the introduction of chapter 1. The response of interest in that example is the 2D circular cross-section of the machined Titanium shaft represented by a set of 64 points measured along that cross-section. The goal of the experiment was to study the effect of depth of cut and cutting speed on the circularity of the shaft. The goal was also, to find the best settings of these two factors that provide the most circular shafts.

Shape representation and description found in the literature can be classified into four categories (Dryden and Mardia, 1998):

1. Landmark-based objects: in this case each object is represented by $k \times m$ configuration matrix where k is the number of landmarks measured on the object and m is the dimension of the landmark measurements.
2. Outline-based objects: in this case the object is viewed as a closed curve and this curve is represented by a set of k measurements in m -dimensional space.
3. Solid objects: in this case the object is viewed as a region in m -dimensional space and a binary function that is equal 1 if and only if a point is within that region.

4. Grey-level objects: in this case the texture of the object is also important and objects are treated as images.

In this dissertation we deal with the first type of shape representation methods, namely the landmark-based objects given that it is very common in manufacturing data obtained via coordinate measuring machines (CMMs).

To better understand statistical shape inference from the designed experimental point of view, we first review some necessary geometric notation in section 2.1.1. Then in sections 2.1.2 through 2.1.4 we review existing Statistical Shape Analysis techniques found in the literature.

2.1.1 Pre-shape and Shape Space

Kendall (1984) defines the shape of an object as all the geometrical information that remains after removing the translation, rotation and scale (size) information from the object. Let \mathbf{X} be a $k \times m$ matrix that contains the k landmark coordinates of an object in m (2 or 3) dimensions, i.e., the configuration matrix. Define the Helmert submatrix \mathbf{H} of size $(k-1) \times k$ such that the j^{th} row is equal to $[\underbrace{h_j, h_j, \dots, h_j}_{j \text{ times}}, \underbrace{0, \dots, 0}_{k-j-1 \text{ times}}]$ where $h_j = -(j(j+1))^{-1/2}$. Notice that $\mathbf{H}\mathbf{H}' = \mathbf{I}_{k-1}$ and the rows of \mathbf{H} are contrasts, since they add up to zero. The *pre-shape*, \mathbf{Z} , of an object \mathbf{X} is all the information remaining after removing location and scale effects. Using the above notation, \mathbf{Z} is defined as

$$\mathbf{Z} = \frac{\mathbf{H}\mathbf{X}}{\|\mathbf{H}\mathbf{X}\|} \quad (2.1)$$

where $\|\cdot\|$ is the Frobenius norm of a matrix such that $\|\mathbf{A}\| = \sqrt{\sum_i \sum_j a_{ij}^2}$. The location effect is removed by the multiplication with \mathbf{H} in the numerator of (A.1). The denominator in (A.1) is the centroid size measure of \mathbf{X} , and hence the pre-shape \mathbf{Z} has a unite centroid size. Kendall (1984) has shown that the pre-shape space (the space of all possible pre-shapes, denoted by S_m^k) is a hypersphere of unit radius in $(k-1)m$ real dimensional space.

The term “pre-shape” was coined by Kendall (1984). To get the shape of \mathbf{X}

(denoted by $[\mathbf{X}]$), rotation is filtered out from the pre-shape \mathbf{Z} , that is:

$$[\mathbf{X}] = \{\mathbf{Z}\mathbf{\Gamma} : \mathbf{\Gamma} \in SO(m)\} \quad (2.2)$$

where $[\cdot]$ is an equivalence class defined over all orthogonal rotation matrices $\mathbf{\Gamma}$ of size $m \times m$, and $SO(m)$ is the special orthogonal group of $\mathbf{\Gamma}$ such that $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{\Gamma}'\mathbf{\Gamma} = \mathbf{I}_m$ and $\det(\mathbf{\Gamma})=+1$, see Appendix A for more information about this and other mathematical notions used in Kendall's SSA. Notice that a shape is a set defined over all values of $\mathbf{\Gamma} \in SO(m)$. The shape space is defined as all the possible shapes a k landmark object in m dimensional space can have. Generally, this shape space is a non-Euclidean M -dimensional manifold which results from mapping the set of pre-shapes that are equivalent under non-reflective rotations into a single point in the shape space. The dimension of the shape space is

$$M = km - m - 1 - \frac{m(m-1)}{2}, \quad (2.3)$$

because we lose m degrees of freedom to remove the location effect, one degree of freedom to remove the size effect, and $\frac{m(m-1)}{2}$ degrees of freedom to remove the rotation effect given that we initially start with km degrees of freedom.

For example, consider the case where $k = 3$ in $m = 2$ dimensional Cartesian space, i.e., a triangle. The pre-shape space of a triangle is a 4-dimensional hypersphere, since $(k-1)m = (3-1)2 = 4$ and its shape space is a 3-dimensional sphere, since $M = km - m - 1 - m(m-1)/2 = 3 * 2 - 2 - 1 - 2(2-1)/2 = 3$.

The size on an object is sometimes of interest and needs not to be filtered always, depending on the application. The remaining information of an object after removing translation and rotation effects but not dilation effects receives the name *shape-and-size* in SSA literature (Dryden and Mardia, 1998). In chapter 4 of this dissertation we deal with the shape-and-size information of objects which we will refer to as the *form* of the objects.

In the analysis of shapes, the observed objects may be oriented or located in space differently, and some type of alignment or "registration" is necessary before conducting a test for mean shape differences. It is precisely this registration what filters the effects of similarity transformations. In the next section we discuss one of the commonly used registration algorithms, namely the Generalized Procrustes

Algorithm. We point out that all SSA methods considered in this dissertation require *corresponding* (or labeled) landmarks between the different objects, i.e., landmarks that “match”. The landmark matching problem has received attention in the Pattern Recognition and computer vision literature in recent years, where it is called the point matching or shape matching problem. For a discussion of landmark matching algorithms, see Appendix B.

2.1.2 Distances Used in Shape Analysis

In this section, we review three different types of distances that are commonly used in the SSA literature to describe how “close” the shapes of two objects, \mathbf{X}_1 and \mathbf{X}_2 , are to each other. These distances are performed in the pre-shape space and they are:

1. The full Procrustes distance: the shortest linear distance between one pre-shape and the other along a tangent plane to the pre-shape space at one of these two pre-shapes. This distance is denoted by $d_F(\mathbf{X}_1, \mathbf{X}_2)$ and is defined mathematically as:

$$d_F(\mathbf{X}_1, \mathbf{X}_2) = \min_{\Gamma \in SO(m), \beta \in \mathbb{R}} \|\mathbf{Z}_2 - \beta \mathbf{Z}_1 \Gamma\| \quad (2.4)$$

Notice here that the minimization is carried over the rotation and scaling of pre-shapes.

2. The partial Procrustes distance: the shortest Euclidean distance between the two objects, that is:

$$d_p(\mathbf{X}_1, \mathbf{X}_2) = \min_{\Gamma \in SO(m)} \|\mathbf{Z}_2 - \mathbf{Z}_1 \Gamma\| \quad (2.5)$$

where minimization is performed over the rotations only. The solution of this problem is well-known in statistics and computer science (Jackson, 2003; Horn et al., 1988) and equals $\hat{\Gamma} = \mathbf{U}\mathbf{V}'$ where \mathbf{U} and \mathbf{V} are obtained through Singular Value Decomposition (SVD) such that $\mathbf{Z}_2' \mathbf{Z}_1 = \mathbf{V} \mathbf{\Lambda} \mathbf{U}$.

3. Procrustes distance: this is the shortest distance between two pre-shapes along the surface of the pre-shape space. Hence, this distance is not linear

and is defined as:

$$\rho(\mathbf{X}_1, \mathbf{X}_2) = 2 \arcsin(d_p(\mathbf{X}_1, \mathbf{X}_2)/2) \quad (2.6)$$

The geometric interpretation of these three distances is shown in Figure 2.2. Notice that both d_F and d_p are linear measures but ρ is not linear. It can be shown that $d_f \approx d_p \approx \rho$ if the two objects are similar, i.e., they are close to each other in the pre-shape space. These three distances are *extrinsic* to the shape space. A statistic is said to be *intrinsic* if it is performed on the shape space and *extrinsic* if it is performed on a mapping (or projection) of that shape space (see Appendix A for more information about these measures).

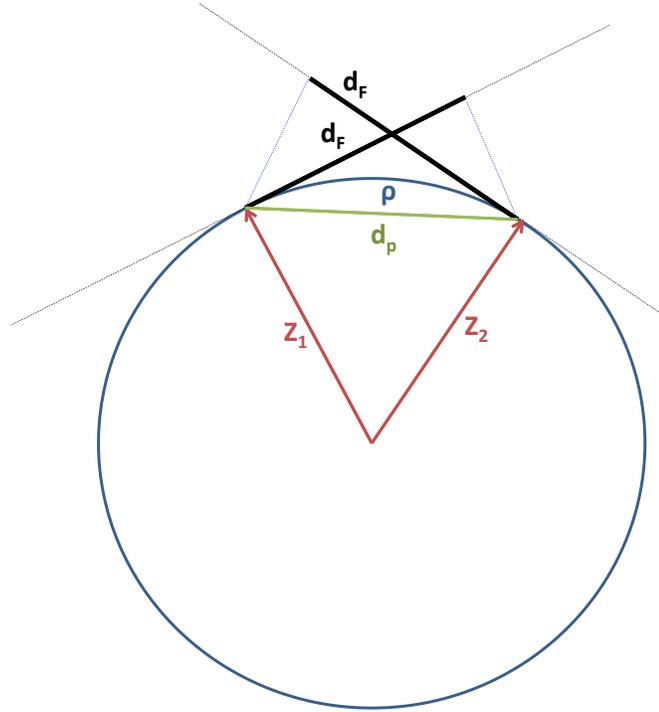


Figure 2.2. Distances between two shapes in preshape space. ρ is the Procrustes distance along the pre-shape surface, d_F is the full Procrustes distance along the tangent plane, and d_p is the partial Procrustes distance (adopted from Del Castillo and Colosimo (2011)).

In SSA, it is assumed that the observed measurements result from the mean shape $\boldsymbol{\mu}$ by applying similarity transformations (rotations -excluding reflections-, translations and dilations). If β , $\boldsymbol{\Gamma}$, and $\boldsymbol{\gamma}$ are a scaling factor, an orthogonal transformation matrix, and a translation vector, respectively, then the usual model that is assumed to generate the configurations is (Dryden and Mardia, 1998):

$$\mathbf{X} = \beta(\boldsymbol{\mu} + \mathbf{E})\boldsymbol{\Gamma} + \mathbf{1}_k\boldsymbol{\gamma}' \quad (2.7)$$

where $\mathbf{1}_k$ is a $k \times 1$ identity vector.

The Procrustes method in (2.5) is used for shape registration of two objects. A generalization of the Procrustes method for the case of n objects is the Generalized Procrustes Algorithm (GPA) developed by Gower (1975) and Ten Berge (1977). GPA estimates the mean shape $\boldsymbol{\mu}$ from a sample of n objects that may have different scales, orientations, and locations in space and as a by-product registers (aligns) the objects in the sample. GPA minimizes the sum of squared distances between every pair of objects in the sample:

$$G(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = \min_{\beta_i, \boldsymbol{\Gamma}_i, \boldsymbol{\gamma}_i} \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \|\beta_i \mathbf{X}_i \boldsymbol{\Gamma}_i + \mathbf{1}_k \boldsymbol{\gamma}'_i - (\beta_j \mathbf{X}_j \boldsymbol{\Gamma}_j + \mathbf{1}_k \boldsymbol{\gamma}'_j)\|^2 \quad (2.8)$$

Once the parameters β_i , $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\gamma}_i$ in (2.8) have been estimated, the registered shapes are found such that

$$\mathbf{X}_i^p = \widehat{\beta}_i \mathbf{X}_i \widehat{\boldsymbol{\Gamma}}_i + \mathbf{1}_k \widehat{\boldsymbol{\gamma}}'_i, \quad i = 1, \dots, n. \quad (2.9)$$

These registered shapes are called the *full Procrustes fits* in the literature (Dryden and Mardia, 1998) as opposed to partial Procrustes of (2.5) since dilation effect is included. Notice that the minimization in (2.8) is carried over the scaling factors β_i 's. But these scaling factors could be set to zeros and a minimum of zero is achieved for $G(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$. To avoid this trivial solution, the minimization in (2.8) should be performed over some constraint on the β_i 's. One common constraint is to keep the same average size of objects before and after registration, that is

$$\frac{1}{n} \sum_{i=1}^n S^2(\mathbf{X}_i^p) = \frac{1}{n} \sum_{i=1}^n S^2(\mathbf{X}_i). \quad (2.10)$$

where the size of a configuration, $S(\mathbf{X})$, is defined as $S(\mathbf{X}) = \sqrt{\sum_{i=1}^k \sum_{j=1}^m (X_{ij} - \bar{X}_j)^2}$. Here X_{ij} is the ij^{th} element of the configuration matrix \mathbf{X} and \bar{X}_j is the average of the j^{th} column of \mathbf{X} .

The GPA algorithm used to solve (2.8) subjected to (2.10) as developed by Del Castillo and Colosimo (2011) is as follows:

1. Center (but do not scale) the configurations $\mathbf{X}_1, \dots, \mathbf{X}_n$ by initially defining

$$\mathbf{X}_i^p = \mathbf{H}\mathbf{X}_i, \quad i = 1, \dots, n$$

2. Let $\bar{\mathbf{X}}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} \mathbf{X}_j^p$, $i = 1, \dots, n$. These are the averaged shapes excluding object i .
3. Do a Procrustes fit (rotation only) of the current \mathbf{X}_i^p 's on to $\bar{\mathbf{X}}_{(i)}$. This yields rotation matrices $\hat{\Gamma}_i$ from which we let

$$\mathbf{X}_i^p \leftarrow \hat{\Gamma}_i \mathbf{X}_i^p, \quad i = 1, \dots, n.$$

We repeat steps 2 and 3 for all i .

4. Compute the $n \times n$ correlation matrix $\Phi = \text{corr}(\mathbf{X}_v)$ where

$$\mathbf{X}_v = [\text{vec}(\mathbf{X}_1^p) \text{vec}(\mathbf{X}_2^p) \dots \text{vec}(\mathbf{X}_n^p)].$$

Note we stack all the m dimensions together.

5. Let $\phi = (\phi_1, \dots, \phi_n)'$ be the eigenvector of Φ corresponding to its largest eigenvalue. Then set

$$\hat{\beta}_i = \sqrt{\frac{\sum_{j=1}^n \|\mathbf{X}_j^p\|^2}{\|\mathbf{X}_i^p\|^2}} \phi_i, \quad i = 1, \dots, n$$

and let $\mathbf{X}_i^p \leftarrow \hat{\beta}_i \mathbf{X}_i^p$. The algorithm repeats steps 2 to 5 until convergence.

This registration method assumes *isotropic errors*, i.e., the variance of the coordinate data is the same at each landmark over each dimension for all objects. The assumption of isotropic errors guarantees the convergence of the GPA algorithm

(Dryden and Mardia, 1998). In many cases this assumption does not hold, and a non-isotropic error structure (i.e., the diagonal elements of Σ are not all the same) is needed. A non-isotropic structure contains different variances in different landmarks or in different dimensions. A modified GPA algorithm called GPA(Σ) (Goodall, 1991) has been suggested to find the mean shape when the error structure is non-isotropic. GPA(Σ) uses a weighted least squares (WLS) method to account for the different variances in minimizing the Procrustes distance in (2.8). Unfortunately, there is no known registration algorithm which guarantees convergence in the non-isotropic case.

2.1.3 Principle Component Analysis of Shapes

A usually performed statistical analysis on shape responses is the Principle Component Analysis (PCA). Performing PCA on shape responses helps to better understand the directions at which the shape landmarks varying the most. If we consider a set of shapes from the same object, but with slight variation, we expect little variation in the distribution of their landmarks. The PCA analysis is performed on the covariance matrix \mathbf{S} of the registered shapes defined as:

$$\mathbf{S} = \sum_{i=1}^N \text{vec}(\mathbf{X}_i^p - \bar{\mathbf{X}}) \text{vec}(\mathbf{X}_i^p - \bar{\mathbf{X}})' \quad (2.11)$$

where \mathbf{X}_i^p is the i^{th} registered shape $i = 1, \dots, N$, N is the total number of observed objects, $\bar{\mathbf{X}}$ is the average shape obtained by GPA and $\text{vec}(\cdot)$ is the operator that concatenates the columns of a matrix into one vector.

The eigenvectors of \mathbf{S} represent the basis vectors, while the eigenvalues are a measurement of the amount of variance in these directions. We note here that since GPA is used here to register shapes, then this PCA analysis is extrinsic to the shape space. If large shape variability is present, GPA estimates will not be efficient and an intrinsic shape registration methods must be used.

2.1.4 Statistical Inference on Shapes

The general problem in SSA is to test the significance of the effects of each factor levels. In the one-way case (one factor case) the hypothesis of interest related to

the effects of the a levels of the factor, just as in ordinary one-way ANOVA:

$$H_o : \tau_1 = \tau_2 = \dots = \tau_a = \mathbf{0}$$

(or equivalently, $H_o : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_a = \boldsymbol{\mu}$) where τ_i is the factor effect at level i and a is the number of the factor levels. The alternative hypothesis is then

$$H_1 : \text{at least one } \tau_i \neq \mathbf{0}.$$

If H_o is true, then all the a groups have the same mean shape $\boldsymbol{\mu}$. H_o shall be tested after registering all configuration matrices using the GPA algorithm.

A one-way ANOVA F-test for shape differences based on the distribution of the minimized Procrustes distance in (2.8) was proposed by Goodall (1991). For the one-way ANOVA model

$$\mathbf{X}_{ij} = \boldsymbol{\mu}_i + \mathbf{E}_{ij} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, n \quad (2.12)$$

the expectation of \mathbf{X}_{ij} is

$$E[\mathbf{X}_{ij}] = \boldsymbol{\mu} + \tau_i, \quad i = 1, \dots, a \quad j = 1, \dots, n \quad (2.13)$$

where n is the sample size. Let $\overline{\mathbf{X}}_{i\bullet}$ be the sample mean of the *registered* profiles of the i^{th} group, and let $\overline{\mathbf{X}}_{\bullet\bullet}$ be the grand sample mean of all registered profiles. Goodall (1991) showed that the statistic:

$$F_0 = n(n-1)a \frac{\sum_{i=1}^a d_F^2(\overline{\mathbf{X}}_{i\bullet}, \overline{\mathbf{X}}_{\bullet\bullet})}{(a-1) \sum_{i=1}^a \sum_{j=1}^n d_F^2(\mathbf{X}_{ij}, \overline{\mathbf{X}}_{i\bullet})} \quad (2.14)$$

follows approximately an F distribution with $(a-1)M$ and $a(n-1)M$ degrees of freedom for a small standard deviation σ . As we will show in chapter 3, the assumption of small σ is crucial for the distributional results above to hold, otherwise the ANOVA identity breaks down. The corresponding one-way ANOVA table (Table 2.1) is then

Table 2.1. One-way ANOVA F test for difference in mean shape.

Effect	SS	dof	MS	F
A (between groups)	$n \sum_{i=1}^a d_F^2(\overline{\mathbf{X}}_{i\bullet}, \overline{\mathbf{X}}_{\bullet\bullet})$	$(a-1)M$	$\frac{SS_A}{(a-1)M}$	$F_0 = MS_A/MS_{Error}$
Error (within groups)	$\sum_{i=1}^a \sum_{j=1}^n d_F^2(\mathbf{X}_{ij}, \overline{\mathbf{X}}_{i\bullet})$	$a(n-1)M$	$\frac{SS_{Error}}{a(n-1)M}$	
Total	$\sum_{i=1}^a \sum_{j=1}^n d_F^2(\mathbf{X}_{ij}, \overline{\mathbf{X}}_{\bullet\bullet})$			

where

$$SS_{Total} \approx SS_A + SS_{Error} \quad (2.15)$$

for small σ .

A more robust method is to test H_o using a permutation test, this was proposed by Del Castillo and Colosimo (2011). Permutation tests are robust with respect to the distribution assumptions made in standard statistical tests, and do not depend on the isotropic variance assumption. Following recommended practice in permutation tests (Edgington, 1995; Klingenberg and McIntyre, 1998), the $N = an$ objects need to be rearranged at random in different groups through resampling. We form several arrangements of a groups (each with N/a objects) to test for the hypothesis of no difference in mean shapes across the levels of the factor. For each random arrangement we compute the test statistic F_0 using equation (2.14). The observed statistic F_0^{obs} is then compared to the distribution of the F statistic obtained through resampling. If r samples result in $F_0 > F_0^{obs}$, then the p-value of the test is given by r/P , where P denotes the number of random permutations generated.

Del Castillo and Colosimo (2011) extended Goodall's one-way ANOVA to the two-way case. Similar to the two-factor designed experiments, suppose we are interested in testing the significance of the two factors A and B and their interaction on the response of interest (part geometry in our case). Let \mathbf{X}_{ijl} be the l^{th} observed configuration at level i of factor A and level j of factor B. The following interaction model is assumed:

$$E[\mathbf{X}_{ijl}] = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + (\boldsymbol{\tau}\boldsymbol{\beta})_{ij}, \quad i = 1, \dots, a; j = 1, \dots, b; l = 1, \dots, n \quad (2.16)$$

Let $\bar{\mathbf{X}}_{i\bullet\bullet}$ be the average shape at the i^{th} level of factor A, $\bar{\mathbf{X}}_{\bullet j\bullet}$ be the average shape at the j^{th} level of factor B and $\bar{\mathbf{X}}_{\bullet\bullet\bullet}$ be the grand average shape, then the corresponding two-way ANOVA table (Table 2.2) is

Table 2.2. Two-way ANOVA table for differences in mean shape (Del Castillo and Colosimo, 2011).

Effect	SS	dof	MS	F
A	SS_A	$(a-1)M$	$SS_A/(a-1)M$	$F_0^{(1)} = MS_A/MS_{Error}$
B	SS_B	$(b-1)M$	$SS_B/(b-1)M$	$F_0^{(2)} = MS_B/MS_{Error}$
AB	SS_{AB}	$(a-1)(b-1)M$	$SS_{AB}/(a-1)(b-1)M$	$F_0^{(3)} = MS_{AB}/MS_{Error}$
Error	SS_{Error}	$ab(n-1)M$	$SS_{Error}/ab(n-1)M$	
Total	SS_{Total}	$(abn-1)M$		

such that

$$SS_{Total} \approx SS_A + SS_B + SS_{AB} + SS_{Error} \quad (2.17)$$

$$\begin{aligned}
SS_{Total} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{l=1}^n d_F^2(\mathbf{X}_{ijl}^p, \bar{\mathbf{X}}_{\bullet\bullet\bullet}), \\
SS_A &= bn \sum_{i=1}^a d_F^2(\bar{\mathbf{X}}_{i\bullet\bullet}, \bar{\mathbf{X}}_{\bullet\bullet\bullet}), \\
SS_B &= an \sum_{j=1}^b d_F^2(\bar{\mathbf{X}}_{\bullet j\bullet}, \bar{\mathbf{X}}_{\bullet\bullet\bullet}), \\
SS_{AB} &= n \sum_{i=1}^a \sum_{j=1}^b d_F^2(\bar{\mathbf{X}}_{ij\bullet} - (\bar{\mathbf{X}}_{i\bullet\bullet} - \bar{\mathbf{X}}_{\bullet\bullet\bullet}) - (\bar{\mathbf{X}}_{\bullet j\bullet} - \bar{\mathbf{X}}_{\bullet\bullet\bullet}), \bar{\mathbf{X}}_{\bullet\bullet\bullet}), \\
SS_{Error} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{l=1}^n d_F^2(\mathbf{X}_{ijl}^p, \bar{\mathbf{X}}_{ij\bullet}).
\end{aligned}$$

Del Castillo and Colosimo (2011) showed that the F statistics in Table 2.2 follow

$$\begin{aligned}
F_0^{(1)} &\sim F_{(a-1)M, ab(n-1)M} \\
F_0^{(2)} &\sim F_{(b-1)M, ab(n-1)M}
\end{aligned}$$

$$F_0^{(3)} \sim F_{(a-1)(b-1)M, ab(n-1)M}$$

Del Castillo and Colosimo (2011) also proposed a two-way ANOVA Permutation test for the two-way interaction case to test for factor and interaction effects.

All of the previously mentioned SSA test (one-way ANOVA, two-way ANOVA and Permutation tests) utilize Kendall’s definition of shape where rotations, locations and dilation effects are filtered out. But, dilation effects might be of interest if one is interested in studying the effect of the experimental factors on the object size. To address this issue, Del Castillo and Colosimo (2011) proposed a separate test of the size effect. Their test is based on summarizing the object size information into a univariate response and then run a standard ANOVA test (Montgomery, 2009) on these univariate responses. The summary statistic they suggested is the object size defined as $S(\mathbf{X}) = \sqrt{\sum_{i=1}^k \sum_{j=1}^m (X_{ij} - \bar{X}_j)^2}$ where X_{ij} is the ij^{th} element of the configuration matrix \mathbf{X} and \bar{X}_j is the average of the j^{th} column of \mathbf{X} .

The ANOVA F-tests in Goodall (1991) and Del Castillo and Colosimo (2011) assume:

1. Normal errors,
2. Isotropic variances, and
3. Small error variances.

The Permutation test eliminates dependency on the first two assumptions, but not the third one. If the third assumption is violated, the ANOVA identities in (2.15) and (2.17) do not hold. Shape similarity is expressed by the extrinsic measure d_F . Distances along the non-Euclidean shape space can be approximated with this extrinsic measure only for the case of small shape variability (small σ). Otherwise, shape space non-linearity must be accounted for by using intrinsic distance measures. The case of large shape variability has been studied recently by Huckemann et al. (2010a). They proposed a Multivariate Analysis of Variance (MANOVA) for mean shape differences that uses an intrinsic mean shape measure. Current shape PCA techniques also assume small shape variability and use the extrinsic GPA registration algorithm to align shapes. Again, this analysis provides accurate results for small shape variability only, otherwise, shape space non-linearity must

be considered. Huckemann et al. (2010b) studied shape PCA analysis for the case of large shape variability where they introduce some intrinsic shape variability measures. Fortunately, the small error variance assumption holds for most manufacturing data, due to the advances in manufacturing process and precision of measuring tools used. Hence, the simpler approximation SSA techniques can be used to analyze manufacturing shape data.

As mentioned before objects can be represented by other methods other than the landmark-based representation. In the remainder of this subsection we review two main tests that are based on the outline-based shape representation where the shape outline is a set of k measurements (also called landmarks) along the surface of the object. Lele and Richtsmeier (1991) suggested a considerably different type of test for the two sample case called EDMA-I (for Euclidean Distance Matrix Analysis) for determining differences in mean shape. They define the *form matrix* $\mathbf{F}(\mathbf{X})$ as the symmetric Euclidean distance matrix between each pair of landmarks in a configuration \mathbf{X} . They also define the *average form difference* matrix of two configurations \mathbf{X} and \mathbf{Y} to be:

$$D_{ij}(\mathbf{E}(\mathbf{X}), \mathbf{E}(\mathbf{Y})) = F_{ij}(\mathbf{E}(\mathbf{X})) / F_{ij}(\mathbf{E}(\mathbf{Y})) \quad (2.18)$$

where D_{ij} denotes the ij -entry of the average form distance matrix \mathbf{D} and $\mathbf{E}(\cdot)$ is the expectation operator. Given n configurations $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ in a first sample and another n configurations $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ in a second sample, the EDMA-I test statistic is computed as follows:

Step 1: Estimate the mean shapes $\mathbf{E}(\mathbf{X})$ and $\mathbf{E}(\mathbf{Y})$. In their first paper about EDMA-I, Lele and Richtsmeier (1991) suggested to use GPA to estimate the mean shapes but later Lele (1993) suggested to use instead Multidimensional Scaling (MDS) for mean shape estimation.

Step 2: Calculate the mean form matrices $\mathbf{F}(\mathbf{E}(\mathbf{X}))$ and $\mathbf{F}(\mathbf{E}(\mathbf{Y}))$ which are simply the Euclidean distance matrices between the landmarks of the estimated mean shapes.

Step 3: Find the average form difference matrix \mathbf{D} using (2.18). Finally, calculate

the EDMA-I test statistic defined by:

$$T = \max \mathbf{D}_{ij}(\mathbf{E}(\mathbf{X}), \mathbf{E}(\mathbf{Y})) / \min \mathbf{D}_{ij}(\mathbf{E}(\mathbf{X}), \mathbf{E}(\mathbf{Y})). \quad (2.19)$$

Step 4: Generate the null distribution of T by permuting the two samples P times and repeating steps 1 through 3 at each permutation. If r permutations result in $T > T^{obs}$, then the p-value of the test is r/P .

The EDMA-I test requires that the variance-covariance matrices be equal for both samples. However, to relax this assumption Lele and Cole (1996) later proposed an alternative test, EDMA-II, for shape differences that does not require equal variance-covariance matrices. The details of the EDMA-II test procedure are as follows:

Step 1: Given two configuration samples $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ and $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ calculate the mean form estimates \mathbf{F}_1 and \mathbf{F}_2 as described in steps 1 and 2 above. Use Lele's (Lele and Cole, 1996) proposed MDS method to estimate for mean forms.

Step 2: Propose a scaling factor (e.g., any edge length or any continuous function of the edge lengths). Standardize \mathbf{F}_1 and \mathbf{F}_2 through dividing them by the scaling factor to obtain the *mean shape matrices* \mathbf{S}_1 and \mathbf{S}_2 :

$$\begin{aligned} \mathbf{S}_1 &= \mathbf{F}_1 / C_1 \\ \mathbf{S}_2 &= \mathbf{F}_2 / C_2, \end{aligned} \quad (2.20)$$

where C_1 and C_2 are the chosen scaling factors obtained from samples \mathbf{X} and \mathbf{Y} respectively.

Step 3: Calculate the EDMA-II test statistic, Z , given by:

$$Z = \min(\mathbf{S}_1 - \mathbf{S}_2) \quad \text{or} \quad \max(\mathbf{S}_1 - \mathbf{S}_2) \quad (2.21)$$

whichever is most different from zero.

Step 4: Generate the null distribution of Z by permuting the two samples P times and repeating steps 1 through 3 at each permutation. If r permutations result in $Z > Z^{obs}$, then the p-value of the test is r/P .

It is important to point out that all distance-based methods apply only to the one-way (one factor) two-levels ($a = 2$) case and this, in itself, is rather restrictive.

It is unclear how this type of tests can be extended to the case of more than one factor.

2.2 Modeling and Robust Parameter Optimization of Profile Response Experiments

Consider an experiment where the response of interest \mathbf{y} is a function of another variable, s , which we will call the “locations” at which the response is measured. In this dissertation, the locations can denote time, an actual physical location along a one dimensional axis (the more general case of higher dimensional locations will not be considered here), or, more generally, some physical scalar variable (such as rotational speed, see Figure 1.1, or temperature, see Figure 1.4). The shape or profile of the continuous function $\mathbf{y} = f(\mathbf{s})$ determines the performance of the system or process under study. This profile is in turn assumed to be modifiable by manipulating controllable factors \mathbf{x}_c , and is also affected by noise factors \mathbf{x}_n . Similarly to traditional RPD for scalar responses, the goal is to find the optimal values of \mathbf{x}_c that make the process robust or insensitive to variations in \mathbf{x}_n . Optimality in this case means achieving a specific target shape or profile for $\mathbf{y}(\mathbf{s}|\mathbf{x}_c)$ with minimum variability due to the existence of noise factors.

In engineering, profile responses are usually sampled at high frequency. Hence, response values sampled at nearby locations tend to be correlated. If the within-profile correlation is strong and neglected, model parameter estimates will be inefficient and will yield less precise predictions (Del Castillo et al., 2011). Inefficient estimators will in turn provide misleading “optimal” solutions when used to solve the RPD problem.

Many authors have considered profile responses when monitoring the quality in a production process (statistical process control) under the name “profile monitoring” (see, e.g., Kang and Albin, 2000; Kim et al., 2003). A more related work to the problem we consider in this dissertation has been done in the “Signal-Response Systems” (sometimes called Dynamic-Response Systems) literature (Miller, 2002; Miller and Wu, 1996; Box, 1988; Nair et al., 2002; Gupta et al., 2010). In signal-response systems literature, the measurement locations s receive the name “signal”

or “signal-factor”.

Two approaches to analyze signal-response systems were proposed by Miller and Wu (1996). These two methods are: the performance measure modeling (PMM) and the response function modeling (RFM). The PMM method is based on summarizing the signal-responses into univariate metrics and then analyzing these metrics. Box (1988) has shown how the PMM method can provide similar results for very different processes, and hence it is not recommended. The RFM method on the other hand is based on modeling the relationship between the signal (measuring locations) and the response in a parametric modeling framework. This approach makes intuitive sense, but should be implemented with care especially when the model parameters are correlated. Furthermore, it requires a linear statistical model for the mean that provides a good fit.

Other existing methodologies to analyze signal-response systems involve a two-stage modeling procedure:

1. Fitting a parametric model for the response as a function of the signal (the locations);
2. Modeling the parameters in the first step as a parametric model in design factors, both noise and control factors.

This type of methodology was followed by Nair et al. (2002) who use a location-dispersion model evaluated at the signal-factor levels such that at each controllable factor settings \mathbf{x}_{ci} and signal level s_j the location μ is

$$\mu(\mathbf{x}_{ci}, s_j) = \mathbf{x}'_{ci}\boldsymbol{\beta}(s_j) \quad (2.22)$$

and the log of the dispersion, σ^2 , is

$$\log \sigma^2(\mathbf{x}_{ci}, s_j) = \mathbf{x}'_{ci}\boldsymbol{\phi}(s_j) \quad (2.23)$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ are the model coefficients for the mean and the variance as a function of the signal, respectively. If noise factors \mathbf{x}_n are present, the following model is used

$$y_{ij} = \mu(\mathbf{x}_{ci}, \mathbf{x}_{ni}, s_j) + \sigma(s_j)\epsilon_{ij} \quad (2.24)$$

such that the location model is

$$\mu(\mathbf{x}_{ci}, \mathbf{x}_{ni}, s_j) = \mathbf{x}'_{ci}\boldsymbol{\beta}(s_j) + \mathbf{x}'_{ni}\boldsymbol{\gamma}(s_j) + \mathbf{x}_{ci}\boldsymbol{\Lambda}\mathbf{x}_{ni} \quad (2.25)$$

where $\boldsymbol{\gamma}$ are the effects of the noise factors and $\boldsymbol{\Lambda}$ the control-noise interaction effects. The dispersion effect is determined through the control-noise interactions. A more recent approach to model signal-response systems is that of Gupta et al. (2010). Their approach is based on modeling the signal-response mean using a Generalized Linear Mixed Model (GLMM). The use of GLMM overcomes the normality assumption of the errors, usually assumed in previous models. Just as for the models above, parametric (regression) models for the mean are needed for GLMM models.

We point out here that all of the previously mentioned approaches assume a small number of signal-factor levels since the estimation methods are frequentist. Hence, model parameters uncertainties are not included which may lead to non-optimal solutions. A more recent approach that considers both a large number of levels for the signal factor and the uncertainty in the model parameters is that of Del Castillo et al. (2011). They proposed a two-stage hierarchal mixed effects model to solve the RPD problem for profile response systems. The proposed approach considers the uncertainty in the model parameters by adopting a Bayesian point of view. Their proposed model is as follows:

$$\mathbf{y}_i = \mathbf{S}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \quad \text{vec}(\boldsymbol{\varepsilon}_i) \sim N_J(\mathbf{0}, \boldsymbol{\Sigma}), \quad (2.26)$$

$$\boldsymbol{\theta}_i = \mathbf{B}\mathbf{f}(\mathbf{x}_i) + \mathbf{w}_i, \quad \text{vec}(\mathbf{w}_i) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_w) \quad (2.27)$$

where \mathbf{y}_i is the i^{th} observed profile, \mathbf{S} is a matrix of regressors in the locations s_j , $\boldsymbol{\theta}_i$ and \mathbf{B} are regression coefficients, $\mathbf{f}(\mathbf{x}_i)$ is a function of design factor settings \mathbf{x}_i , and $\boldsymbol{\varepsilon}_i$ and \mathbf{w}_i are error terms. Substituting (2.27) into (2.26) we get

$$\mathbf{y}_i = \mathbf{S}(\mathbf{B}\mathbf{f}(\mathbf{x}_i) + \mathbf{w}_i) + \boldsymbol{\varepsilon}_i \quad (2.28)$$

$$= \mathbf{S}\mathbf{B}\mathbf{f}(\mathbf{x}_i) + \mathbf{S}\mathbf{w}_i + \boldsymbol{\varepsilon}_i \quad (2.29)$$

Here, $\mathbf{S}\mathbf{B}\mathbf{f}(\mathbf{x}_i)$ is a vector of size $J \times 1$. For three conformable matrices \mathbf{Z}_1 , \mathbf{Z}_2 and \mathbf{Z}_3 we have $\text{vec}(\mathbf{Z}_1\mathbf{Z}_2\mathbf{Z}_3) = (\mathbf{Z}'_3 \otimes \mathbf{Z}_1)\text{vec}(\mathbf{Z}_2)$ (Henderson and Searle, 1979)

where $vec(\cdot)$ is the operator that concatenates the matrix columns into a vector and \otimes is the Kronecker product. Applying this property to (2.29) we get

$$\mathbf{y}_i = (\mathbf{f}(\mathbf{x}_i)' \mathbf{S}) vec(\mathbf{B}) + \mathbf{S} \mathbf{w}_i + \boldsymbol{\varepsilon}_i. \quad (2.30)$$

Letting $\mathbf{F}_i = \mathbf{f}(\mathbf{x}_i)' \mathbf{S}$ and $\boldsymbol{\beta} = vec(\mathbf{B})$, this model can be written as:

$$\mathbf{y}_i = \mathbf{F}_i \boldsymbol{\beta} + \mathbf{S} \mathbf{w}_i + \boldsymbol{\varepsilon}_i \quad (2.31)$$

where $\mathbf{F}_i \boldsymbol{\beta}$ is the fixed effect term and $\mathbf{S} \mathbf{w}_i$ is a random effects term. Lange et al. (1992) and Chib and Carlin (1999) provide the full conditionals of the Bayesian posterior density of this model in closed form. However, as discussed by Ware (1985), model (2.31) is unnecessarily restricted, since the design matrices for the fixed effect and the random effect terms are linked (they both depend on \mathbf{S}). The random effects term $\mathbf{S} \mathbf{w}_i$ induces a within-profile correlation structure that may not be what the data shows. To better estimate the within-profile correlation, the use of different matrices \mathbf{S} and \mathbf{S}^* was suggested (Del Castillo et al., 2011). In their paper, Del Castillo et al. (2011) suggested to use the first p columns of \mathbf{F}_i to construct \mathbf{S}^* . A model selection criterion such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) is then used to select the best value for p .

As a more flexible alternative to the optimization of profile response processes, a Gaussian Random Function model is presented in chapter 5 of this dissertation.

Chapter 3

Statistical Performance of Tests for Mean Shape Difference with Application in Manufacturing

This chapter considers experiments in manufacturing where the response of interest is the geometrical shape of a part and the goal is to determine whether and how the experimental factors affect the resulting shape. The usual approach in practice is to estimate the form error of the part and conduct an ANOVA on the form errors. Instead, we study the performance of several SSA techniques to analyze this class of experiments.

This chapter extends the preliminary results found in Del Castillo and Colosimo (2011) for performance analysis of SSA tests (described in section 2.1.2) in several directions. The performance of the different SSA tests for shape difference are analyzed for 2D and 3D objects. For cases of interest in manufacturing where a definition of form error exists (circularity and cylindricity), the SSA tests are compared to the standard approach (ANOVA based on the form errors). There exists considerable debate about the merits of the different SSA techniques, and it is not clear which one to use in a given situation (Dryden and Mardia, 1998, section 12.2.5). One of the goals of this chapter is to elucidate this debate.

The chapter is organized as follows. In section 3.1 we study the performance of the shape difference tests for data sets of arbitrary geometries previously considered in the literature. Section 3.2, instead, extends these results to shapes frequently oc-

curing in manufactured parts, namely, 2D circular and 3D cylindrical shapes. The performance comparisons in sections 3.1 and 3.2 deal with the so-called *isotropic* case, i.e., the case where the variances at each landmark and along each dimension are the same. Section 3.3 studies the performance of mean shape difference tests in the *anisotropic* (or *non-isotropic*) case, including cases where different covariance structures are present. The performance of these tests under non normal errors is investigated in section 3.4. Summary and conclusions are provided in section 3.5.

3.1 Performance Analysis of Tests for Non-manufacturing Shapes of Arbitrary Ge- ometry, Isotropic Errors

The tests for detecting differences in mean shape studied in the past SSA literature consist mainly of instances of arbitrary polygons. These are not shapes commonly occurring in manufacturing but are included here to allow comparison between the SSA tests described in section 2.1.2 for the cases in the extant literature on SSA.

3.1.1 Triangles

The simplest non-trivial geometrical shape that has been studied in shape tests is the case of $k = 3$ landmarks in $m = 2$ dimensions, i.e., a triangle whose shape space is a 2-sphere (see section 2.1.1 and Appendix A), shown in Figure 3.1. This is the Σ_2^3 shape space, first studied by Kendall (1984). For two-dimensional (planar) data, Kendall coordinates are homogeneous coordinates in \mathbb{C} (the complex plane), given that a complex representation of each landmark coordinates simplifies the algebraic computations, especially the rotations (see Appendix A for more information about the mathematical notation and terminology used in shape theory). For k landmarks on the plane, the Kendall coordinates are:

$$u_j + iv_j = \frac{z_j - 1}{z_1}, \quad j = 3, \dots, k \quad (3.1)$$

In the case of triangles ($k = 3$) each triangle is parameterized by a single Kendall coordinate:

$$z_3 = u_3 + iv_3. \quad (3.2)$$

Since if $k = 3$, $\Sigma_2^3 = CP^1 = S^2(\frac{1}{2})$ (Hopf's submersion, see Appendix A), the shape space of triangles can be defined in terms of the coordinates:

$$x = \frac{1 - r^2}{2(1 + r^2)} \quad y = \frac{u_3}{1 + r^2} \quad z = \frac{v_3}{1 + r^2} \quad (3.3)$$

where $r^2 = |z_3|^2 = u_3^2 + v_3^2$ so that $x^2 + y^2 + z^2 = \frac{1}{4}$ which implies these points are on a sphere of radius $\frac{1}{2}$ ($S^2(\frac{1}{2})$).

The coordinates above are *extrinsic* to the sphere $S^2(\frac{1}{2})$, since they refer to the ambient \mathbb{R}^3 space where S^2 is embedded in. Kendall's spherical coordinates are instead *intrinsic* (see Appendix A) to the shape space:

$$\frac{1}{2} \sin \theta \cos \phi = \frac{1 - r^2}{2(1 + r^2)} \quad (3.4)$$

$$\frac{1}{2} \sin \theta \sin \phi = \frac{u_3}{(1 + r^2)} \quad (3.5)$$

$$\frac{1}{2} \cos \theta = \frac{v_3}{(1 + r^2)} \quad (3.6)$$

where the two angles correspond to the 'latitude' $\theta \in [0, \pi]$ and the 'longitude' $\phi \in [0, 2\pi]$.

The circular plot in Rohlf (2000) (see Figure 3.2 below) is called the Schmidt net or Lambert projection of the 2-sphere (see Dryden and Mardia, 1998, page 37). This projection from $S^2(\frac{1}{2})$ to \mathbb{R}^3 preserves areas and is defined by the angles (ξ, ψ) such that

$$\xi = 2 \sin \left(\frac{\theta}{2} \right), \quad 0 \leq \xi \leq \sqrt{2} \quad (3.7)$$

$$\psi = \phi, \quad 0 \leq \psi \leq 2\pi \quad (3.8)$$

which result in a circle whose center is an equilateral triangle (i.e., the projection is on a plane on the 'north pole' of $S^2(\frac{1}{2})$). The Schmidt-Lambert projection, shown in Figure 3.2, is useful to study the power of a statistical test for detecting

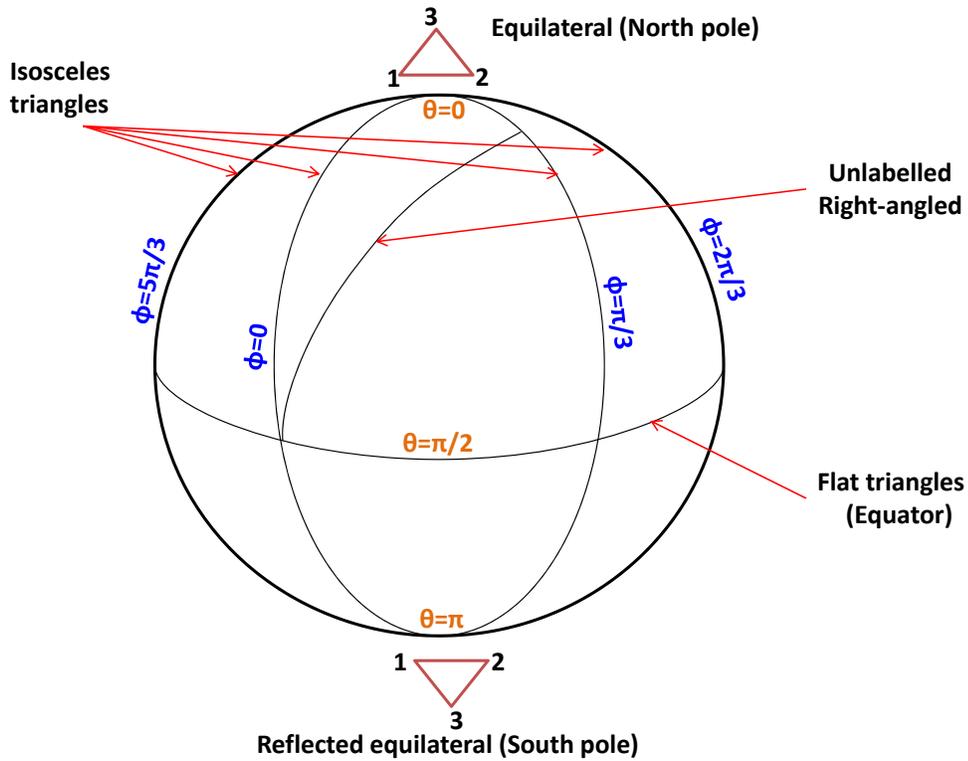


Figure 3.1. Kendall's shape space for triangles in $m = 2$ dimensions. This is $S^2(\frac{1}{2})$, a 2-sphere with radius $\frac{1}{2}$. The shape coordinates are the latitude θ and the longitude ϕ in equations (3.4-3.6). Adapted from Dryden and Mardia (1998).

significant deviations from an equilateral triangle, as done by Rohlf (2000) and by us in what follows.

Since the shape space for triangles (Figure 3.1) 'repeats' itself 6 times over $S^2(\frac{1}{2})$, forming six 'lunes', it is customary to show the Schmidt projection of $S^2(\frac{1}{2})$ only for one of these lunes ((see figure 28 in Dryden and Mardia, 1998)). Also since points with $\theta > \pi/2$ are reflections of points for $\theta < \pi/2$, it is customary to display the Schmidt projection only for the upper hemisphere.

Figure 3.2 shows the Schmidt projection (Dryden and Mardia, 1998) of the shape space of triangles ($S^2(\frac{1}{2})$) for the upper "hemisphere" of Figure 3.1, to avoid reflections. In the figure, When $\theta = 0$, we have an equilateral triangle; when $\theta = \pi/2$ the triangle degenerates into a line, which corresponds to points on the

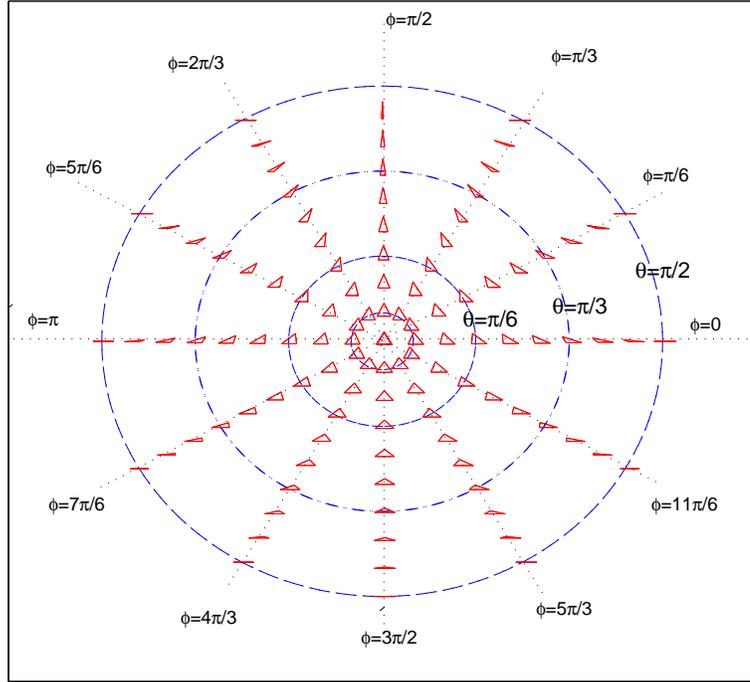
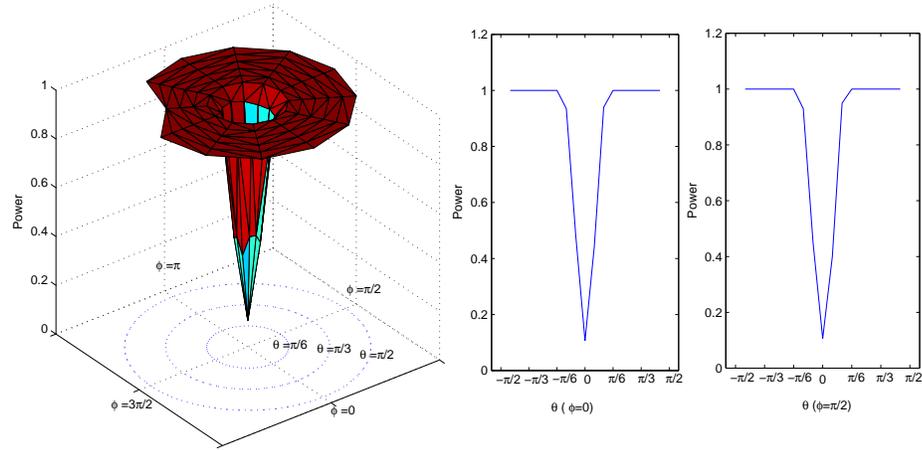


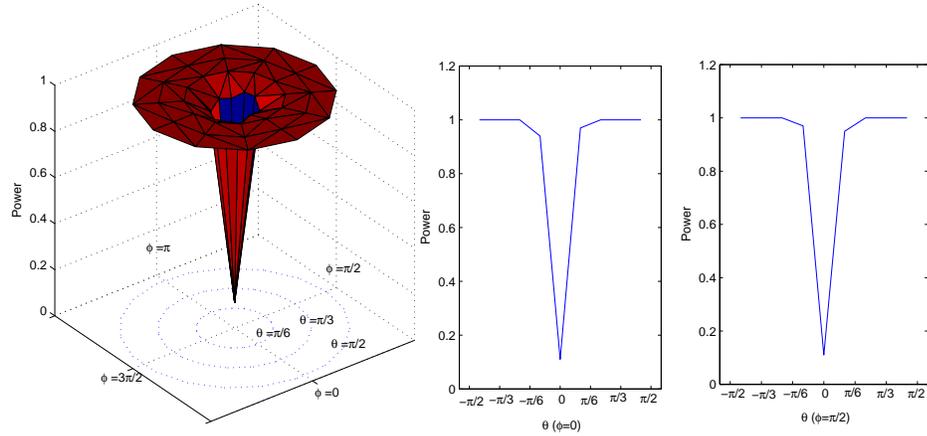
Figure 3.2. A representation of different equilateral triangles projected on the tangent plane shown in Figure 3.1 as a function of the rotational angles ξ and ψ , given by equations (3.7 and 3.8).

equator of Figure 3.1. When $\phi = \pi/2$, we have an isosceles triangle; when $\phi = \pi$ the isosceles triangle becomes right-angled, when $\theta > \pi/2$ the triangles are simply reflected, a case we do not consider here.

The power surface of the ANOVA Permutation test under different possible values of (θ, ϕ) was computed. For each point on the power surface, two ($a = 2$) samples were compared: one with parameters $\theta = \phi = 0$ (equilateral triangle) and a second sample whose (θ, ϕ) values were taken within the ranges $(-\pi/2, \pi/2)$ and $(0, 2\pi)$ respectively. The simulation setup was chosen to be the same as in Rohlf (2000) in order to have comparable surfaces. Our results are shown in Figure 3.3, which displays also the ANOVA F-test power surface previously obtained by Rohlf. It can be seen that the ANOVA Permutation test behaves for this case very closely to the ANOVA F-test in terms of power and type-I error rate.



(a) Power surface of the ANOVA Permutation test for detecting deviations from an equilateral triangle as a function of rotational angles θ and ϕ .



(b) Power surface of the ANOVA F-test for detecting deviations from an equilateral triangle as a function of rotational angles θ and ϕ (see Rohlf, 2000).

Figure 3.3. Power surface for detecting differences in shape between triangles using the ANOVA Permutation test and Goodall's ANOVA F-test, isotropic variance. The simulation parameters were: error standard deviation σ was set at 0.15, the test size α was set at 0.1, twenty ($= n$) configurations per sample were used, each test was replicated one hundred times, and one hundred permutations were performed. θ is restricted to $[-\pi/2, \pi/2]$ to exclude reflections. On the right, the marginal power plots for fixed ϕ .

3.1.2 Performance of Tests for Arbitrary Polygons (Rohlf, 2000; Coward and McConathy, 1996)

Rohlf (2000) gives one landmark data set, depicted in Figure 3.4(a), used to compare the power of tests for shape differences for the case of $a = 2$ samples. The ANOVA F-test (eq. 2.14), the ANOVA Permutation test and the distance-based tests developed by Lele (EDMA-I and EDMA-II, eqs. (2.19) and (2.21)) were compared using simulation for the shapes in this figure. The simulation setup was as follows: the test size α was set to 0.05, one hundred ($= n$) configurations per sample were used, each test was replicated 1000 times, and 150 permutations for each test that requires permutation were performed. $N(0, \sigma^2)$ *i.i.d* errors were added to each coordinate at each landmark and both $\sigma^2 = 1$ and $\sigma^2 = 1.5^2 = 2.25$ were tested. To estimate the type-I error rate, the mean shape under the null hypothesis was equal to the shape labeled sample 1 in Figure 3.4(a). The scaling factor for EDMA-II was chosen to be the total sum of distances between consecutive landmarks. As summarized in Table 3.1, no test failed to detect the difference in mean shape when $\sigma = 1$, but as expected, the power decreases as σ increases without affecting the type-I error rate.

Table 3.1. Performance of tests for shape difference for the data set provided by Rohlf (Figure 3.4(a)). Results were obtained using a test size α of 0.05, one hundred configurations ($= n$) per sample. Each test was replicated one thousand times, and 150 permutations for the ANOVA Permutation and EDMA tests. Numbers in parentheses are the standard errors of the corresponding simulation estimates.

Test	Type-I error	Power when $\sigma = 1$	Power when $\sigma = 1.5$
ANOVA F-test	0.04(0.04)	1(0)	0.92(0.0271)
ANOVA Permutation test	0.042(0.0063)	1(0)	0.99(0.0099)
EDMA-I	0.042(0.0063)	1(0)	0.93(0.0255)
EDMA-II	0.042(0.0063)	1(0)	0.93(0.0255)

Another data set published by Rohlf (2000) for 3-dimensional shapes can also be used to study the performance of the different tests. The data set consists of two samples with configurations of four landmarks in 3D space. The two mean shapes are depicted in Figure 3.4(b). Simulations were carried out using the same parameter settings as in Rohlf (2000), namely: a test size α of 0.05, one hundred

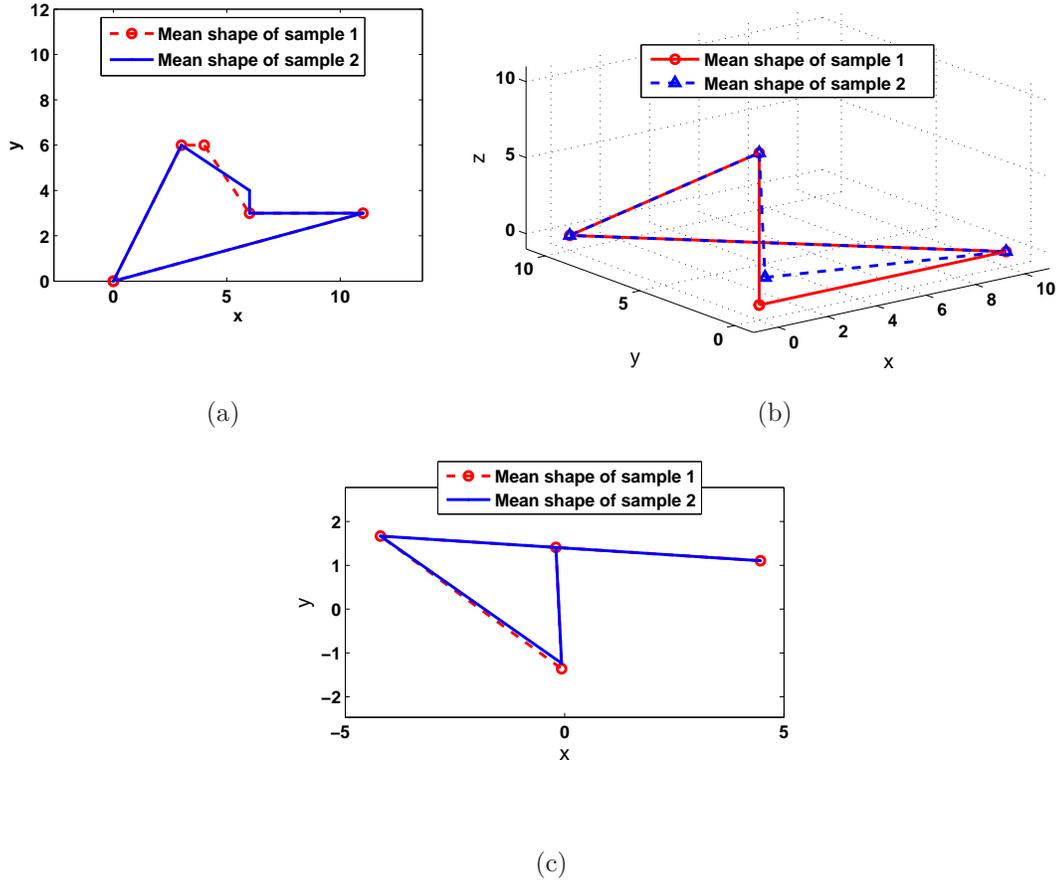


Figure 3.4. (a) The two mean shapes of the 2D data set used by Rohlf (2000). (b) The two mean shapes of the 3D data set used by Rohlf (2000). (c) The two mean shapes of the 2D data set used by Coward and McConathy (1996).

configurations per sample ($= n$), each test was replicated one thousand times, and one hundred permutations for the ANOVA Permutation and EDMA tests were performed. The simulations were performed at two values of the σ parameter, namely, $\sigma = \{1, 1.5\}$. The results are summarized in Table 3.2 and show that the ANOVA Permutation test is the most powerful while providing the closest to advertised type-I error probability.

Notice that in Tables 3.1 and 3.2 the ratio of the standard deviation of the type-I error rate for the ANOVA F-test to the type-I error rate itself is large compared to the same ratio for other tests. As seen in Figures 3.4(a) and 3.4(b), the two mean shapes in this case differ largely from each other and from the grand mean

Table 3.2. Performance of tests for shape difference for the data set of Table 3 in Rohlf (2000) (3-dimensional shape). Results were obtained using: the test size α was set at 0.05, one hundred configurations per sample ($= n$) were used. Each test was replicated one thousand times, and 150 permutations were performed for the ANOVA permutation test and the EDMA tests. Numbers in parentheses are the standard errors of the corresponding simulation estimates.

Test	Type-I error	Power when $\sigma = 1$	Power when $\sigma = 1.5$
ANOVA F-test	0.003(0.0017)	0.704(0.0144)	0.22(0.0414)
ANOVA Permutation test	0.05(0.0069)	0.937(0.0077)	0.55(0.0497)
EDMA-I	0.048(0.0068)	0.401(0.0155)	0.04(0.0196)
EDMA-II	0.048(0.0068)	0.401(0.0155)	0.04(0.0196)

shape ($\mathbf{X}_{\bullet\bullet}$). Large differences between the mean shapes and the grand mean shape affect the distribution of the sum of squares in Table 2.1 making them *not* follow the assumed χ^2 distributions (see section 3.1.3 for more details) and hence, this affects the type-I error rate.

Coward and McConathy (1996) studied the effect of the sample size n on the power and type-I error rates of the EDMA-I and Goodall's F-test. Figure 3.4(c) shows the 2D mean shapes they considered. For these shapes the authors reported that the EDMA-I test has an inflated type-I error rate which decreases by increasing the sample size. We observed a similar behavior when applying the ANOVA Permutation test. Our results, shown in Figure 3.5, indicate that increasing the sample size decreases the type-I error rate until it converges to the nominal value.

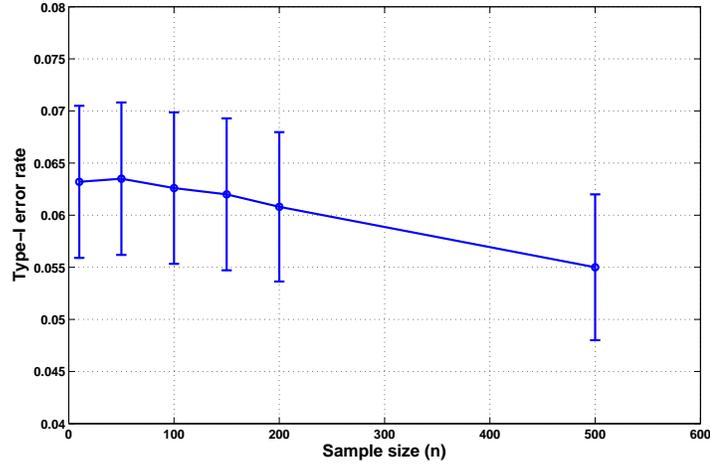


Figure 3.5. Simulated type-I error rate of the ANOVA Permutation test using the data set in Coward and McConathy (1996).

3.1.3 The Assumption of a Small Error Variance (σ^2) in the F-test

Goodall's F-test requires that the magnitude of the standard deviation of the error σ be small. Otherwise, the distribution of the test statistic in (2.14) is not an F, since the Procrustes sums of squares are approximately distributed as a χ^2 only when σ is small, and as was mentioned in section 2.1.1, the ANOVA identity breaks down when σ is large. This is because the analysis on the tangent space of the pre-shape space becomes a worse linear approximation of the nonlinear pre-shape space when σ is large, i.e., when shapes differ considerably. Dryden and Mardia (1998) indicate that a value of $\sigma \leq 0.1$ is small enough for this approximation to hold well but do not provide any empirical evidence behind this assertion. Therefore, to understand how the distribution of the scaled procrustes G statistic deviates from a χ_M^2 as σ increases, a simulation study was conducted. In the simulation, $n = 100$ parts each with $k = 20$ landmarks following a circular mean shape were generated to which *i.i.d.* $N(0, \sigma^2)$ noise was added at each coordinate. It was found (see Figure 3.6) that as σ increases beyond 0.1 (for unit size scaled objects, that is, $\mathbf{X}_{scaled} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$ where $\|\mathbf{X}\| = \sqrt{trace(\mathbf{X}^T \mathbf{X})}$) the distribution begins to deviate noticeably from a χ_M^2 , the assumed distribution in the ANOVA F-test.

Specifically, the distribution of G thins at the tails and begins to move to the left of a χ^2_M distribution. This confirms Dryden and Mardia's suggestion.

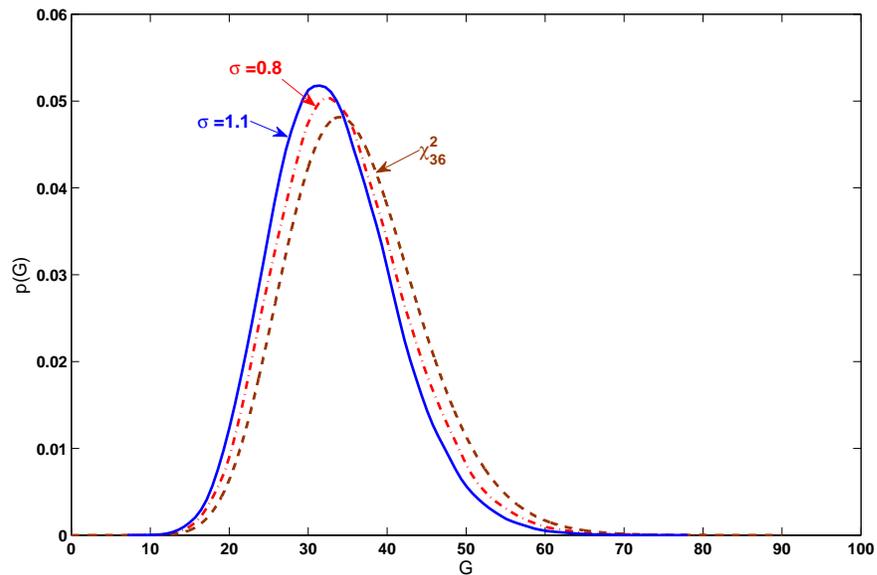


Figure 3.6. The distribution of the scaled G statistic as σ changes.

To show numerically how the ANOVA identity breaks down when σ is large, a simulation was conducted for one factor at two-levels. The mean shape for each of the two samples is shown in Figure 3.7 below. Each of these two mean shapes has 64 landmarks. Ten configurations were simulated for each sample where normal i.i.d errors with standard deviation σ were added at each of the 64 landmarks. Table 3.3 below shows the sum of squares values at different values of σ . This table clearly shows that when σ is large $SS_{Total} \neq SS_{Factor} + SS_{Error}$.

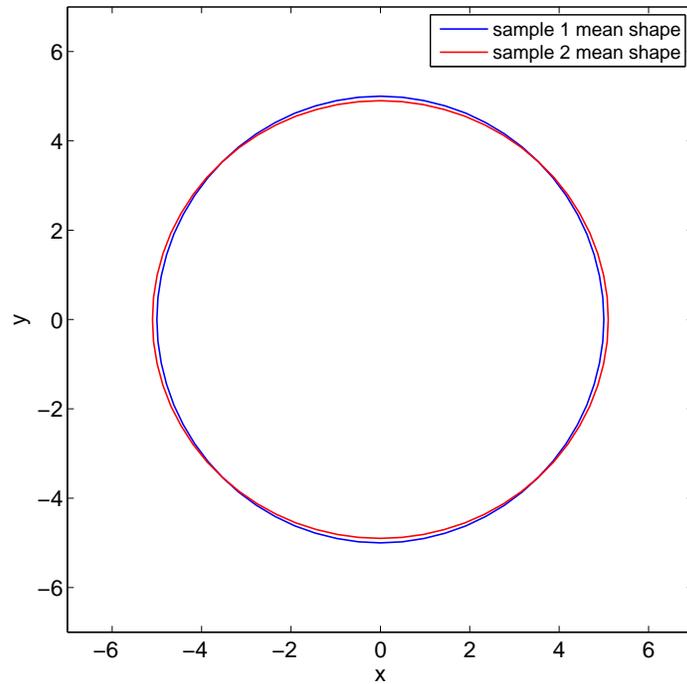


Figure 3.7. The two mean shapes used in the ANOVA identity calculations simulation study.

Table 3.3. Sum of squares for the two-sample simulation study of the ANOVA identity. As σ increases, the identity breaks down.

σ	SS_{Factor}	SS_{Error}	$SS_{Factor} + SS_{Error}$	SS_{Total}
0.1	0.0018	0.0144	0.0162	0.0162
0.3	0.0083	0.1305	0.1388	0.1387
0.5	0.0211	0.333	0.3541	0.3538
1	0.0794	1.2133	1.2927	1.2878
1.5	0.1701	2.6265	2.7966	2.7739
2	0.2898	4.0218	4.3116	4.2526
2.5	0.6198	6.2539	6.8737	6.6772

3.2 Performance Analysis for Shapes with a Geometry of Interest in Manufacturing, Isotropic Errors

The triangles and arbitrary polygons studied in section 3.1 are not common shapes found in manufacturing. In this section we consider the performance of shape analysis tests for two types of shapes widely found in many manufacturing processes, namely, 2D circles and 3D cylinders, under the assumption of isotropic errors.

3.2.1 Detection of Differences in Mean for 2D Circular Parts

Circular shapes are common in manufacturing, specially in parts that are produced using lathe or drill machine tools. Recent applications in micro-machining require analysis of circular shapes, too. Factors such as cutting speed, feed rate, etc. can affect the circularity of these parts. Due to the employment of CMMs in manufacturing, hundreds of measurements (landmarks) can be taken from a single part. A simulation study was conducted to determine the performance of the tests considered in the previous section for determining differences in mean shape for circular shapes. In this case, the usual ANOVA on the (circularity) form error is available, and was included in the comparisons as well. International standards (ANSI Y14.5M-1982) exist for measuring form errors in circular (or cylindrical) geometries. These are frequently calculated in tolerancing practice as the smallest difference between the radii of two coaxial circles (or cylinders) that enclose all the measurements in a part (Henzold, 2006; Krulikowski, 1996, see Figure 1.5). A standard one-way ANOVA is then used to test for equal means of form errors for the two samples.

Circular shape data that are less circular as the level of a single factor increases was simulated. As the value of the controllable factor changes from low to high, a second harmonic with amplitude δ was added to the circular profiles to simulate bilobed shapes (a common problem in lathe machining, see Figure 3.8). Amplitude values were chosen such that $\delta = w\sigma/r$, where σ is the noise (error) standard deviation, r is the radius of the true circle (set equal to 5) and w is the “non-

circularity” parameter $w = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ we desire the tests to detect. Figure 3.8 shows the mean shapes for two values of δ . The simulation setup was as follows: 20 configurations per sample were simulated each having $k = 64$ landmarks. One hundred replications and 100 permutations (for tests that require them) were conducted. The test size $\alpha = 0.05$, and $N(0, 0.05^2)$ *iid* errors were added at each coordinate and at each landmark and a fixed radius $r = 5$. The scaling factor for EDMA-II was chosen to be the circumference of the circle, which is the total sum of the distances between the consecutive landmarks. The simulation results are shown in Figure 3.9. The power curves for the F-test, ANOVA form error and ANOVA Permutation tests for this specific case were found previously by Del Castillo and Colosimo (2011). Comparing the curves, it is seen that the ANOVA Permutation test possesses as good power as the ANOVA F-test, while the EDMA-I test has very low power for this type of shapes. The ANOVA on the form errors performed considerably worse than the ANOVA F-test and the ANOVA Permutation test.

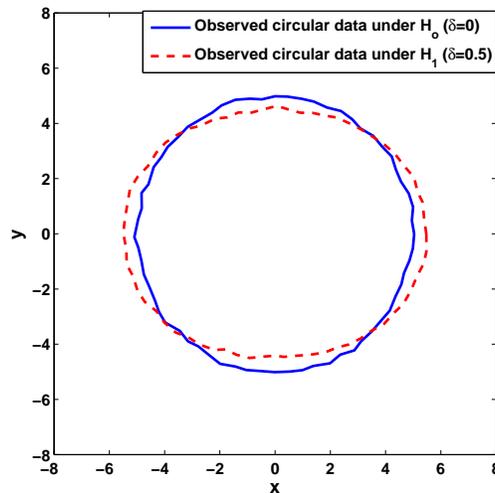


Figure 3.8. Circular and bilobed profiles with parameters $\sigma = 0.05$ and $r = 5$.

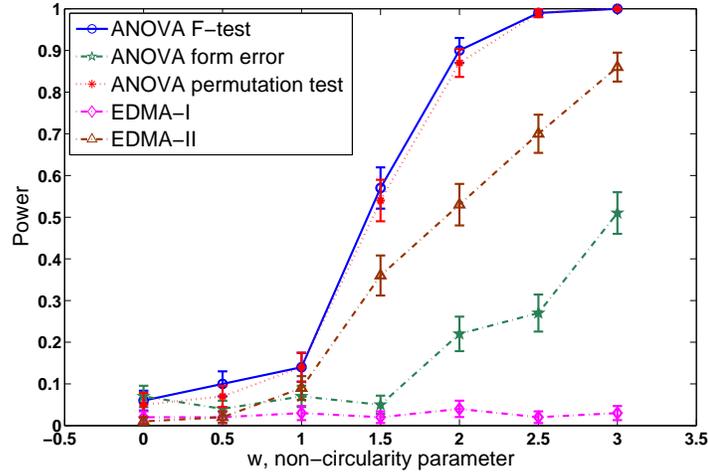


Figure 3.9. Power curves using 2D Circular profiles.

3.2.2 Detection of Differences in Mean Shape for 3D Cylindrical Parts

Since manufactured parts are of three dimensions, it is of interest to study the power of the statistical shape tests described earlier in the case $m = 3$. A common instance of a 3D shape of interest in manufacturing is a cylinder (see e.g., Traband et al., 2004). For example, in lathe machining, different problems in the process can result in non-cylindrical shapes in the form of a “barrel”, a “banana”, etc. (see Colosimo et al., 2007).

We consider a cylinder that can degenerate into a “barrel” shape in our performance study of 3D tests for shape difference (see Figure 3.10). The study consisted of two samples: the first sample contains parts with a cylindrical mean shape of radius r and height h ; the parts in the second sample have a barrel mean shape with radius $r_1 = r$ at both the top and the bottom and radius $r_2 > r_1$ at the middle of the cylinder, which has height h . In analogy to the circularity case, we refer to the difference $\delta = r_2 - r_1$ as the “amplitude”. The change from r_1 to r_2 along the height was considered to be a sine wave with amplitude δ and a period of $2h$. The amplitude δ can be calculated as $\delta = w\sigma/r_1$ where w is a “non-cylindricity” parameter, in analogy to the non-circularity parameter described in 3.2.1. Figure 3.10 shows the mean shape of the two samples when $\delta = \{0, 1\}$.

Twenty ($n = 20$) parts were generated at each value of the non-cylindricity parameter $w = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$. A total of 320 landmarks were simulated at five levels of the cylinder height $\{0, h/4, h/2, 3h/4, h\}$. Independent and identically distributed $N(0, 0.05^2)$ errors were added at each coordinate and at each landmark. To estimate the power of the EDMA-II test the scaling factor was chosen to be the sum of the five circumferences of the five circular cross sections along the height of the cylinder. The power results are shown in Figure 3.11. Comparing the power curves it can be said that the ANOVA F-test and the ANOVA Permutation test are the most powerful and control type-I error well, while the EDMA-I, EDMA-II and the ANOVA on the cylindricity form error (computed analogously to the circularity error by using two concentric cylinders, see Krulikowski (1996) and Figure 1.5) all have very low power.

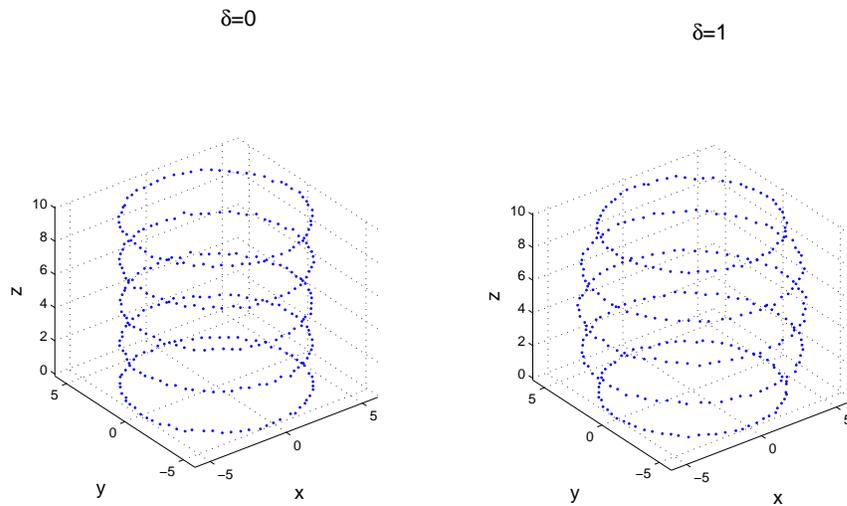


Figure 3.10. Sketch of two different mean shapes of cylindrical parts with $r = 5$ and $h = 10$. Left: perfect cylinder ($\delta = 0$), right: “barrel” ($\delta = 1$).

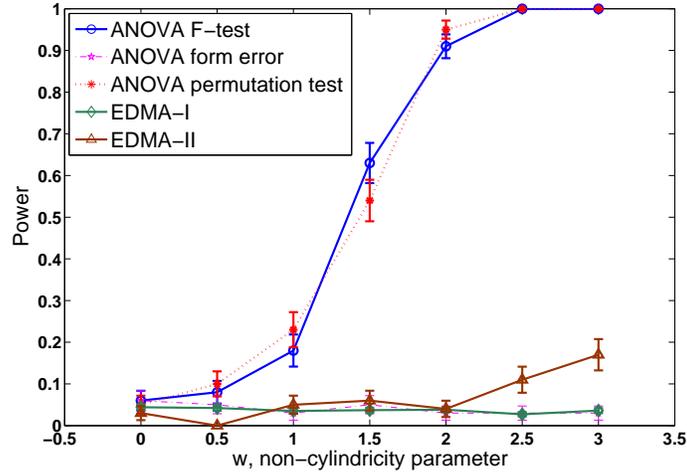


Figure 3.11. Statistical power for detecting mean shape difference in cylinders for the F-test, the Permutation ANOVA test, the ANOVA form error method, the EDMA-I and the EDMA-II tests. Simulation parameters were: 320 landmarks ($= k$) per configuration, test size α of 0.05, error standard deviation σ of 0.05, each test was replicated one thousand times, and 150 permutations for each test requires permutation.

3.2.3 Two Way ANOVA Tests for Shapes; Test for Interaction Effect

Del Castillo and Colosimo (2011) extended Goodall’s ANOVA F-test to a two-way ANOVA test with interaction, see section 2.1.4. They also provided a two-way ANOVA permutation test as a more robust alternative when the distributional assumptions required by the ANOVA F-test do not hold. In this test, main effects permutations are performed based on the suggestions by Edgington (1995) and the interaction effect is computed following recommendations given by Jung et al. (2006). In this section, we compare the power of the ANOVA Permutation test and the ANOVA F-test (only tests that have been extended to the two-way case) for detecting an interaction effect. Suppose there are two factors, A and B, each varied at two levels $\{-1, 1\}$ that affect the circularity of a 2D circle. Let $\delta = w\sigma/r$ be the non-circularity parameter as described in subsection 3.1. Let δ change according to the following model:

$$\delta = \delta_1 A + \delta_2 B + \delta_3 AB \quad (3.9)$$

where A and B represent the levels of the two factors, either -1 or 1. Then to simulate an interaction effect on the mean shape, we changed the value of δ_3 while fixing the values of both δ_1 and δ_2 .

The parameter δ at each of the four cells of the two-way ANOVA table is shown in Table 3.4. Parameters δ_1 and δ_2 were fixed at σ/r and $-\sigma/r$ respectively with r equal to the radius of the true circle (5). The parameter δ_3 was changed according to $\delta_3 = w\sigma/r$ over the set of values $w = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ with $w = 0$ indicating no AB interaction effect.

Table 3.4. Circularity parameter δ at the four possible combinations of A and B.

		A	
		-1	1
B	-1	$\delta = -\delta_1 - \delta_2 + \delta_3$	$\delta = \delta_1 - \delta_2 - \delta_3$
	1	$\delta = -\delta_1 + \delta_2 - \delta_3$	$\delta = \delta_1 + \delta_2 + \delta_3$

The EDMA-I and EDMA-II tests were not considered here since they are defined for one factor (two-levels) problems only. One hundred replications were performed for each test, and 100 permutations were used in the ANOVA Permutation test. Ten configurations per combination were allocated and the size of the tests (α) was set to 0.05. The error variance was set to 0.05^2 . The simulation results are shown in Figure 3.12. As it can be seen, under the ideal case of normal i.i.d. errors the ANOVA Permutation test and the ANOVA F-test provide the same type-I error rate. They also have very similar power for detecting a two-factor interaction effect on the mean shape of circular data.

3.3 Performance of Tests for Difference in Shape Under Non-isotropic Variance (2D and 3D)

Recall from section 2.1 that the usual assumed model in SSA is:

$$\mathbf{X} = \beta(\boldsymbol{\mu} + \mathbf{E})\boldsymbol{\Gamma} + \mathbf{1}_k\boldsymbol{\gamma}^T, \quad \mathbf{E} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (3.10)$$

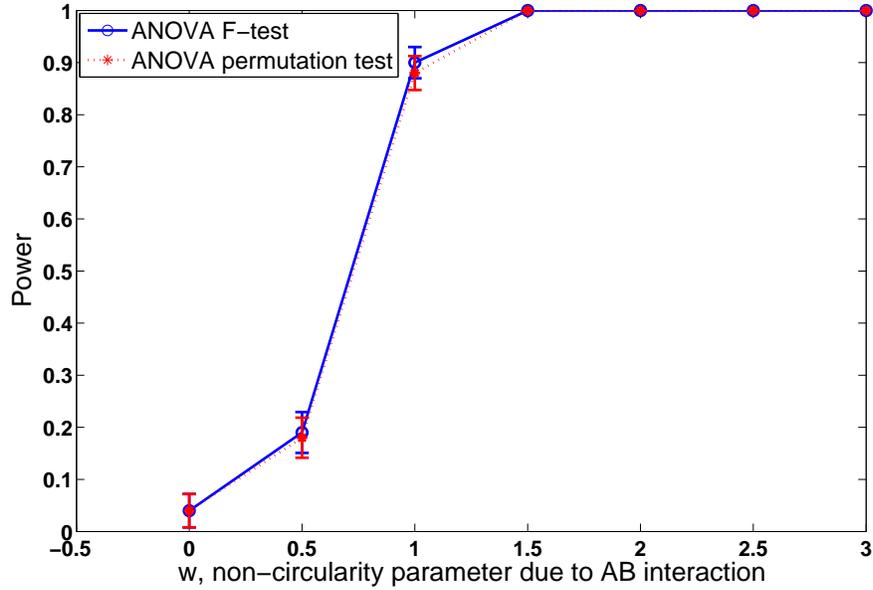


Figure 3.12. Power results for detecting an interaction effect in a two-way shape ANOVA for circular data.

Goodall (1991) suggested to decompose the variance-covariance matrix of the errors Σ into two parts, the $(k \times k)$ landmarks covariance matrix Σ_K , and the $(m \times m)$ covariance matrix Σ_D . Then Σ is the Kronecker product of Σ_K and Σ_D :

$$\Sigma = \Sigma_K \otimes \Sigma_D. \quad (3.11)$$

This is a plausible model in manufacturing data, where it is reasonable to assume that $\Sigma_D = I_m$ since the measurements acquired with a CMM can be considered to have independent errors in each direction (axis) and to have about the same variance. Furthermore, Procrustes methods do not provide an explicit mean shape when $\Sigma_D \neq I$ (Goodall, 1991; Lele, 1993). For this reason, we confine our simulations in this section only to the case when $\Sigma_D = I$.

In this section, three different models for the covariance matrix Σ_K are considered. The first model simply assumes that the landmarks are divided into g groups with $g \leq k$. Each group has a different variance which is constant within

the group and no correlation is assumed. Then $\Sigma_{\mathbf{K}}$ can be written as:

$$\Sigma_{\mathbf{K}} = \sigma^2 \mathbf{I} + \mathbf{C} \quad (3.12)$$

where \mathbf{C} is a $k \times k$ diagonal matrix with diagonal elements that vary giving different variances to the g groups. Note that for the isotropic case $\mathbf{C} = \mathbf{0}$.

In the second covariance model, the variance between landmarks is constant but some covariance is allowed. An exponential covariance structure is a common model in the Spatial Statistics literature (Banerjee et al., 2004, section 2.1.3). It considers the covariance between every two landmarks i and j as a function of the distance between them. Then $\Sigma_{\mathbf{K}}$ can be written as:

$$(\Sigma_{\mathbf{K}})_{ij} = \sigma^2 \exp(-\phi d_{ij}) + \tau^2 I(i = j), \quad \sigma^2 > 0, \phi > 0, \tau^2 > 0 \quad (3.13)$$

where $(\Sigma_{\mathbf{K}})_{ij}$ is the ij entry in matrix $\Sigma_{\mathbf{K}}$, d_{ij} is the Euclidean distance between landmark i and landmark j , $I(i = j)$ is an indicator function which equals 1 whenever $i = j$ and 0 otherwise and σ^2 , ϕ and τ^2 are the *sill*, *range* and *nugget* parameters of the covariance function (notice that when $d_{ij} = 0$ then $(\Sigma_{\mathbf{K}})_{ij} = \sigma^2 + \tau^2$ and that is the variance at that point since $i = j$). We note how this model is isotropic, since the variances are the same at each landmark.

In the third and last covariance model geometric anisotropy is considered. This model is used in the Spatial Statistics literature (Ecker and Gelfand, 1999). The model uses an $m \times m$ positive definite matrix \mathbf{B} to account for the unequal correlations, and can be written as:

$$(\Sigma_{\mathbf{K}})_{ij} = \sigma^2 \exp(-\phi(\mathbf{d}'_{ij} \mathbf{B} \mathbf{d}_{ij})^{1/2}) + \tau^2 I(i = j), \quad \sigma^2 > 0, \phi > 0, \tau^2 > 0 \quad (3.14)$$

where \mathbf{d}_{ij} is an $m \times 1$ vector of Euclidean distances along each of the m dimensions between landmarks i and j . This model provides geometric anisotropy since the correlation between any two landmarks depends on the separation vector \mathbf{d}_{ij} rather than merely on its length. As an example, consider three landmarks x, y and z in a two dimensional space, with $x = (0, 0)$, $y = (0, 5)$ and $z = (5, 0)$. The Euclidian distance $d_{xy} = 5 = d_{xz}$. Assuming $\sigma^2 = 1, \phi = 1, \tau^2 = 0.5$ and $B =$

$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$ we get $cov(x, y) = 0.5 + exp(-37.5)$ and $cov(x, z) = 0.5 + exp(-25)$. This example illustrates how in this model a pair of points equally distanced from a third reference point can have different correlations with that reference point.

An iterative weighted least squares (WLS) procedure to estimate the mean shape when $\Sigma_K \neq \mathbf{I}$ and $\Sigma_D = \mathbf{I}$ was suggested by Goodall (1991). Although no theoretical guarantee of convergence exists for this algorithm, our tests indicate that $GPA(\Sigma)$ is able to correctly register the shapes for all the non-isotropic error models considered herein (an instance of this can be seen in Figure 3.13). We noticed that $GPA(\Sigma)$ converges in all cases we tested but in a larger number of iterations than ordinary GPA.

Under the first covariance model described by equation (3.12), we set $g = 3$ with group variances of C_{low} , C_{med} and C_{high} . Two dimensional circular shapes and three dimensional cylindrical shapes were generated as described in previous sections. Normal random errors were added to each landmark according to the diagonal values of $\sigma^2\mathbf{I} + \mathbf{C}$. Power results for all five tests of interest are shown in Figures 3.14 and 3.15. Comparing to Figures 3.9 and 3.11, it can be seen that there is a dramatic loss in power for the ANOVA Permutation test and the ANOVA F-test once the equal landmark variance assumption is violated, but the power increases as more landmarks per configuration are considered.

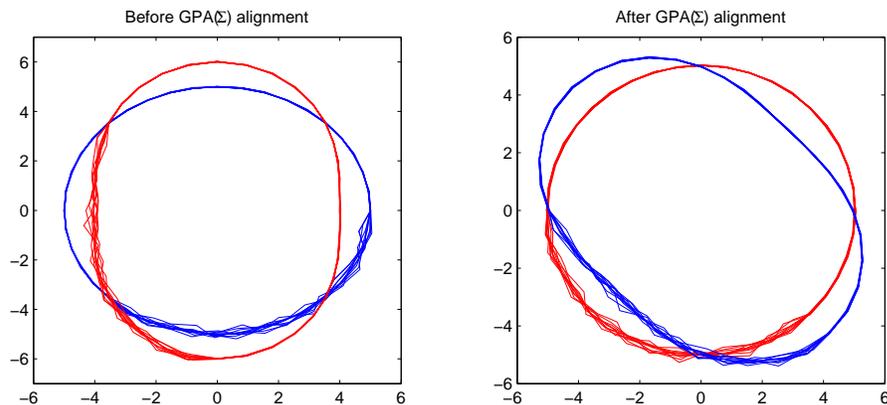


Figure 3.13. Mean shape estimated by the iterative $GPA(\Sigma)$ algorithm applied to 2D circular shapes with non-isotropic error structure (model 1 in equation (3.12) with $g = 2$ groups were used). Left: before $GPA(\Sigma)$ alignment, right: after $GPA(\Sigma)$ alignment.

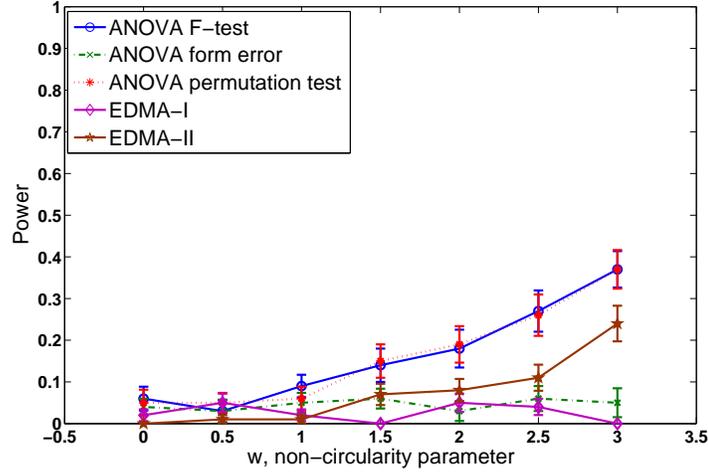


Figure 3.14. Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under non-isotropic error structure described by equation (3.12). Simulation parameters were: 64 landmarks ($= k$) per configuration, test size α of 0.05, parameters $\{C_{low}, C_{med}, C_{high}\}$ were set to 0.001, 0.005 and 0.01 respectively, each test was replicated one hundred times, and 100 permutations for each test that requires permutations were performed.

Under the second covariance model described by equation (3.13) we set σ, ϕ and τ to be 0.05, 0.05 and 0.05 respectively. Two dimensional circular shapes and three dimensional cylindrical shapes were generated. Correlated $N(\mathbf{0}, \Sigma)$ errors with covariance matrix $\Sigma = \Sigma_K \otimes \Sigma_D$ were generated and added to each landmark. Power results for all five tests of interest are shown in Figures 3.16 and 3.17. It can be seen that adding correlations reduces the power of the ANOVA Permutation test and the ANOVA F-test, but their power apparently increases as we move from the 2D to the 3D case. In reality, there are more landmarks per configuration in the 3D case than in the 2D case, and this explains the increase in power.

Finally, under the third covariance model described by equation (3.14) we set σ, ϕ and τ to be 0.05, 0.05 and 0.05 respectively. Matrix \mathbf{B} was chosen to include non-isotropic correlations (see Ecker and Gelfand, 1999). For the 2D circular shapes and the 3D cylindrical shapes, \mathbf{B} was set at $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$ and

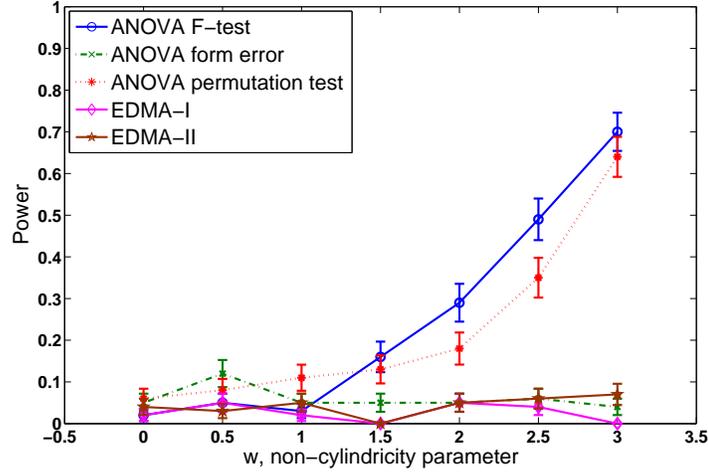


Figure 3.15. Statistical power for detecting changes in 3D cylindrical shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under non-isotropic error structure described by equation (3.12). Simulation parameters were: 320 landmarks ($= k$) per configuration, test size α of 0.05, parameters $\{C_{low}, C_{med}, C_{high}\}$ were set to 0.001, 0.005 and 0.01 respectively, each test was replicated one hundred times, and 100 permutations for each test that requires permutations were performed.

$$\begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1.5 & 0.5 \\ 0.5 & 0.5 & 2 \end{pmatrix}$$
 respectively. Correlated $N(\mathbf{0}, \Sigma)$ errors with covariance matrix $\Sigma = \Sigma_K \otimes \Sigma_D$ with $\Sigma_D = \mathbf{I}$ were generated and added to each landmark. Power results for all five tests of interest are shown in Figures 3.18 and 3.19.

Figures 3.14 through 3.19 show that whenever unequal variances or correlations are present in the error structure, the ANOVA F-test and the ANOVA Permutation test lose some of their power with more loss due to unequal variances than to the presence of correlations. However, the F-test and the ANOVA Permutation test always have higher power than the distance-based and ANOVA form error tests under all covariance models tested. It is also noticeable that under the different covariance models these two tests appear to possess higher power for the 3D case than the 2D case. Since it was thought this might be due to the larger number of landmarks considered in the 3D cylinders relative to the 2D circles, we repeated the same simulation for the 2D case except that the number of landmarks was

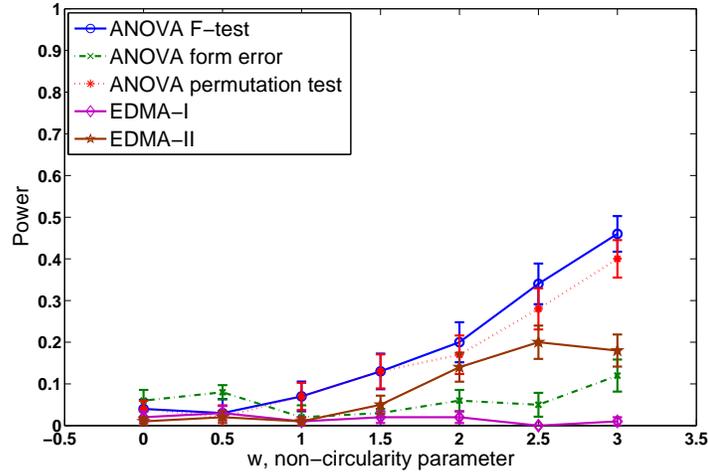


Figure 3.16. Statistical power for detecting changes in 2D circular shapes for the F-test, the Permutation ANOVA test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential isotropic errors as described by equation (3.13). Simulation parameters were: 64 landmarks ($= k$) per configuration, test size α of 0.05, 10 configurations ($= n$) per sample, the parameters $\{\tau, \sigma, \phi, r\}$ were set to 0.05, 0.05, 0.05 and 5 respectively, each test was replicated one hundred times, and 100 permutations for tests that require them were performed.

increased to 320. Results are shown in Figure 3.20. Comparing with the results in Figure 3.16 (where $k = 64$), it can be said that as the number of landmarks increases the power of the F-test and the ANOVA Permutation test increases indeed, although for model (3.13) the F-test provides a higher than advertised type-I error rate. This is due to the larger number of landmarks per configuration than in Figure 3.16 (320 compared to 64). Increasing the number of landmarks reduces the inter-landmark distances, and for correlation models (3.13) and (3.14) this makes the between landmarks correlation more evident. This in turn affects the ANOVA F-test since it assumes independence. On the other side, the ANOVA Permutation test assumes exchangeability between configuration matrices. But it does not assume independent landmarks, hence it is not affected by the between landmarks correlation (i.e., when k is large).

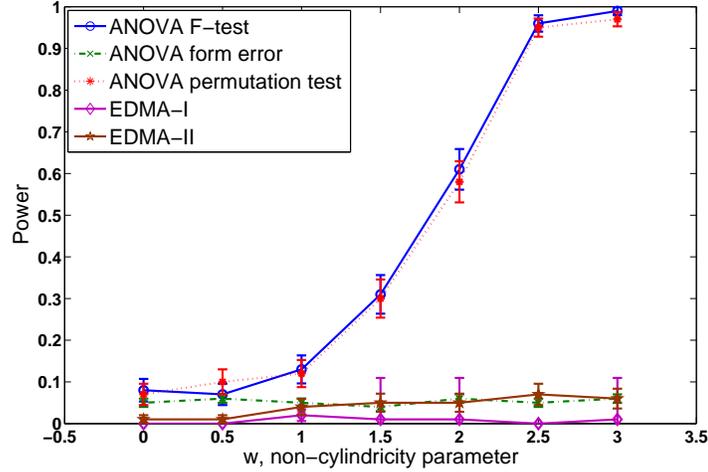


Figure 3.17. Statistical power for detecting changes in 3D cylindrical shapes for the F-test, the Permutation ANOVA test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential isotropic errors as described by equation (3.13). Simulation parameters were: 320 landmarks ($= k$) per configuration, test size α of 0.05, 10 configurations ($= n$) per sample, the parameters $\{\tau, \sigma, \phi, r, h\}$ were set to 0.05, 0.05, 0.05, 5 and 10 respectively, each test was replicated one hundred times, and 100 permutations for tests that require them were performed.

3.4 Performance of Tests Under Non-normal Errors

It has been shown how the two SSA ANOVA tests for shapes provide higher power than all other tests discussed in section 3. Together with the ANOVA test on the form errors, the ANOVA F-test is based on the assumption of normally distributed errors. Here we consider the robustness of all the ANOVA tests described in section 2.2 (ANOVA F-test, ANOVA Permutation test, and ANOVA on the form errors) under two cases of non-normal errors: a uniform distribution, and a t-distribution with 5 degrees of freedom. A simulation study similar to the one in section 5.1 was run except that the added errors were distributed as $\text{Uniform}(a, b)$ or $t_5(0, \sigma^2)$. The values of a and b were chosen such that the mean is zero and the variance is σ^2 . This yields $b = -a$, $a = -\sqrt{3}\sigma$ and $b = \sqrt{3}\sigma$. The fitted kernel density functions for the distributions of form errors under the ideal case of normal errors ($\text{vec}(\mathbf{E}) \sim N(\mathbf{0}, 0.05^2 \mathbf{I}_{km})$), the case of uniform errors

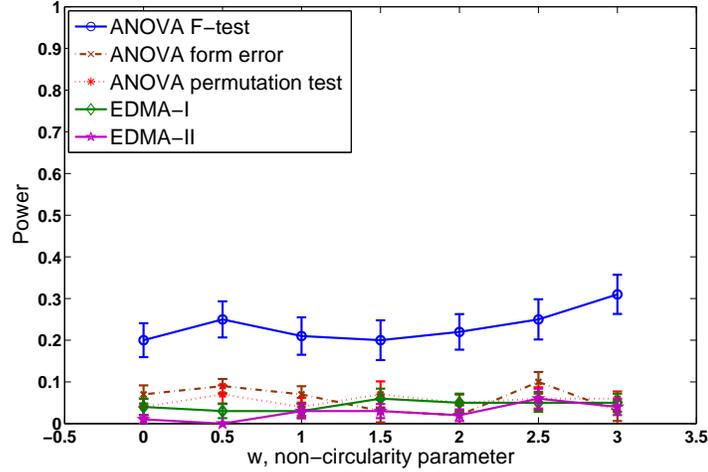


Figure 3.18. Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential non-isotropic errors as described by equation (3.14). Simulation parameters were: 64 landmarks ($= k$) per configuration, 10 configurations ($= n$) per sample, test size α of 0.05, the parameter $\{\tau, \sigma, \phi, r\}$ were set to 0.05, 0.05, 0.05 and 5 respectively, the matrix \mathbf{B} was set to $(1 \ 0.5; 0.5 \ 1.5)$, each test was replicated one hundred times, and 100 permutations were performed for each test that requires permutation.

($\text{vec}(\mathbf{E}) \sim \text{Uniform}((-0.05\sqrt{3})\mathbf{1}_{km}, (0.05\sqrt{3})\mathbf{1}_{km})$) and the case of t-distributed errors ($\text{vec}(\mathbf{E}) \sim t_5(\mathbf{0}, 0.05^2\mathbf{I}_{km})$) are shown in Figure 3.21. This figure shows how the variance of the distribution of the form errors increases as the tails in the distribution of the errors become thicker.

The estimated power curves for this case are shown in Figures 3.22 and 3.23. It is seen that the ANOVA F-test and the ANOVA Permutation test still show the best relative performance. However, the ANOVA F-test does not have the advertised type-I error rate. The ANOVA on the circularity form errors has a good power performance under uniform errors but shows a poor performance under t-distributed errors. This is not a surprising result since the form error is a range statistic and hence it will be more sensitive to a distribution with “outlier” errors than to a bounded distribution of the errors.

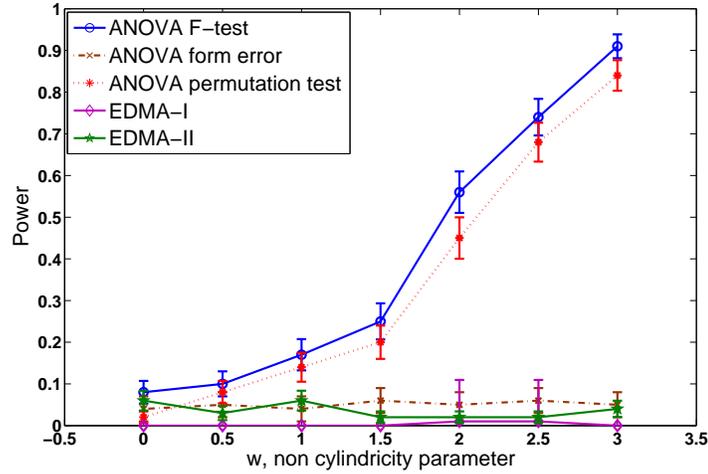


Figure 3.19. Statistical power for detecting changes in 3D cylindrical shapes for the F-test, the ANOVA Permutation test, the ANOVA form error method, and the EDMA-I and EDMA-II tests under exponential non-isotropic errors as described by equation (3.14). Simulation parameters were: 320 landmarks ($= k$) per configuration, 10 configurations ($= n$) per sample, test size α of 0.05, the parameter $\{\tau, \sigma, \phi, r, h\}$ were set to 0.05, 0.05, 0.05, 5 and 10 respectively, the matrix \mathbf{B} was set to $(1 \ 0.5 \ 0.5; 0.5 \ 1.5 \ 0.5; 0.5 \ 0.5 \ 2)$, each test was replicated one hundred times, and 100 permutations were performed for each test that requires permutation.

Table 3.5. Means and standard deviations of the simulated form error density functions used in Figure 3.21.

	$N(\mathbf{0}, 0.05^2 \mathbf{I}_{km})$	$\text{Uniform}((-0.05\sqrt{3})\mathbf{1}_{km}, (0.05\sqrt{3})\mathbf{1}_{km})$	$t_5(\mathbf{0}, 0.05^2 \mathbf{I}_{km})$
Mean	0.2343	0.1945	0.3528
Standard deviation	0.0316	0.0137	0.0939

3.5 Conclusions and Summary of Results of This Chapter

The performance analyses conducted in this chapter were classified according to the types of mean shapes utilized: first, the performance of tests for detecting differences in mean shape was studied for shapes with arbitrary geometry that are not commonly found in manufacturing, such as triangles and arbitrary polygons. These types of shapes have been considered by previous SSA authors, and hence

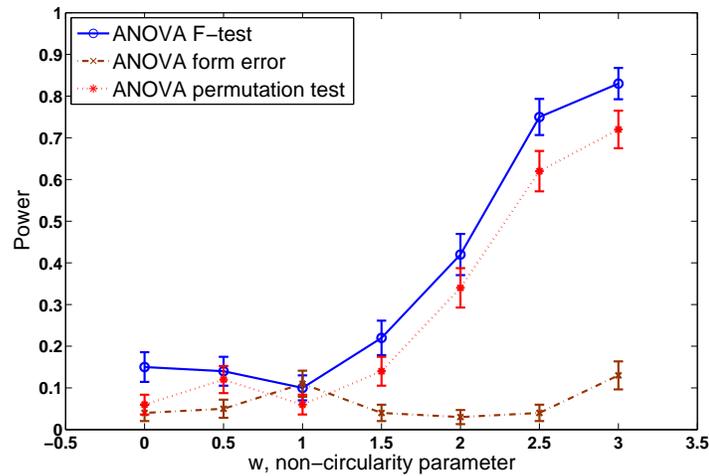


Figure 3.20. Effect of increasing the number of landmarks: Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test and the ANOVA on form errors under exponential isotropic errors as described by equation (3.13). Simulation parameters were: 320 landmarks ($= k$) per configuration, 10 configurations ($= n$) per sample, test size α of 0.05, the parameter $\{\tau, \sigma, \phi, r\}$ were set to 0.05, 0.05, 0.05 and 5 respectively, each test was replicated one hundred times, and 100 permutations were performed for each test that requires permutation (compare to Figure 3.16).

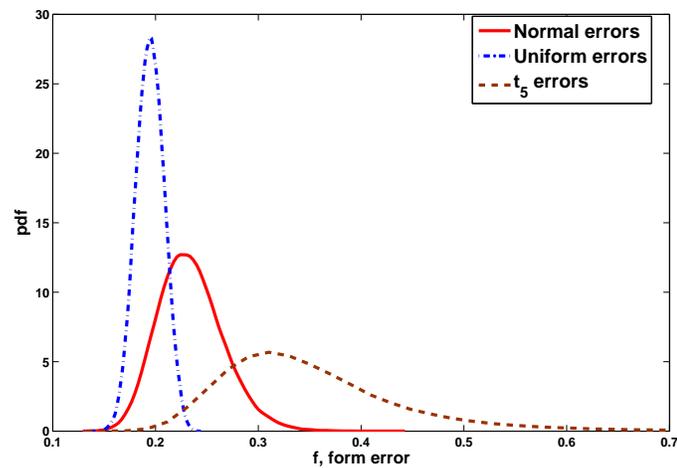


Figure 3.21. The fitted kernel density functions from 100,000 simulated circularity form errors ($k = 64$, isotropic errors).

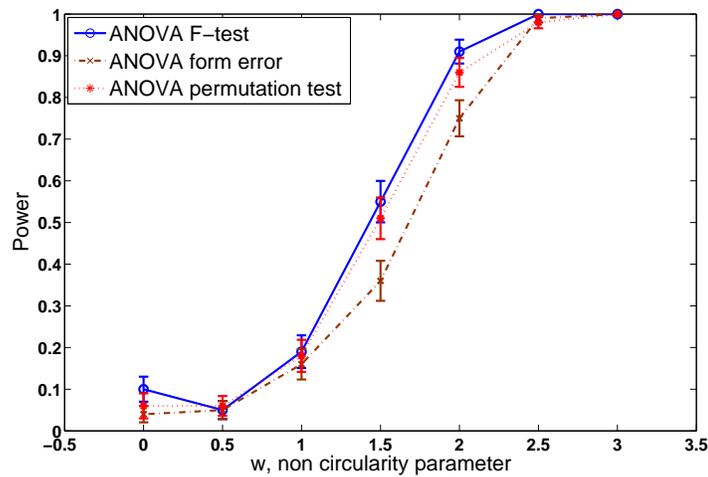


Figure 3.22. Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test and the ANOVA form error method under Uniform errors. Simulation parameters were: 320 landmarks ($= k$) per configuration, 20 configurations ($= n$) per sample, test size α of 0.05, the parameters $\{a, b\}$ were set to -0.0866 and 0.0866 respectively, each test was replicated 100 times, and 100 permutations were performed for each test that requires them.

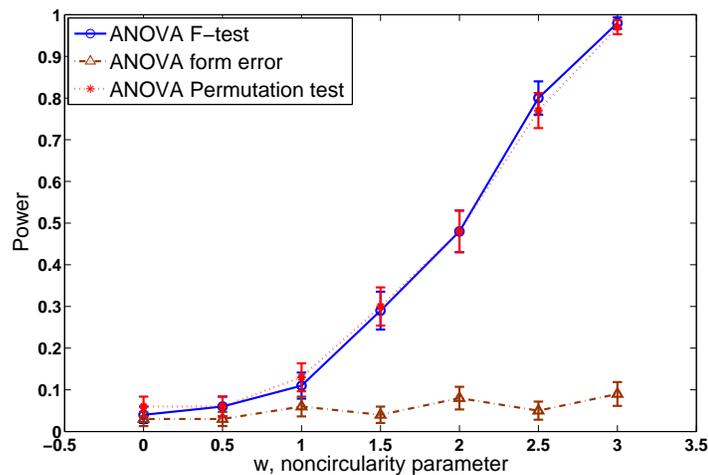


Figure 3.23. Statistical power for detecting changes in 2D circular shapes for the F-test, the ANOVA Permutation test and the ANOVA form error method under error terms simulated from $t_5(0, 0.05^2)$. Simulation parameters were: 320 landmarks ($= k$) per configuration, 20 configurations ($= n$) per sample, test size α of 0.05, each test was replicated 100 times, and 100 permutations were performed for each test that requires them.

the reason for their inclusion in the present study. Second, we studied the performance of these tests for shapes more commonly found in manufacturing, such as circles and cylinders. In the first case, the data sets all have a small number of landmarks per configuration, 5 at the maximum. The second type of shapes, those of interest in manufacturing, have not been studied from an SSA point of view before, and hence, simulated data sets were used, with considerably more landmarks per configuration.

The best-known test found in the SSA literature, the ANOVA F-test by Goodall (1991), extended to the two-way case by Del Castillo and Colosimo (2011), assumes the following:

1. Normal i.i.d. errors;
2. Constant error variances across all landmarks and dimensions;
3. Small error variance.

Fortunately, the third condition usually holds in advanced manufacturing processes, where SSA techniques are likely to be used. Related to this, we found that the rule of thumb in Dryden and Mardia (1998) used to determine when the variance is “small” (namely, $\sigma \leq 0.1$ is considered “small”) is appropriate. The robustness of the different SSA tests with respect to assumptions 1 and 2 above was studied in sections 3.2 and 3.3.

A summary of the results we found related to the statistical power and type-I error rate of each test is as follows:

- *Mean shapes of arbitrary geometry under isotropic errors*: the ANOVA Permutation test and the ANOVA F-test showed the highest power performance. The ANOVA F-test showed slightly lower than nominal type-I error rate.
- *Circles (2D) and cylinders (3D) under isotropic errors*: both the ANOVA F-test and the ANOVA Permutation test showed the best power performance with type-I error rates close to nominal.
- *Circles (2D) and cylinders (3D) under non-isotropic errors*: again, we found the ANOVA F-test and the ANOVA Permutation test to give best power performance among all considered tests with higher power for the ANOVA

F-test in some cases, but also with higher type-I error rate than assumed. We observed that the power of these two tests improves as the number of landmarks per configuration increases.

- *Circles under non-normal errors*: the ANOVA Permutation test showed the best power performance and type-I error rate control. The ANOVA F-test showed similar power to the ANOVA Permutation test but with a higher than nominal type-I error rate. In contrast, the ANOVA on the form errors provides good power performance only in cases where the distribution of the errors is bounded since it is based on a range statistic.

From these results, the following conclusions and recommendations can be made:

- Based on the simulation results it can be concluded that Goodall's ANOVA F-test and the ANOVA Permutation test are the most powerful tests for the types of shapes tested under isotropic and non-isotropic error structures. In particular, simulation evidence presented shows that the ANOVA Permutation test has better control of the type-I error rate, and hence it is the test we recommend.
- The EDMA-I and EDMA-II tests exhibited very low power in three dimensions under both isotropic and non-isotropic errors. We can not generalize our results to all three dimensional shapes since some other studies have shown that these tests do have good power in detecting shape differences for some geometrical shapes (however, this seems to be true only for shapes with very few landmarks). This raises the point that these tests depend on the shape being tested. This behavior is obvious for EDMA-II since its power depends also on the chosen baseline for scaling.
- Neither the ANOVA on the form errors nor the EDMA-I test is recommended since they exhibited very low power in most of the cases considered. Our results show that EDMA-I has the true designed type-I error under isotropic errors, but exhibits low power. Our results related to the ANOVA on the circularity and cylindricity form errors are relevant in manufacturing (tolerancing) practice. Practitioners should utilize the ANOVA Permutation test

to detect lack of circularity/cylindricity and for the estimation of factors effects on the resulting shape (for more on the later, see Del Castillo and Colosimo, 2011).

- A non-isotropic covariance worsens notably the performance of all tests considered. In particular, the power of the ANOVA Permutation test is reduced with respect to the isotropic case, but this test still performs relatively well if the number of landmarks per configuration is large. Fortunately, large numbers of landmarks per configuration are common in manufacturing data.

Areas of future research related to this chapter are given in chapter 6.

Chapter 4

Robust Parameter Optimization of Shape Responses

In this chapter, we extend the landmark-based shape optimization methods proposed by Del Castillo and Colosimo (2011) to allow for the inclusion of noise factors and for the case when objects of different sizes are considered different, i.e., when dilation effects should not be filtered out, and interest is in ‘shape-and-size’ or the *form* of the objects rather than their shape.

This chapter is organized as follows. In section 4.1, we introduce the problem of shape optimization, review the existing optimization method proposed by Del Castillo and Colosimo (2011) and extend it to the case of noise factors as well as the case when shape sizes differ. Section 4.2 provide two real manufacturing data examples where the extended method can be applied. Finally, we provide summary and conclusions of this chapter in section 4.3.

4.1 Shape Optimization

Let $\{A, B, \dots, L\}$ be controllable factors varied in a designed experiment. These factors might have an effect on the geometry of some manufactured part. We wish to find the settings of these L factors that provide a mean shape as close as possible to the target geometry of the part. At the same time we would like to minimize the variability of manufactured parts from the mean part geometry. One useful measure to determine optimal design factor settings is the Generalized Procrustes

distance G

$$G(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = \min_{\beta_i, \mathbf{\Gamma}_i, \gamma_i} \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \|\beta_i \mathbf{X}_i \mathbf{\Gamma}_i + 1_k \gamma_i' - (\beta_j \mathbf{X}_j \mathbf{\Gamma}_j + 1_k \gamma_j')\|^2 \quad (4.1)$$

found by the Generalized Procrustes Algorithm (GPA). We note here that the GPA provides reliable mean shape estimates when the errors variance σ^2 is small, attainable assumption in manufacturing applications. If σ^2 is large, intrinsic mean shape estimation techniques (Huckemann et al., 2010a,b) should be used instead. Let $\mathbf{X}_{(ab\dots l)}$ be $k \times m$ matrix of measurements representing an observed configuration at treatment combination $ab\dots l$ of design factors where a is the level of factor A , b is the level of factor B , etc. In their paper, Del Castillo and Colosimo (2011) solve the following problem to optimize manufactured parts geometry:

$$\min_{ab\dots l} \{G_{Mean} = G(\overline{\mathbf{X}}_{(ab\dots l)}, \mathbf{T})\} \quad (4.2)$$

where \mathbf{T} is the desired target shape and $\overline{\mathbf{X}}_{(ab\dots l)}$ is the mean shape at treatment $ab\dots l$. In other words, they choose treatment $ab\dots l$ that attains the statistic G_{Mean} a minimum as the optimal settings. This method considers the mean shape at each treatment and the target in the comparison. Hence, the variability of objects at each treatment combination is not considered.

The classical definition of shape, according to Kendall, filters dilations, rotations (excluding reflections) and location. If objects of different size are to be deemed different, then dilation should not be filtered out. Del Castillo and Colosimo (2011) provide an analysis of variance (ANOVA) test for differences in size, since it is common in tolerancing practice to separate size and shape.

To preserve size information and avoid a separate test for differences in size, we first assume that the measurements result from the mean form $\boldsymbol{\mu}$ by applying rotations and translations only, that is:

$$\mathbf{X} = (\boldsymbol{\mu} + \mathbf{E})\boldsymbol{\gamma} + 1_k \boldsymbol{\gamma}' \quad (4.3)$$

where $\boldsymbol{\gamma}$ is an $m \times m$ rotation matrix, $\boldsymbol{\gamma}'$ is a translation vector and \mathbf{E} is a $k \times m$ matrix of errors. This model is called ‘size-and-shape’ by which we will refer from

now on simply as a model for the 'form' of objects, rather than their shape (Dryden and Mardia, 1998). The form of an object is therefore defined as all the information remaining after removing rotation and location effects from an object.

Consider the following form Procrustes statistic

$$G_{size}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = \min_{\mathbf{\Gamma}_i, \gamma_i} \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \|\mathbf{X}_i \mathbf{\Gamma}_i + 1_k \gamma_i' - (\mathbf{X}_j \mathbf{\Gamma}_j + 1_k \gamma_j')\|^2 \quad (4.4)$$

This is similar to the G statistic in (4.1) but with scale minimization omitted. This optimization problem can be solved by following the same procedure provided in section 2.1.2 to solve for GPA in (4.1), but skipping steps 4 and 5 where size optimization is performed.

To find the best treatment combination $ab \dots l$ that minimizes variation around a target form in addition to the within-treatment variability, we solve the following optimization problem:

$$\min_{ab \dots l} \left\{ G_{Total} = G_{size}(\overline{\mathbf{X}}_{(ab \dots l)}, \mathbf{T}) + \frac{1}{n-1} G_{size}(\mathbf{X}_{(ab \dots l1)}, \dots, \mathbf{X}_{(ab \dots ln)}) \right\} \quad (4.5)$$

where $\mathbf{X}_{(ab \dots li)}$ represents the i^{th} object observed at treatment $ab \dots l$, $i = 1, \dots, n$ where n is the sample size. The first part of the objective function to be minimized in (4.5) accounts for the distance between the mean form and the target, whereas the second term accounts for form variability at that treatment. This is similar to minimizing the mean squared error (MSE) in classical statistics.

Following common convention in Robust Parameter Design (RPD) (Myers and Montgomery, 1995), *noise factors* are assumed to be modifiable during a carefully controlled experiment, but in actual use of the process or product they are not controllable and vary randomly. Similar to traditional RPD for scalar responses, the goal is to find the optimal settings of the controllable factors that make the process robust, or insensitive, to variations in the noise factors. Optimality in the case we are concerned with means achieving a specific target form with minimum variability.

To extend our method to the case of noise factors, we adapt the usual assumption in RPD literature that a significant interaction between control and noise factors is present. Hence, variability can be controlled by adjusting control factors only. Assume that additional R noise factors are considered in the experiment.

Define $\mathbf{X}_{(ab\dots l)(R)}$ to be a $k \times m$ matrix of measurements representing an observed configuration at combination $ab\dots l$ of the controllable factors and combination R of the noise factors, where we assume n configurations have been observed at each experimental run in the controllable factors. We solve :

$$\begin{aligned} \min_{ab\dots l} \{ & G_{Total} = G_{size}(\overline{\mathbf{X}}_{(ab\dots l)(\cdot)}, \mathbf{T}) \\ & + \frac{1}{n-1} G_{size}(\mathbf{X}_{(ab\dots l1)(\cdot)}, \dots, \mathbf{X}_{(ab\dots ln)(\cdot)}) \} \end{aligned} \quad (4.6)$$

where $\mathbf{X}_{(ab\dots li)(\cdot)}$ is the i^{th} observed configuration at combination $ab\dots l$ of control factors and any combination of noise factors, $i = 1, \dots, n$, and $\overline{\mathbf{X}}_{(ab\dots l)(\cdot)}$ is the average form of these n configurations observed at treatment $(ab\dots l)(\cdot)$ obtained by solving (4.1). We choose combination $ab\dots l$ which yields a minimum G_{Total} as the optimal factor settings. Consider the following example for illustration.

Example: Consider the 2-dimensional perfect circle of radius 5 mm and a notch of 1 mm depth shown in Figure 4.1 as a target part geometry. Suppose there are two design factors A and B each at two levels, low and high, that affect the geometry of the manufactured parts. Factor A introduces ellipticity along the y -axis (measured by the amplitude change δ from a perfect circle of radius 5 mm) as it changes from low to high while factor B controls the depth of the notch, d , in mm. Two noise factors, namely C and D , were found to affect the resulted part geometry. Factor C introduces ellipticity in an orthogonal direction as factor A does. Factor D affects the variability of the manufactured part (variability within each landmark). It also introduces ellipticity in the same direction as factor C does. Both of these factors are costly to control. We would like to reduce the effect of these two factors by controlling factors A and B . A 2^4 designed experiment was simulated with sample size $n = 20$ configurations at each treatment combination. Figure 4.2 shows all of the observed configurations while Figure 4.3 shows the mean part geometry at all treatments. It can be seen how the variability increases as factor D changes from high to low.

If noise factors were ignored, the experiment could be reduced to a 2^2 factorial design in factors A and B . The mean forms are estimated as shown in Figure 4.4; they all look close to the target. Applying the G_{Mean} method in (4.2) results in G_{Mean} statistic values shown in Table 4.1 with combination $A = 1$ and $B = 1$

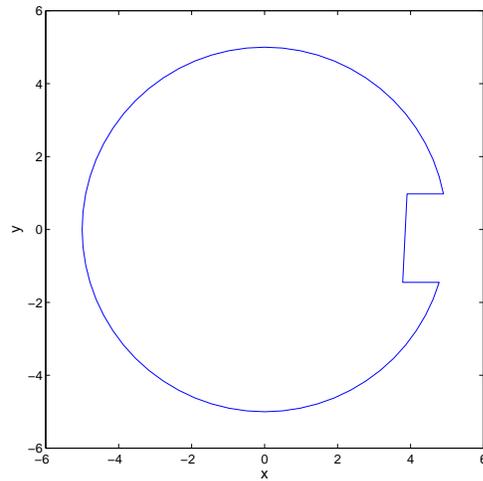


Figure 4.1. Target part geometry of example 2.

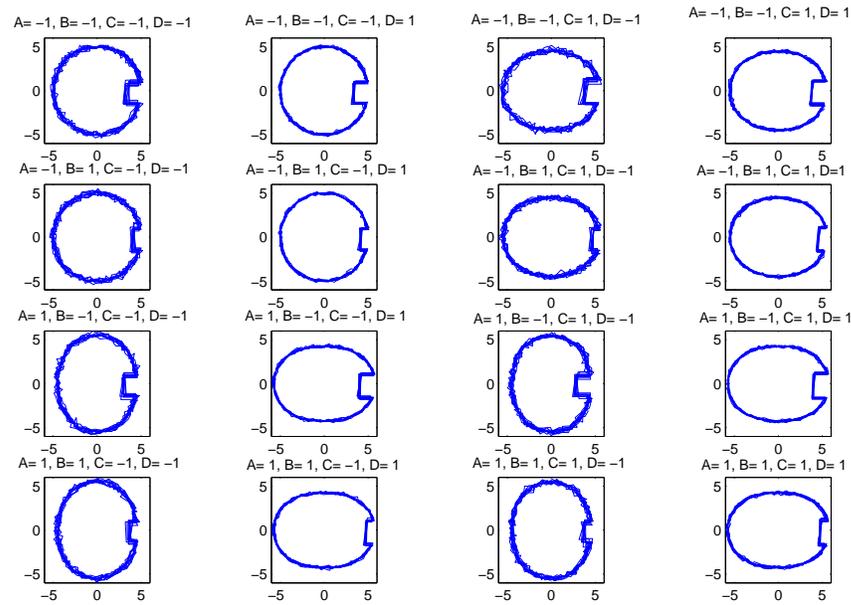


Figure 4.2. Observed configurations of example 2 (low=-1 and high=1).

chosen as the optimal factor settings. Figure 4.5 shows the simulated configurations at all combinations of control factors. As seen in Figure 4.5, treatment $A = 1$ and $B = 1$ generates parts with higher variability and might result in a high rejection

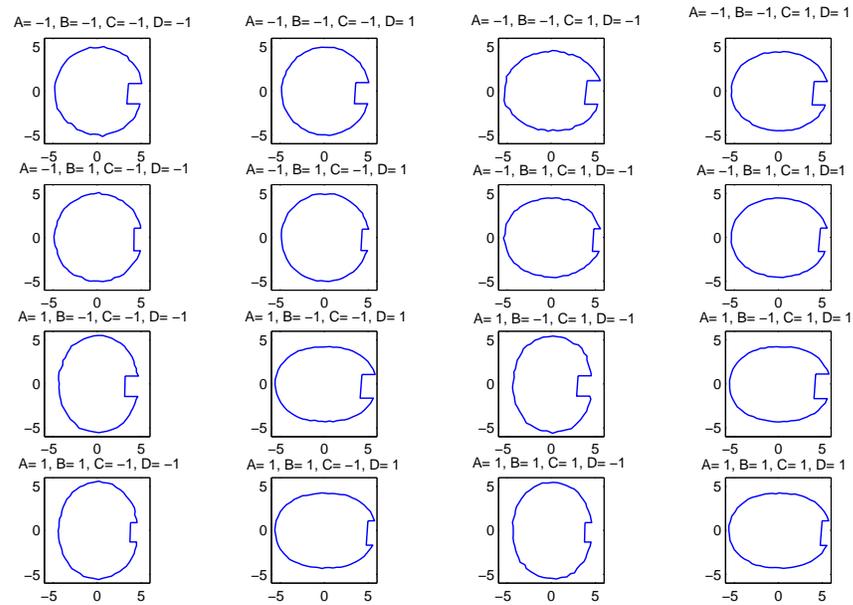


Figure 4.3. Mean part geometry at each treatment of control and noise factors of example 2 (low=-1 and high=1).

rate. The results of the G_{Total} method are shown in Table 4.2. Table 4.2 shows that $A = -1$ and $B = 1$ is a better combination since it provides the lowest G_{Total} value.

Table 4.1. The G_{Mean} values for data in example 2.

		Factor B	
		-1	1
Factor A	-1	2.33	1.87
	1	1.44	0.50

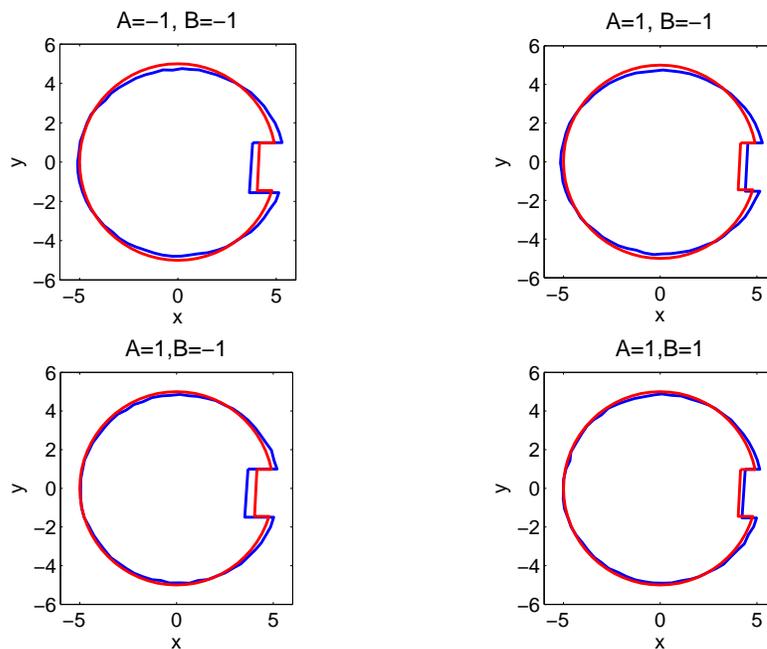


Figure 4.4. Mean part geometry (blue) overlaid by the target geometry (red) at all control factor combinations.

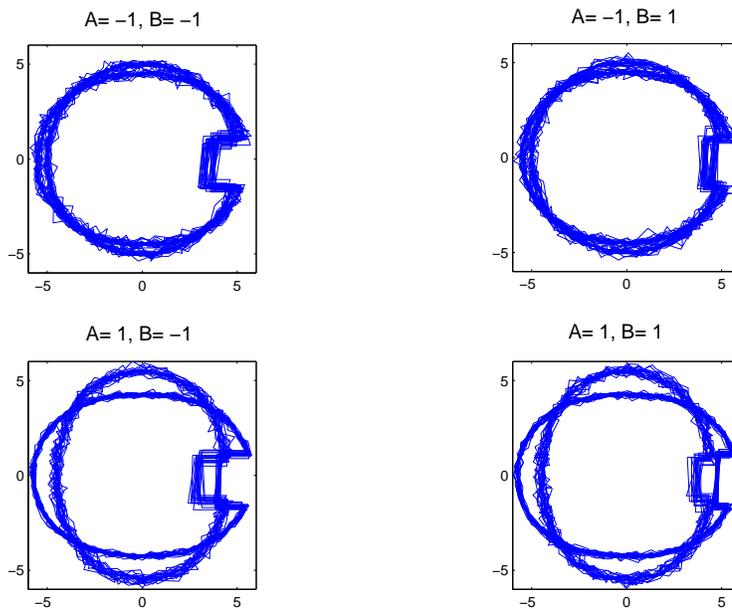


Figure 4.5. Observed configurations by ignoring noise factors ($low = -1, high = 1$).

Table 4.2. G_{Total} values obtained using (4.6) for example 2 data.

Factor A	Factor B	$G_{size}(\mathbf{X}_{(ab1)(\cdot)}, \dots, \mathbf{X}_{(ab80)(\cdot)})$	$G_{size}(\overline{\mathbf{X}}_{(ab)(\cdot)}, \mathbf{T})$	G_{Total}
-1	-1	204.1379	2.327984	4.912008
-1	1	195.6328	1.866385	4.34275
1	-1	595.6815	1.436474	8.976746
1	1	601.7431	0.495439	8.112439

4.2 Real Manufacturing Process Examples

4.2.1 Titanium Lathe-turning

Consider the Titanium lathe-turning example mentioned in section 2.2 and discussed by Del Castillo and Colosimo (2011). A set of 90 titanium alloy (Ti-6Al-4V) specimens was machined by lathe-turning. Two cutting steps were performed to reduce the initial diameter of 20 mm to the final diameter of 16.8 mm (as shown in Figure 4.6). The original specimens were obtained by vacuum arc remelting followed by forging, rolling, hardening (1 hour at 780 C and then air cooling) and a last phase of centerless grinding. Lathe-turning of the external surface was then performed considering a full factorial 3^2 design, where each of the 9 treatments was replicated 10 times. The two factors under study were A =depth of cut [mm] and B =cutting speed [mm/rev] of the final (finishing) machining step. Values assumed for these two parameters in each of the treatments are shown in Table 4.3. The machining feed [mm/rev] was selected at specific levels depending on the cutting speed (specifically a feed equal to 0.07, 0.11 and 0.14 mm/rev corresponded to a cutting speed of 80, 70 and 65 m/min, respectively). This type of dependency between the feed and the speed was suggested by the tool supplier in order to keep the tool life constant. For each specimen, the roundness profile was obtained at a fixed distance of 5 mm from the left-hand side of the specimen (shown with a dotted line in Figure 4.6). The profile was obtained using a CMM that measured a set of 64 equally spaced points on each profile. The goal of the experiment is to determine the best settings of the design factors to achieve the most circular parts.

Measured circular profiles are depicted in Figure 4.7. It can be seen from Figure 4.7 that all factor combinations provide almost the same mean form, but

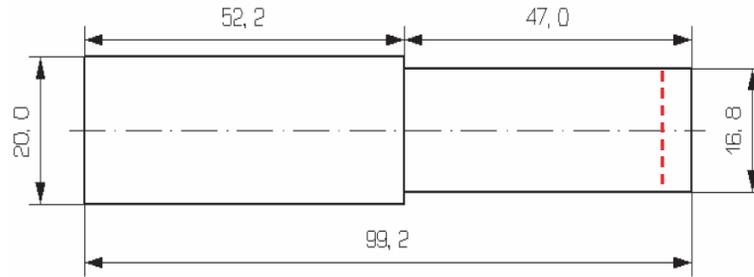


Figure 4.6. The desired geometry of the final Titanium specimen obtained by lathe-turning.

Table 4.3. Factors and levels in the Titanium machining experiment.

Treatment	A : depth of cut [mm]	B : cutting speed [m/min]
1	0.4	65
2	0.4	70
3	0.4	80
4	0.8	65
5	0.8	70
6	0.8	80
7	1.2	65
8	1.2	70
9	1.2	80

the variability decreases as both factor A and factor B increases. Applying the G_{Total} method, we get the values shown in Table 4.4. These results suggest that $A = 1.2$ and $B = 80$ is the optimal design setting at which a close mean form to the target is achieved with minimized variability around that mean form. This result is confirmed by engineering practices where it is usually suggested to cut at high speed and low metal removal rate in order to gain better cut quality and size control.

For illustration purpose, let us assume that factor B is a noise factor that is hard or expensive to control. The observed data at all levels of factor A are shown in Figure 4.8. As it is seen in Figure 4.8, we get the minimum variability at $A = 1.2$. This is confirmed by the calculated G_{Total} values shown in Table 4.5.

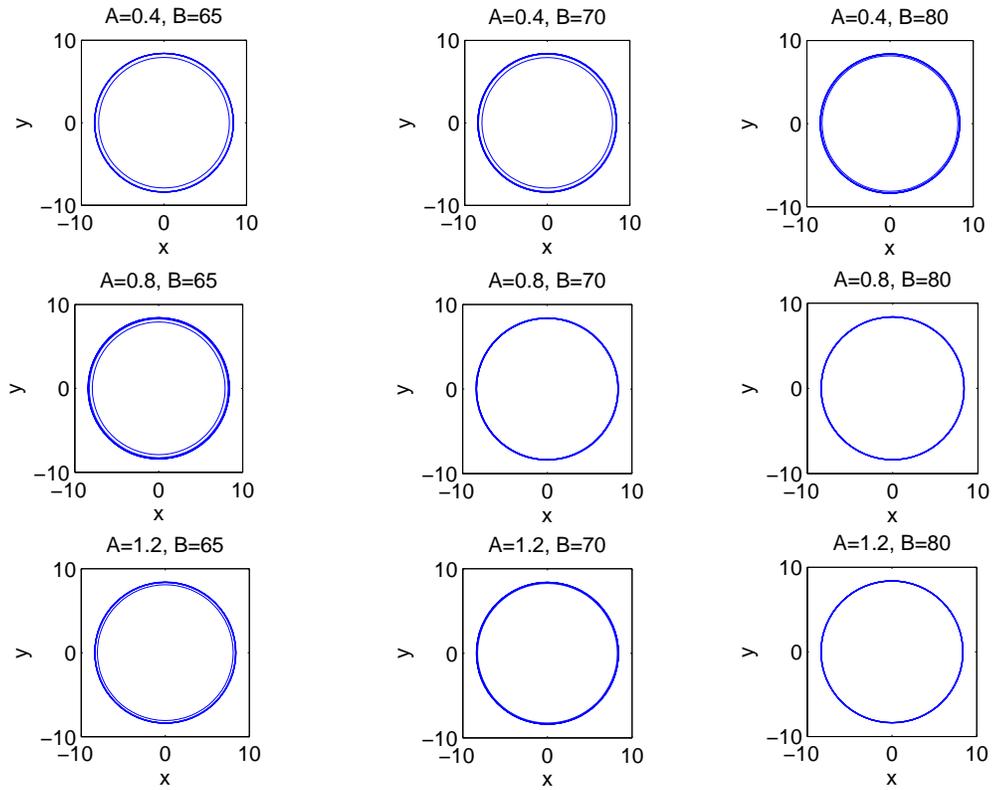


Figure 4.7. Observed shape data of the Titanium lathe-turning experiment.

Table 4.4. Calculated G_{Total} values for the Titanium lathe-turning data.

Treatment	A : depth of cut [mm]	B : cutting speed [m/min]	G_{Total}
1	0.4	65	1.5429
2	0.4	70	1.5401
3	0.4	80	0.32001
4	0.8	65	1.6456
5	0.8	70	0.045513
6	0.8	80	0.020104
7	1.2	65	0.59855
8	1.2	70	0.11562
9	1.2	80	0.016857

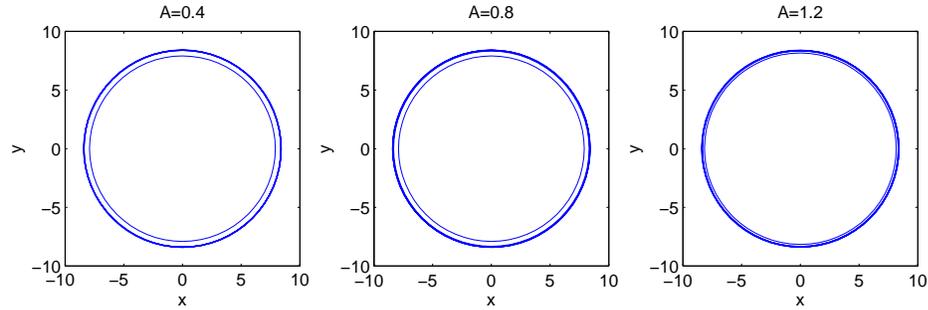


Figure 4.8. Observed shape data of the Titanium lathe-turning experiment assuming factor B is noise factor.

Table 4.5. Calculated G_{Total} values for the Titanium lathe-turning data assuming factor B is noise factor.

A : depth of cut [mm]	G_{Total}
0.4	3.7912
0.8	1.7519
1.2	0.39922

4.2.2 Aluminum Cylinders Lathe-turning

A set of 40 Aluminum alloy (6061) cylindrical specimens was machined by lathe-turning in the FAME LAB at Penn State by the author. Each specimen was 4" in length and was machined to reduce the diameter from 1" to 0.8" for the first 2.5" of its length. Three factors were considered in the experiment, A =spindle rotational speed [RPM], B =depth of cut [in] and C =machine type. Both factors A and B were set at two levels each. Two different types of machines were utilized: a 50-years-old LeBlond Regal GSA E800-25 lathe-turning machine and a HAAS SL30 tail-stock computerized CNC machine. The machine type factor was considered as a noise factor. This situation could happen in manufacturing where large quantities of a certain part are needed but not enough computerized CNC machines are available. Thus, old manual cutting machines can be used to prevent time delays. Values assumed for the three factors in each of the eight treatments

are shown in Table 4.6. The machining feed [in/rev] was set at 0.002 based on the recommendations of the tool manufacturer to maximize tool life. The experiment was replicated five times at each of the eight treatments. For each machined part, a set of 179 landmarks was measured along the machined cylindrical surface using a Ziess Vista C400 CMM and Calypso version 3.5 measurement software. All of the observed 40 configurations are depicted in Figure 4.9. The goal of this experiment is to find the optimal settings of design factors that generate the most cylindrical shapes of diameter 0.8” with minimum shape variability.

Table 4.6. Factors and levels in the Aluminum machining experiment.

Treatment	$A =$ cutting speed [RPM]	$B =$ depth of cut [in]	$C =$ machine
1	179	0.02	old
2	179	0.02	CNC
3	179	0.05	old
4	179	0.05	CNC
5	1800	0.02	old
6	1800	0.02	CNC
7	1800	0.05	old
8	1800	0.05	CNC

For form optimization, we first calculate the G_{Total} values using equation (4.6). These values are shown in Table 4.7 and suggest choosing $A = 179$ and $B = 0.02$ as the optimal design setting. As seen in Table 4.7 and Figure 4.10, this combination provides the lowest variability. Lower cutting speed and lower depth of cut makes the cutting operation more controllable and hence, less variability is generated.

4.3 Summary and Conclusions to This Chapter

In RPD problems, experimental factors are divided in two types, controllable and noise factors, and the goal is to find the values of the controllable factors that yield response values that are robust to variation in noise factors. In shape optimization, the response to be optimized is the whole part geometry.

SSA has been recently introduced to manufacturing by Del Castillo and Colosimo (2011). They have provided a statistical tool for shape optimization

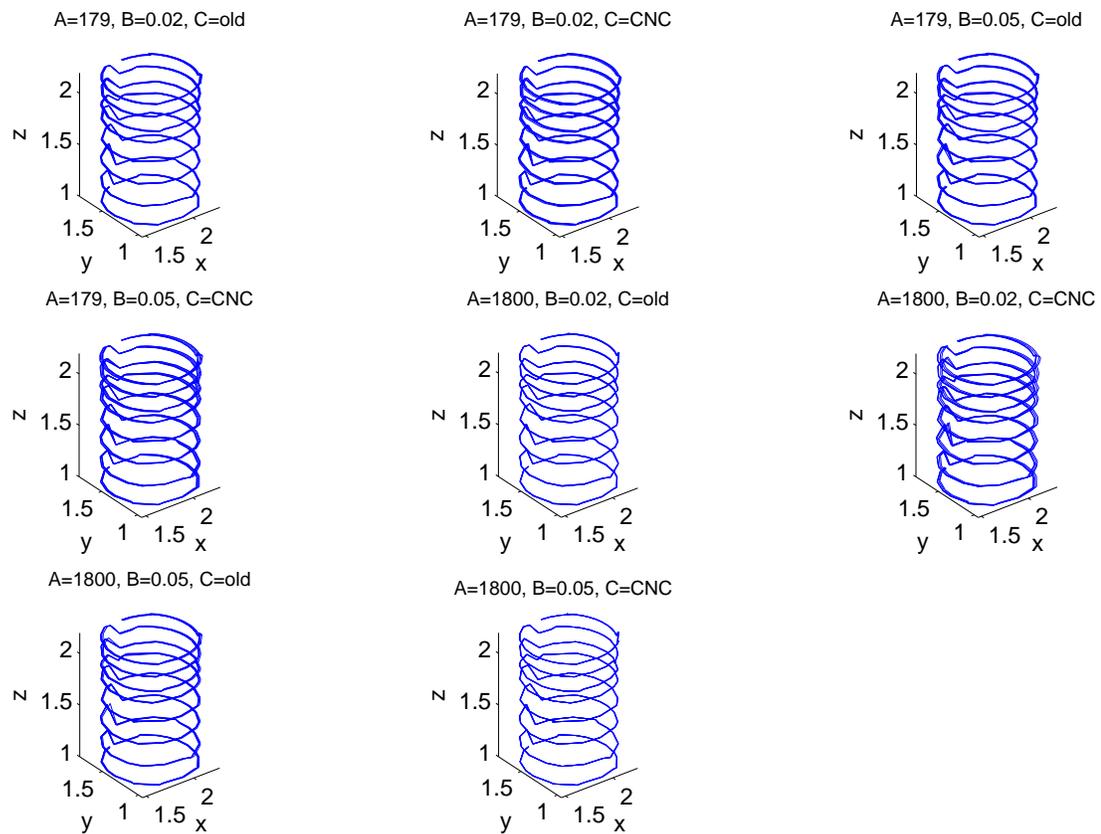


Figure 4.9. Observed shape data of the Aluminum lathe-turning experiment.

Table 4.7. Calculated G_{Total} values for the Aluminum lathe-turning data with control factors only.

Treatment	$A =$ cutting speed [RPM]	$B =$ depth of cut [in]	G_{Total}
1	179	0.02	0.011028
2	179	0.05	0.018547
3	1800	0.02	0.037036
4	1800	0.05	0.018381

where they compare the mean shape with the target and choose the treatment combination with the minimum difference as the optimal one. In their optimiza-

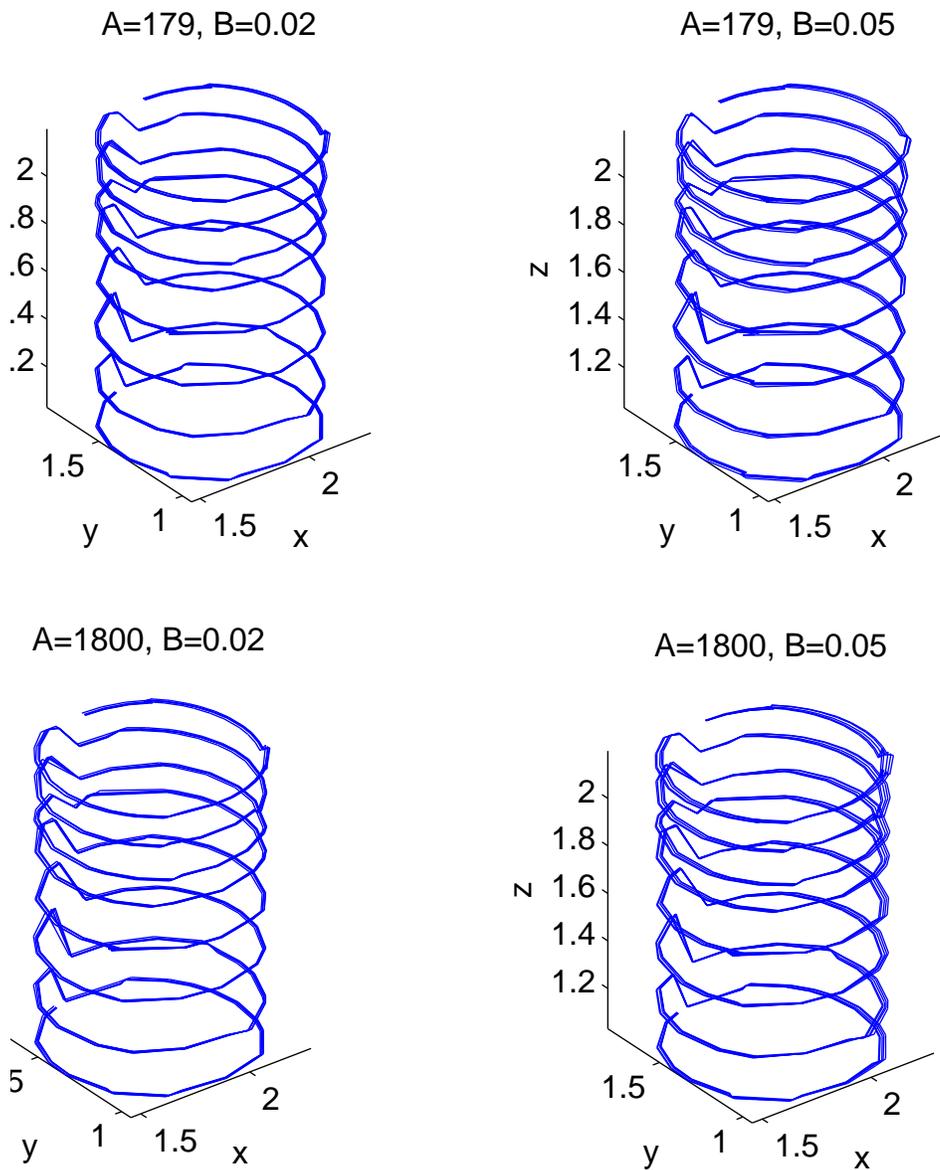


Figure 4.10. Observed shape data of the Aluminum lathe-turning experiment considering control factors only.

tion, they adopt the classical deflection of shape where size information are filtered out, and do provide a separate test for the size effect. Also, they consider the case where all design factors are controllable.

We propose a statistical tool for form optimization that considers both controllable and noise factors. The proposed approach uses the object forms for optimization, and hence size information is preserved and no separate test for differences in size is required. The proposed approach is illustrated with several simulated and real manufacturing shape data. The results obtained show that our method provides good solutions in terms of close mean form to the target and minimum variability.

Spatio-Temporal Modeling and Optimization of Profile Response Experiments

This chapter discusses the third and last problem studied in this dissertation, the robust parameter optimization of profile response experiments, considered in this dissertation. In this chapter, a Gaussian Random Function (GRF) process model is adopted to study and optimize profile responses. The chapter is organized as follows. The proposed model is explained in section 5.1. Section 5.2 discusses model robustness and validation. Finally, conclusion remarks are given in section 5.3. Some details about convergence assumptions and about the particular Markov Chain Monte Carlo techniques utilized are shown in the Appendices.

5.1 A Spatio-Temporal Gaussian Random Function Process Model

Consider an experiment where the response of interest $\{y_j\}_{j=1}^J$ is observed at several locations $\mathbf{s} = \{s_1, s_2, \dots, s_J\}$, after each experimental run. The locations s_j can refer to instances in time when the profile response is observed, or in general, they can refer to some other variable the observed profiles depend on, so we have, in effect, a functional response (Ramsay and Silverman, 2005). Assume the process

performance depends on the shape of the sampled profile, where a given target profile shape is desired. The shape of the sampled profile is modifiable through a set of control factors \mathbf{x}_c and is also affected by noise factors \mathbf{x}_n . Similarly to classical Robust Parameter Design (RPD) experiments, noise factors are assumed to be controllable in a carefully designed experiment but are uncontrollable once the product or process under study is optimized. The goal is to find the best settings of \mathbf{x}_c that make the process achieve a desired target profile shape with maximum probability.

Existing profile modeling techniques (Del Castillo et al., 2011; Nair et al., 2002) assume that a profile response can be modeled, at a first stage, as a regression in the sampling locations. Profile responses can have arbitrary functional forms and it may be hard sometimes to model the shape of the mean profile with a linear regression in the s_j 's. A more flexible alternative is to model the within- and between-profiles correlation using a GRF process and use these correlations to improve the predictions. We will assume that the process that generates the observed profiles follows an infinite dimensional normal distribution. Due to the properties of the normal distribution, any finite set of observations will also be normally distributed (Banerjee et al., 2004, section 2.4).

Assume that all N profiles each measured (sampled) at the same J locations. Let y_{ij} be the observed response value of profile i at location j . Also, let $\mathbf{x} = (\mathbf{x}_c, \mathbf{x}_n)$ and define $\mathbf{f}(\mathbf{x}_i, s_j)$ to be a function of the design factor settings \mathbf{x} under which the j^{th} location in the i^{th} profile was observed. Then the observed response values can be modeled as

$$\begin{aligned} y_{ij} &= \mu(\mathbf{x}_i, s_j) + \epsilon_{ij}, & \epsilon_{ij} &\sim N(0, \sigma^2(\mathbf{x}_i, s_j)) \\ &= \mathbf{f}(\mathbf{x}_i, s_j)' \boldsymbol{\beta} + \epsilon_{ij} \end{aligned} \tag{5.1}$$

Notice that both the mean and the variance are functions of the design factor settings and measurement location. Let $\boldsymbol{\epsilon}_i$ be the vector of errors ϵ_{ij} observed along the i^{th} profile, that is:

$$\boldsymbol{\epsilon}_i = [\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{iJ}]'$$

and let

$$\mathbf{F}_i(\mathbf{x}_i, \mathbf{s}) = \begin{bmatrix} \mathbf{f}(\mathbf{x}_i, s_1)' \\ \mathbf{f}(\mathbf{x}_i, s_2)' \\ \vdots \\ \mathbf{f}(\mathbf{x}_i, s_J)' \end{bmatrix}.$$

Then the i^{th} profile, \mathbf{y}_i , can be expressed as

$$\mathbf{y}_i(\mathbf{x}_i) = \mathbf{F}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_J(\mathbf{0}, \boldsymbol{\Sigma}_s(\mathbf{x}_i, \mathbf{s})) \quad (5.2)$$

where $\boldsymbol{\Sigma}_s$ is the within-profile covariance matrix. If we define the matrices

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1J} \\ y_{21} & y_{22} & \cdots & y_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NJ} \end{pmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \vdots \\ \mathbf{F}_N \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix}, \quad (5.3)$$

then \mathbf{Y} , the matrix of all N profiles can be represented as:

$$\begin{aligned} \text{vec}(\mathbf{Y}') &= \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_{NJ}(\mathbf{0}, \boldsymbol{\Sigma}(\mathbf{x}, \mathbf{s})) \quad \text{or} \\ \text{vec}(\mathbf{Y}') &\sim N_{NJ}(\mathbf{F}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \end{aligned} \quad (5.4)$$

where \mathbf{F} of size $NJ \times q$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of the regression parameters, and $\text{vec}(\cdot)$ is the operator that concatenates matrix columns into one vector.

The covariance structure, $\boldsymbol{\Sigma}$, must capture both the within- and the between-profile correlations. In order to reduce the number of covariance parameters to be estimated one could assume a spatial covariance model over both \mathbf{x} and \mathbf{s} . However, the design factor space (x -space) and the measurement location space (s -space) are usually measured in different scales (Wikle and Berliner, 2005), and a single spatial covariance model would be inadequate. Therefore, we assume instead two separate spatial covariance models, one for the x -space and another for the s -space, in such a way that:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s \quad (5.5)$$

where the $N \times N$ matrix Σ_x models the between-profiles correlations due to the change in design factors, the $J \times J$ matrix Σ_s models the within-profile correlations due to the proximity of any two locations in \mathbf{s} , and \otimes denotes the Kronecker product. This assumption is referred to as *separability* in Spatial Statistics (Genton, 2007). The rationale is that if operating conditions \mathbf{x}_1 and \mathbf{x}_2 are close in x -space, then they will tend to result in similar profile responses. Likewise, if responses y_{ij} and y_{ik} are such that j and k are close in s -space, then they will tend to be similar. For example, Figure 5.1 below shows a schematic representation of a profile response experiment in a two dimensional x -space. The two points \mathbf{x}_1 and \mathbf{x}_2 , that are close to each other in the x -space, are expected to generate two profiles \mathbf{y}_1 and \mathbf{y}_2 in the s -space that are correlated. In contrast, the point \mathbf{x}_3 that is farther away (in the x -space) from \mathbf{x}_1 and \mathbf{x}_2 should generate a profile response \mathbf{y}_3 that is not that correlated with \mathbf{y}_1 and \mathbf{y}_2 . The Kronecker product decomposition of Σ makes the model more attractive computationally for large size problems (Genton, 2007), since one deals with separate $N \times N$ and $J \times J$ covariance matrices Σ_x and Σ_s instead of a single $NJ \times NJ$ covariance matrix, Σ . Another reason behind the wide use of separable covariance structures is that they provide an easy way for generating positive definite covariance matrices (Genton, 2007; Gneiting, 2002). We note that this Kronecker product decomposition is not unique since

$$\Sigma_x \otimes \Sigma_s = (c\Sigma_x) \otimes \left(\frac{1}{c}\Sigma_s\right). \quad (5.6)$$

which results in a non-identifiability problem. To reduce the number of parameters in the separable covariance structure and to eliminate the non-identifiability problem, we assume the following exponential covariance functions:

$$\Sigma_s = \exp\{-\mathbf{D}_s/\phi_s\} \quad (5.7)$$

$$\Sigma_x = \kappa \exp\{-\mathbf{D}_x/\phi_x\} + \psi_x \mathbf{I}. \quad (5.8)$$

where \mathbf{D}_s is a $J \times J$ matrix of Euclidean distances between measuring locations and \mathbf{D}_x is an $N \times N$ matrix of distances between the design factor settings. Notice that we have only four parameters to determine, $(\phi_s, \kappa, \phi_x, \psi_x)$. The covariance functions (5.7-5.8) set the diagonal elements of Σ_s to ones, restricting the constant

c to be one in (5.6). Additional notes on this Kronecker product decomposition of covariance matrices are provided in Appendix C.

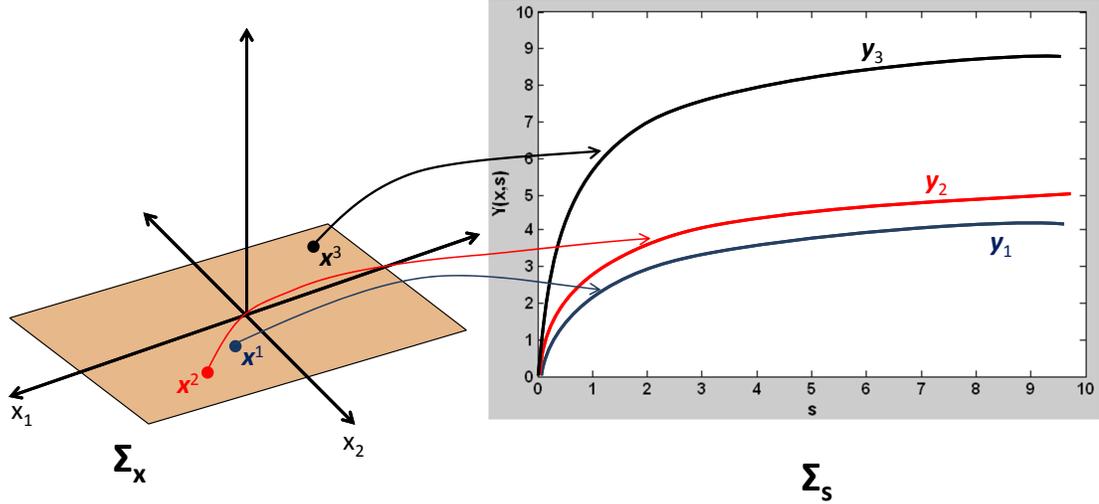


Figure 5.1. A schematic representation of the spatio-temporal data structure. Points x_1 and x_2 are close in the x -space and hence they are expected to generate similar profiles, y_1 and y_2 . In contrast, a point x_3 in x -space that is farther away from x_1 and x_2 is expected to generate a profile response y_3 that is not that correlated with y_1 and y_2 .

Letting $\boldsymbol{\theta} = (\phi_s, \psi_x, \phi_x, \kappa)$ and assuming the priors $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\beta})$ are available, the joint posterior density is:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta} \mid \mathbf{Y}, \mathbf{F}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\beta})|\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s)^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\} \quad (5.9)$$

For this model, full conditionals of the parameters are usually hard or even impossible to derive in a closed form (see Banerjee et al., 2004, section 5.1.1), and hence Gibbs sampling is not possible in general. Therefore, a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm is required to draw samples from the posterior distribution in (5.9). Full Bayesian estimation of a GRF model is complicated due to convergence problems (Besag and Green, 1993). We have achieved good convergence behavior with the parametrization (5.7-5.8) and the

adaptive Metropolis method of Appendix D.

If \mathbf{z} is a new observed profile at factor settings \mathbf{x}^* then, due to the GRF process assumption we have:

$$\left(\begin{bmatrix} \text{vec}(\mathbf{Y}') \\ \mathbf{z}' \end{bmatrix} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{F}, \mathbf{x}^* \right) \sim N_{(N+1)J} \left(\begin{bmatrix} \mathbf{F}\boldsymbol{\beta} \\ \mathbf{F}_z(\mathbf{x}^*, \mathbf{s})\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

where $\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s$, $\boldsymbol{\Sigma}_{22} = (\psi_x + \kappa)\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_{12}$ is an $NJ \times J$ matrix such that the i^{th} $J \times J$ block is $[\psi_x + \kappa \exp(-d(\mathbf{x}_i, \mathbf{x}^*)/\phi_x)]\boldsymbol{\Sigma}_s$ for $i = 1, \dots, N$. The distribution of $\mathbf{z} \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\beta}$ is easily derived using basic results from Multivariate Normal Theory (see, e.g., Johnson and Wichern, 2009) and is equal to:

$$\mathbf{z}' \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\beta} \sim N \left(\mathbf{F}_z(\mathbf{x}^*, \mathbf{s})\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta}), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \right) \quad (5.10)$$

Using the composition rule (see, e.g., Gelman et al., 2004) we can integrate numerically the posterior predictive density $\pi(\mathbf{z} \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}^*)$:

$$\pi(\mathbf{z} \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}^*) \propto \int \int \pi(\mathbf{z} \mid \mathbf{x}^*, \boldsymbol{\beta}, \boldsymbol{\theta}) \pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{F}) d\boldsymbol{\beta} d\boldsymbol{\theta} \quad (5.11)$$

The integration in (5.11) needs to be carried out by first generating samples for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ from their joint posterior density (5.9). Appendix D provides an adaptive MCMC algorithm to generate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ samples from the posterior distribution in (5.9). In all examples shown in this chapter the prior distributions shown in equations (D.1-D.5) in Appendix D were used for the model parameters. The generated $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ samples are then used to generate as many samples as needed from the density $\pi(\mathbf{z} \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}^*, \boldsymbol{\theta}, \boldsymbol{\beta})$.

To find the optimal control factor settings that makes the process robust to variability in the noise factors, we maximize the “probability of conformance” to the given specification limits, $p(\mathbf{x}_c)_{RPD}$, with respect to \mathbf{x}_c where

$$\begin{aligned} p(\mathbf{x}_c)_{RPD} &= E_{\mathbf{x}_n} [P(\mathbf{z} \in \mathbf{T} \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}_c, \mathbf{x}_n)] \\ &= \int P(\mathbf{z} \in \mathbf{T} \mid \mathbf{Y}, \mathbf{F}, \mathbf{x}_c, \mathbf{x}_n) \pi(\mathbf{x}_n) d\mathbf{x}_n \end{aligned} \quad (5.12)$$

with \mathbf{T} being a given set of specifications (e.g., low and high values of y at each location s_j) for the desired values for the response, and $\pi(\mathbf{x}_n)$ is the probability density of the noise factors which we assume known following standard RPD assumptions. This is the Bayesian predictive optimization approach of Peterson (2004) and Miro et al. (2004) applied to functional or profile responses, where $p(\mathbf{x}_c)_{RPD}$ is estimated by Monte Carlo integration. The Bayesian predictive approach has the advantage of considering the correlation structure of the data, the variability of the noise factors, and the uncertainty in the model parameters.

The GRF model in (5.4) assumes a zero mean and a stationary error ϵ . If the observed data has a trend in the x -space, the s -space or both, then the data need to be detrended in order to obtain stationarity. Furthermore, noise \times control interactions for the mean may be needed in a Robust Parameter Design problem (Myers and Montgomery, 1995). Therefore, we suggest to use an additive mean structure in both \mathbf{x} and \mathbf{s} to gain such a stationary error structure. Furthermore, if a significant interaction between the x -space and the s -space is anticipated, this interaction should also be included in the mean structure to enhance the Kronecker product approximation to Σ . Finally, when noise factors are present, it is recommended to include all two factor control \times noise interaction terms in the mean structure to allow for RPD optimization.

Example 1, no noise factors: metal injection moulding process. Gov-aerts and Noel (2005) report an experiment where 25 profiles of the elastic modulus (\mathbf{Y}) of green parts (products before a sintering operation) were observed in a metal injection moulding process. The elastic modulus was measured for each of the 25 parts at 701 values (locations) of the debinding temperature ranging from 10 to 80° C. The experiment consisted of two controllable factors in the ingredients of the binder, namely, Xanthan concentration (denoted by x_1 and varied at 5 levels from 1 to 5) and Chromium/Xanthan concentration ratio (denoted by x_2 and varied at 4 levels from 1:1 to 4:1). To speed up computations, the number of locations was reduced to 78 locations by sampling every 9th observed value. It was reported by the authors that one of the profiles was a clear outlier, so it was excluded from the analysis. The objective of the experiment as discussed by the authors is to obtain a large elastic modulus at low temperature values while using low Chromium concentration, given it is a pollutant. The specification limits were therefore set

considering the range of the observed profiles and the conditions above. Figure 5.2 shows the observed 24 profiles (after removing the outlier profile) along the 14 distinct design factor settings.

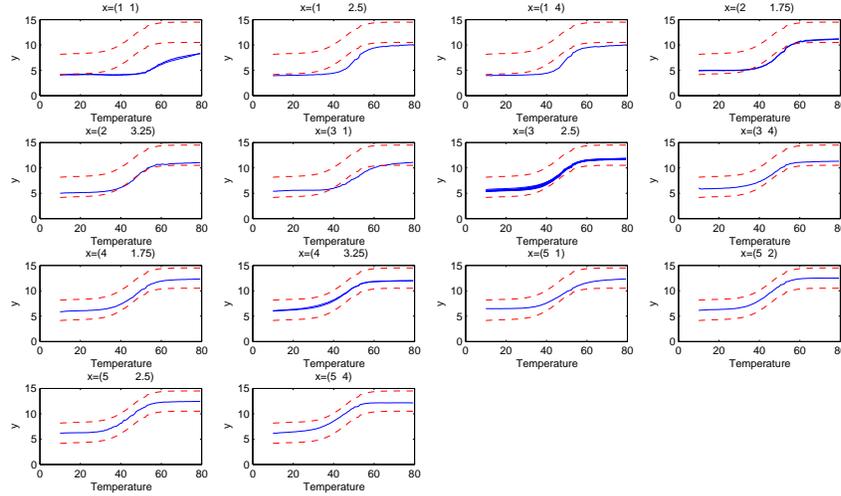


Figure 5.2. Observed elastic modulus profiles at each of the 14 control factor treatments in the metal injection moulding process. The lower and upper specification limits are shown by dashed lines.

We assume that the mean structure has a first order form in design factors and locations, i.e., $\mu(\mathbf{x}_i, s_j) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 s_j$, $i = 1, \dots, 24$ and $j = 1, \dots, 78$. 1000,000 samples from the posterior distribution were generated and thinned by keeping every 10^{th} sample resulting in 100,000 samples. The thinned samples were checked for convergence through five types of plots for each of the four covariance parameters, a trace plot, an autocorrelation function plot (ACF), a plot for the expected value vs. the sample number, a plot of the MCMC standard error of the posterior mean estimate calculated using the batch means method (see e.g., Flegal et al., 2008), and a plot for the posterior variance vs. the sample number. As it can be seen from Figure 5.3, trace plots show steady behavior and ACF plots all decay. In addition, the expected value and variance plots also converge to an asymptote, and the MCMC standard error approaches zero as the sample number increases. The generated MCMC samples were used in MATLAB's `ga` genetic optimization algorithm to maximize $p(\mathbf{x}_c)_{RPD}$. The optimization was constrained

to $1 \leq x_1 \leq 5$ and $1 \leq x_2 \leq 4$. The optimal solution found was at $x_1^* = 4.995$ and $x_2^* = 2.006$ with probability of conformance of 0.574 ($SE = 0.0157$) and can be verified through the surface plot of $p(\mathbf{x}_c)_{RPD}$ shown in Figure 5.4(a). In more general multi-factor experiments, either using an algorithm which attempts to find a global optimum (like `ga` does) or using a local non-linear optimizer started from a large number of initial points is necessary given $p(\mathbf{x}_c)_{RPD}$ is not concave.

Figure 5.4(b) shows the mean, 10th and 90th percentiles of the posterior predictive density at the optimal solution found. Luckily, this optimal solution satisfies the criteria for a low Chromium requirement otherwise the upper bound for x_2 would need to be lowered. The optimal solution found coincides with that found by Del Castillo et al. (2011), but they report higher conformance probability since they used wider specification limits. It is important to mention that the fitted model needs to be validated before it is used for process optimization, otherwise optimization results may not be accurate. Model validation is discussed in section 5.2. We first present a second example where noise factors are considered in the experiment.

Example 2, with noise factors: electric alternator design. In the electric alternator design example by Nair et al. (2002), presented in the introduction, the goal is to find the controllable factor settings that maximize the probability the electric current profile lies within the specified limits U and L given in Table 5.1.

Table 5.1. Specification limits at each of the 7 RPM locations for the electric alternator example.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7
$U(s_j)$	190	210	215	220	225	230	230
$L(s_j)$	120	140	155	170	185	200	200

To fit model (5.4-5.8) we assume a mean model with an intercept, all main effects and all two factor control×noise interaction terms, i.e., $\mu(\mathbf{x}_i, s_j) = \beta_0 + \sum_{k=1}^{10} \beta_k x_{ik} + \sum_{k=1}^8 \beta_{ik9} x_{ik} x_{i9} + \sum_{k=1}^8 \beta_{ik10} x_{ik} x_{i10} + \beta_{28} s_j$, $i = 1, \dots, 108$ and $j = 1, \dots, 7$. 1000,000 samples from the posterior distribution were generated using the MCMC algorithm in Appendix D and were thinned to 100,000 samples by keeping every 10th sample. The posterior densities for the covariance parameters along with the plots used to check for convergence of the MCMC algorithm are shown in

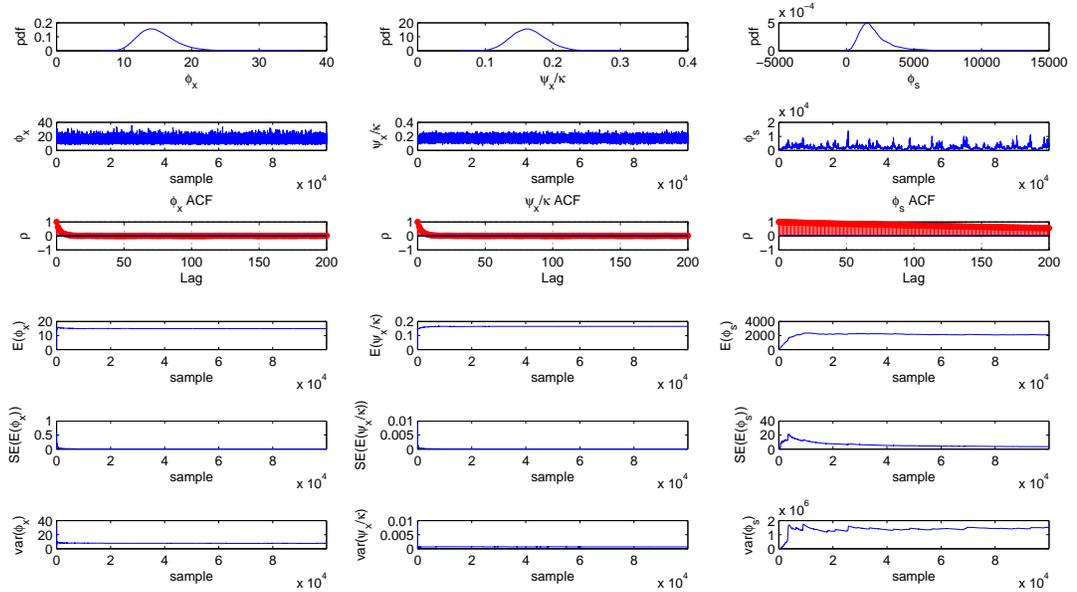


Figure 5.3. MCMC convergence plots, metal injection example. From top to bottom: posterior densities, trace plots, autocorrelation functions (ACF), expected value plots, MCMC standard error of the estimated posterior mean, and posterior variance plots of the generated covariance parameters (ϕ_x , ψ_x/κ , ϕ_x) (one column of plots per parameter). In here, we show the plots of the ratio ψ_x/κ since it is known in the Spatial Statistics literature (see e.g., Banerjee et al., 2004) that the ratio of these two parameters converges to its true distribution but not each of the two parameters separately.

Figure 5.5. The MCMC samples were used in MATLAB's *ga* optimization routine to maximize $p(\mathbf{x}_c)_{RPD}$. The optimization was constrained to $-1 \leq x_k \leq 1$ for $k = 1, \dots, 8$ and the noise factors were assumed to be independent *unif*(-1,1) random variables. Joint and marginal probabilities of conformance of the optimal solution found are shown in Table 5.2. The mean, 10th and 90th percentiles of the posterior predictive density at the optimal solution found are shown in Figure 5.6. The optimal solution found is different than the one given by Nair et al. (2002) due to the inclusion of the uncertainty in the model parameters. It is noticeable in Table 5.2 and Figure 5.6 that the specifications are violated mostly at the last two locations. If the specification limits at these two locations can be widened, the conformance probability evidently will improve.

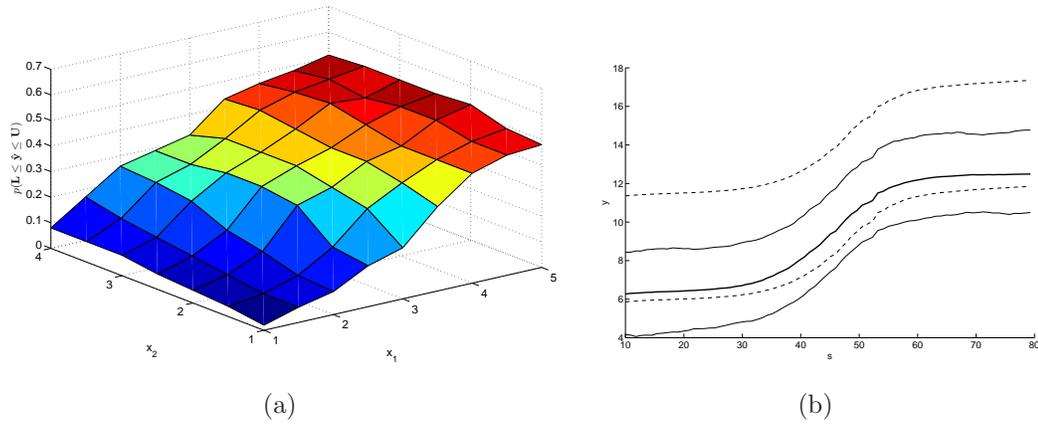


Figure 5.4. Metal injection example: (a) Probability of conformance to specifications as a function of design factors. (b) Predicted profiles at the optimal design factor settings. The mean of the posterior predictive density is shown with thick solid line. The other two solid lines are the 10th and 90th percentiles of the posterior predictive density. Upper and lower specification limits are shown in dashed lines.

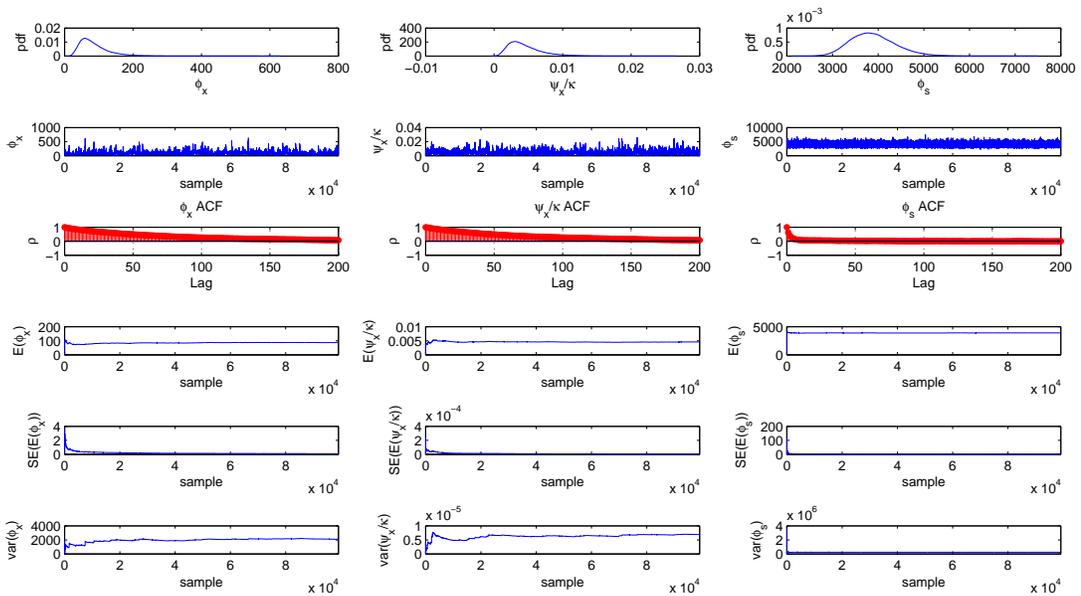


Figure 5.5. MCMC convergence plots, electric alternator example. From top to bottom: posterior densities, trace plots, autocorrelation functions (ACF), expected value plots, MCMC standard error plots and variance plots of the generated covariance parameters ($\phi_x, \psi_x/\kappa, \phi_s$) (one column of plots per parameter).

Table 5.2. Optimization results for the electric alternator example. Numbers in parentheses are the standard errors associated with probability estimation.

x_{opt}	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
	0.059	0.640	0.759	-0.966	0.558	0.846	-0.847	0.977
x_{opt} rounded	0	1	1	-1	1	1	-1	1
$p(L < \mathbf{y} < U)$	0.325(0.0148)							
$p(L_i < y_i < U_i)$	y_1	y_2	y_3	y_4	y_5	y_6	y_7	
	0.8	0.83	0.79	0.69	0.69	0.58	0.50	
	(0.0126)	(0.0119)	(0.0129)	(0.0146)	(0.0146)	(0.0156)	(0.0158)	

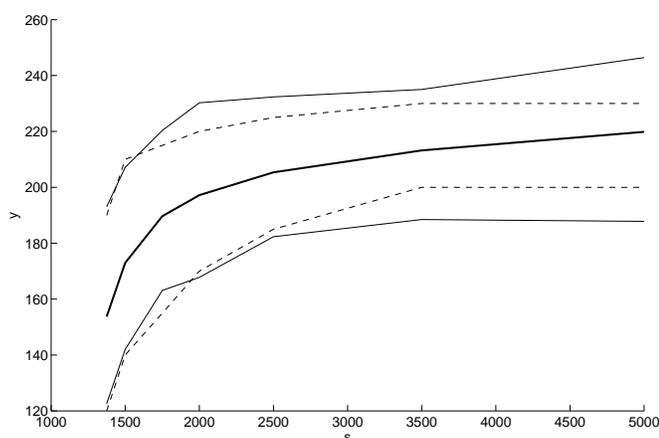


Figure 5.6. Predicted profiles at the optimal controllable factor settings for the electric alternator example. The mean of the posterior predictive density is shown in solid thick line. The other two solid lines are the 10th and the 90th percentiles of the posterior predictive distribution. The upper and lower specification limits are shown by dashed lines.

Example 3, with noise factors: plastic injection moulding process. Wu and Hamada (2000) report an experiment where the amount of plastic injected in a plastic injection process needs to be controlled. The response of interest was the part weight measured at eight values of the injection pressure, namely, 650, 700, 750, 800, 850, 900, 950 and 1000. The experiment consisted of seven controllable factors each at two levels, shown in Table 5.3, and four noise factors: melt index, percent regrind, operator and resin moisture. To reduce the experiment cost, a single compound noise factor at two levels was used instead (see Table 5.4). The conducted experiment was a 2^{7-4} replicated four times resulting in a 64 profiles each of eight locations. The experimental goal is to achieve part weights between

650 and 750 at different injection pressure values. The observed data profiles along with the target limits are shown in Figure 5.7. This Figure shows clearly that the data is not stationary and need to be detrended. The model was fitted assuming a first order mean structure in design factors and locations and all two factor control-noise interactions, that is $\mu(\mathbf{x}_i, s_j) = \beta_0 + \sum_{k=1}^8 \beta_k x_{ik} + \sum_{k=1}^7 \beta_{k8} x_{ik} x_{i8} + \beta_{17} s_j$, $i = 1, \dots, 64$ and $j = 1, \dots, 7$. 100,000 MCMC samples were generated. The MATLAB's `ga` optimization routine was used to maximize $p(\mathbf{x}_c)_{RPD}$. The optimization was constrained to $-1 \leq x_k \leq 1$ for $k = 1, \dots, 7$ (coded units). Noise factor was assumed to have *uniform*(-1, 1) distribution. The optimal solution found along with the optimal conformance probability are shown in Table 5.5. Figure 5.8 shows the mean, 10th and 90th percentiles of the posterior predictive density at the found optimal solution. It can be seen in Table 5.5 that the first injection pressure location (650) is the one drives the joint conformance probability. If a restriction can be imposed on the injection pressure value to not get below 700, then a 100% conformance probability can be achieved.

Table 5.3. Control factors in the plastic injection moulding experiment by Wu and Hamada (2000).

Factor	Level	
	-1	1
x_1 : injection speed	0	2
x_2 : clamp time (seconds)	44	49
x_3 : high injection time (seconds)	6.3	6.8
x_4 : low injection time (seconds)	17	20
x_5 : clamp pressure (psi)	1700	1900
x_6 water cooling ($^{\circ}$ F)	70	80
x_7 : low injection pressure (psi)	550	650

5.2 Model Robustness and Validation

The modeling approach discussed in section 5.1 has the following two main assumptions:

1. The data is multivariate normally distributed with mean $\mathbf{F}\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$.

Table 5.4. Noise factors in the plastic injection moulding experiment by Wu and Hamada (2000).

Label	Noise factor	Level
$x_8 = -1$	melt index	18
	percent regrind	5
	operator	new
	resin moisture	high
$x_8 = 1$	melt index	22
	percent regrind	0
	operator	experienced
	resin moisture	low

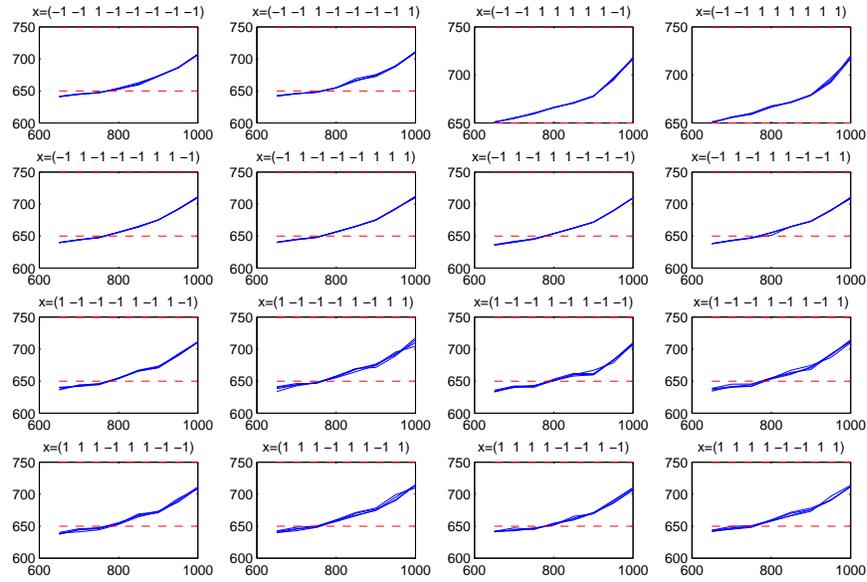


Figure 5.7. Observed profiles at each of the 16 design factor treatments in the plastic injection process example. The specifications $L(s_j)$ and $U(s_j)$ are shown by dashed lines.

2. The covariance matrix Σ can be decomposed into a Kronecker product of Σ_x and Σ_s .

In this section, we discuss the robustness of the proposed methodology with respect to these two assumptions, followed by a discussion on how to verify the second assumption.

Table 5.5. Optimal solution and point wise probabilities of conformance for the plastic injection moulding data. Numbers in parentheses are the standard errors associated with probability estimation.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
x_{opt}	-1	-0.9997	0.9997	0.9997	1	0.9997	1	
x_{opt} rounded	-1	-1	1	1	1	1	1	
$p(L < \mathbf{y} < U)$	0.544(0.0158)							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
$p(L_i < y_i < U_i)$	0.544 (0.0158)	1 (0.0)						

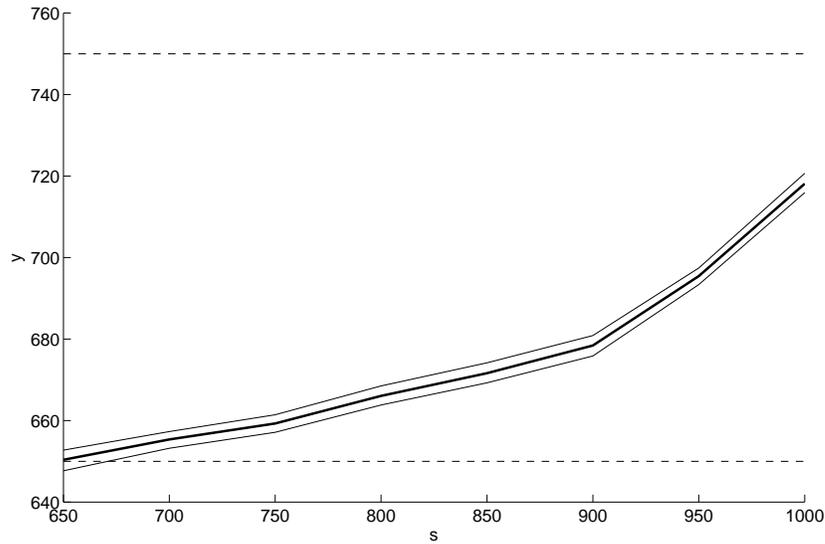


Figure 5.8. Predicted profiles at the optimal controllable factor settings. The mean of the posterior predictive density is shown in solid thick line. The other two solid lines are the 10th and 90th percentiles of the posterior predictive density. The upper and lower specification limits are shown in dashed lines.

5.2.1 Robustness to Normality Assumption

As usually performed in linear regression analysis (e.g., Bastos and O’Hagan, 2009; Shen and Xu, 2007), the normality assumption can be verified through normal probability plots of the residuals. Let \mathbf{R} be an $NJ \times 1$ vector of residuals such

that

$$\mathbf{R} = \text{vec}(\mathbf{Y}') - \mathbf{F}\hat{\boldsymbol{\beta}} \quad (5.13)$$

where $\hat{\boldsymbol{\beta}}$ is the mean of the posterior distribution of $\boldsymbol{\beta}$. Then

$$\begin{aligned} \text{var}(\mathbf{R}) &= \text{var}(\text{vec}(\mathbf{Y}') - \mathbf{F}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{\Sigma} + \mathbf{F}\text{var}(\hat{\boldsymbol{\beta}})\mathbf{F}'. \end{aligned} \quad (5.14)$$

Under a non-informative (flat) prior for $\boldsymbol{\beta}$ we have that $\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}$ (see Gelman et al., 2004, page 374), and therefore

$$\text{var}(\mathbf{R}) = \boldsymbol{\Sigma} + \mathbf{F}(\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}'. \quad (5.15)$$

Hence, if \mathbf{G} is the standard deviation matrix such that $\text{var}(\mathbf{R}) = \mathbf{G}\mathbf{G}'$, the elements of the vector $\mathbf{R}_G = \mathbf{G}^{-1}\mathbf{R}$ are standardized uncorrelated residuals with unit variance. A normal probability plot can then be used to check the normality of \mathbf{R}_G .

To check how robust model (5.4-5.8) is with respect to non-normal errors, we simulated the following two stage model to generate data we then fit to our model:

$$\begin{aligned} y_{ij} &= \mathbf{h}(s_j)\boldsymbol{\theta}_i + \sigma_\epsilon\epsilon \\ \boldsymbol{\theta}_i &= \mathbf{B}\mathbf{g}(\mathbf{x}_i) + \sigma_w\mathbf{w}, \quad \mathbf{w} \sim^{iid} N(0, 1). \end{aligned} \quad (5.16)$$

Assume there are two factors x_1 and x_2 each at three levels $\{1, 2, 3\}$ and the response is measured at fifty locations, namely $\mathbf{s} = \{1, 2, 3, \dots, 49, 50\}$. Let $\mathbf{g}(\mathbf{x}_i)$ be a full quadratic model in x_1 and x_2 such that $\mathbf{g}(\mathbf{x}_i) = [1, x_{i1}, x_{i2}, x_{i1}x_{i2}, x_{i1}^2, x_{i2}^2]$ and let $\mathbf{h}(s_j) = [1, \sin(0.1s_j), \sin(0.1s_j)^2]$. Two distributions were assumed for ϵ , a $N(0, 1)$ and a t with 5 degrees of freedom. The constant σ_ϵ was changed at four levels of $\{0, 0.5, 1, 1.5\}$ while σ_w kept constant at 0.5. To fit our model, an additive linear model in \mathbf{x} and \mathbf{s} was assumed for the mean structure, i.e., $\mathbf{f}(\mathbf{x}_i, s_j) = \{1, x_{i1}, x_{i2}, s_j\}$. The MCMC algorithm was run for 100,000 iterations. The mean, 10th and 90th percentiles of the posterior predictive density were estimated based on 1000 simulated profiles at each of the design treatments. Residuals

were then calculated as the difference between the profiles used for model fitting and the mean of the posterior predictive density. Figures 5.9(a) and 5.9(b) show the observed (simulated) and predicted profiles when $\sigma_e = 1.5$ for both cases of errors distribution. These two figures show that model (5.4-5.8) provides a good fit for the simulated data even under the case of t distributed errors. Figure 5.10 shows the normal probability plot of the residuals at the 8 combinations of σ_e and error distribution. It can be seen in Figure 5.10 that the model provides acceptable normal probability plots under normal and t distributed errors.

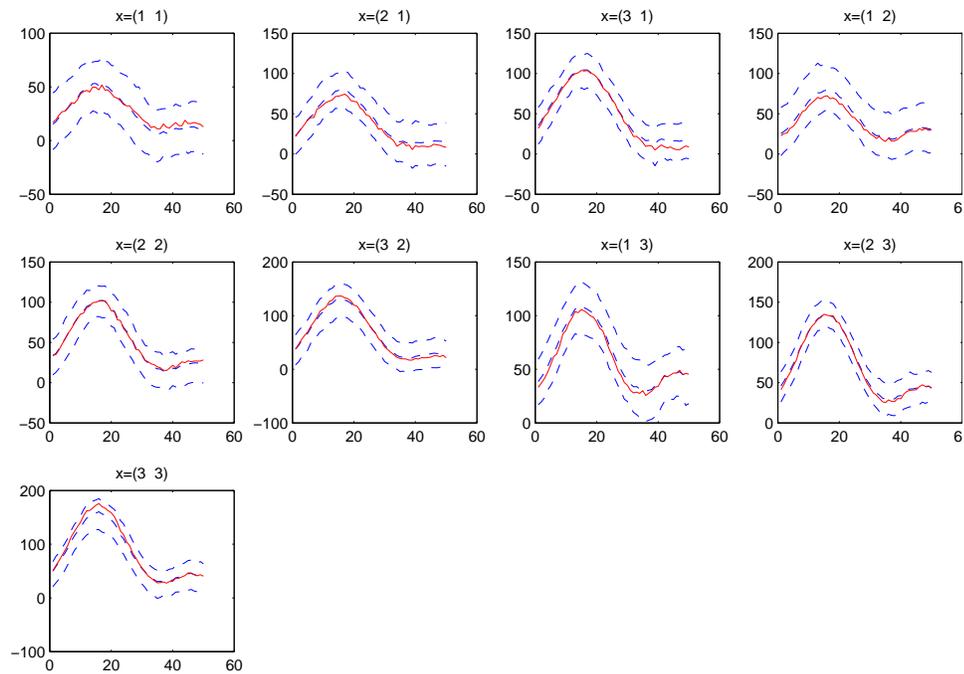
5.2.2 Robustness with Respect to Separability and Mean Structure Misspecification

Since in the proposed methodology the fitted model is used for process optimization, the reliability of the estimated optimal solutions is important. In this section, we test the robustness of the proposed model for

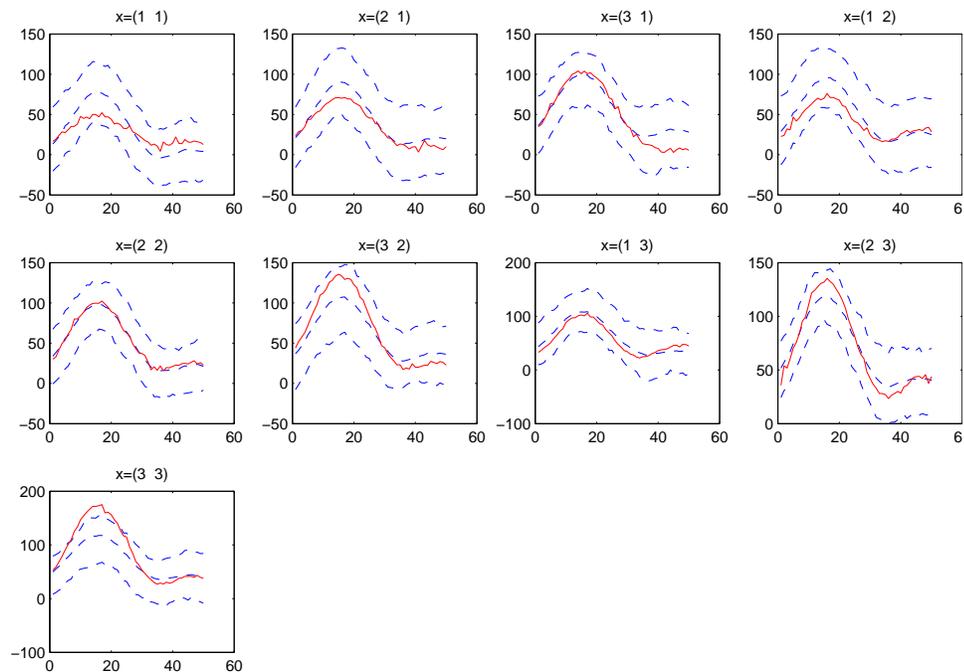
1. Mean structure misspecification,
2. The separability assumption.

For the purposes of the robustness analyses conducted in this section, we assumed there are two factors x_1 and x_2 , each at five levels $\{1, 2, 3, 4, 5\}$, and each profile is measured at 50 locations $\mathbf{s} = \{1, 2, \dots, 50\}$. Three mean structures were assumed. The first one is an intercept only, that is $\mu(\mathbf{x}_i, s_j) = \beta_0$. The second mean structure is an additive model in design factors and locations, that is $\mu(\mathbf{x}_i, s_j) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 s_j$. The third mean model is an interaction mean structure, that is $\mu(\mathbf{x}_i, s_j) = \beta_0 + \beta_1 s_j + \beta_2 x_{i1} + \beta_3 x_{i1} s_j + \beta_4 x_{i2} + \beta_5 x_{i2} s_j$.

To assess the proposed model performance under mean structure misspecification, we assume an iid $N(\mathbf{0}, 0.2\mathbf{I})$ error structure. For each mean structure, the simulated profiles were fit once to each of the three same mean models. Hence, we have 9 model fit combinations (3 forms for the true mean structure used for data generation and 3 forms for model fitting). The fitted and actual models were then used to find the best design factor settings that maximizes the probability the generated/predicted profiles fall between the specification limits. The specification



(a)



(b)

Figure 5.9. Observed vs predicted profiles at (a) $\sigma_e = 1.5$ and $\epsilon \sim N(0, 1)$, (b) $\sigma_e = 1.5$ and $\epsilon \sim t_5$ used in the normality robustness analysis. Solid lines represent the observed profiles used for model fitting, the mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines.

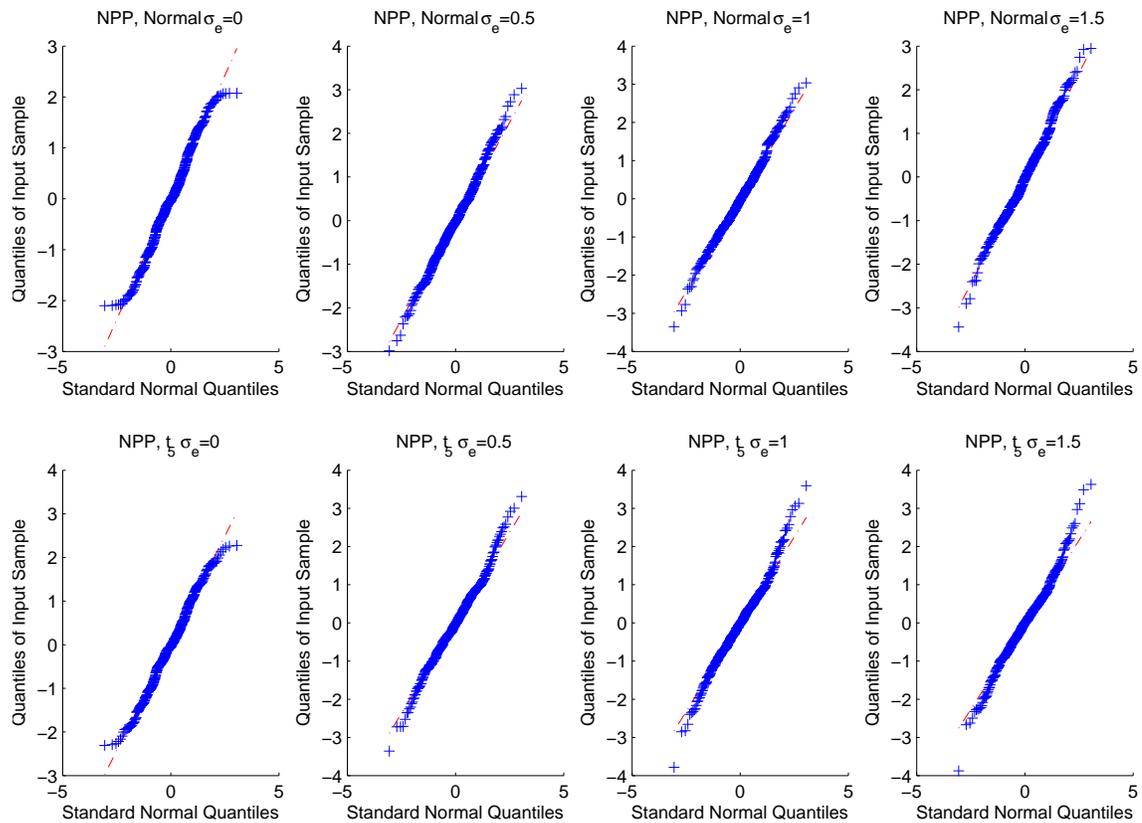


Figure 5.10. Normal probability plots for the different error distributions considered in the normality robustness analysis.

limits were set to ± 2 of the mean profile at factor settings $x_1 = 2$, $x_2 = 2.5$. Specification limits were kept fixed for all simulations. Optimization results, shown in Table 5.6 below, show that the proposed model is robust to misspecifications of the mean structure since it provided low relative errors in the estimated optimal solution. Here, the relative error is calculated as the norm of the difference between the true and estimated optimal solutions divided by the norm of the true optimal solution. They also show that the performance of the model is improved once the data is detrended to obtain stationary residuals, i.e., when the true mean structure has a trend, a first order model or an interaction model is needed rather than an intercept term only model.

Table 5.6. True vs. estimated optimal solutions for the 9 model combinations. Numbers in parentheses are the standard errors associated with the conformance probability estimation. The relative error is calculated as $e_x = \|\mathbf{x}_{true}^* - \mathbf{x}_{estimated}^*\|_F / \|\mathbf{x}_{true}^*\|_F$. In case the optimizer returns multiple optimal solutions, only one of them is listed.

True Mean	True optimal solution			Fitted Mean	Estimated optimal solution			e_x
	x_1^*	x_2^*	\hat{p}		x_1^*	x_2^*	\hat{p}	
Intercept	2.3104	1.000	1(0.0)	Additive	2.310	1.000	1(0.0)	0.000
				Interaction	2.310	1.000	1(0.0)	0.000
				Intercept	2.310	1.000	1(0.0)	0.000
additive	1.2803	2.819	1(0.0)	Additive	1.405	2.757	1(0.0)	0.002
				Interaction	1.155	2.757	1(0.0)	0.002
				Intercept	2.310	1.000	0(0.0)	0.456
interaction	1.187	2.827	1(0.0)	Additive	1.405	2.757	0.01(0.00031)	0.006
				Interaction	1.155	2.757	1(0.0)	0.001
				Intercept	3.666	2.827	0.01(0.00031)	0.654

To assess the proposed model robustness to the separability assumption, we used the following model for data generation:

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_{NJ}(\mathbf{0}, \boldsymbol{\Sigma})$$

where \mathbf{F} is defined as in (5.3) and the ij^{th} element of $\boldsymbol{\Sigma}$ has the following structure:

$$\Sigma_{ij} = \frac{\sigma^2}{(a(d_s)^{2\alpha} + 1)^{\delta + \gamma d/2}} \exp\left(-\frac{cd_x}{(a(d_s)^{2\alpha} + 1)^{\gamma/2}}\right) \quad (5.17)$$

where d_x is the distance in the x -space between the factor settings at which the i^{th} and j^{th} elements of $vec(\mathbf{Y}')$ were observed, d_s is the distance in the s -space between the locations at which the i^{th} and j^{th} elements of $vec(\mathbf{Y}')$ were observed, d is the dimension of the x -space, and a , c and α are constants define the smoothness of this covariance function. Notice that $\boldsymbol{\Sigma}$ is the multiplication of two terms. The first is $\sigma^2/(a(d_s)^{2\alpha} + 1)^{\delta + \gamma d/2}$ which is a function of d_s only. The second is $\exp(-cd_x/(a(d_s)^{2\alpha} + 1)^{\gamma/2})$ which is a function of both d_s and d_x . In this model proposed by Genton (2007), if $\gamma = 0$ the second term reduces to $\exp(-cd_x)$ which depends on d_x only, and hence $\boldsymbol{\Sigma}$ will be separable. Otherwise ($\gamma > 0$) the second term is not separable, and then $\boldsymbol{\Sigma}$ has a non-separable structure. The covariance structure parameters were set to $\sigma^2 = 0.2$, $a = 1$, $c = 1$, $d = 2$, $\delta = 1$ and $\alpha = 1$. Again, we assumed three mean structures as before and the generated profile data

at each mean structure was fit once to the three same mean models. Three values for the constant γ were assumed, $\{0.1, 0.5, 1\}$, resulting in nonseparable covariance structures.

Table 5.7 shows the estimated optimal solutions for the 27 combinations (of the true mean model, assumed mean model and the nonseparability parameter γ) tried along with the true optimal solution for each case. Note that in all cases in the table, a nonseparable covariance matrix was assumed, since $\gamma > 0$. As can be seen when the correct mean structure model is fit (i.e., only the covariance structure separability is assumed and no mean structure misspecification is present), the relative error is smallest. More interestingly fitting a wrong mean structure model (i.e., misspecified mean structure and nonseparable covariance structure) provides optimal solutions with low relative error values. This shows that the fitted model (5.4-5.8) is robust with respect to the assumed mean structure and is also robust with respect to this nonseparable covariance structure.

5.2.3 Checking the Covariance Separability Assumption

If the stochastic process that generates the observed profile data is separable (i.e., no significant spatial-temporal interaction), then the covariance structure of that process is separable (Genton, 2007). A separable process can be written as the product of two independent processes, one purely spatial and a second purely temporal. Let \mathbf{Y}^* be the detrended observed data matrix. Then using singular value decomposition (SVD), \mathbf{Y}^* can be written as

$$\mathbf{Y}^* = \mathbf{U}\mathbf{\Delta}\mathbf{V}' = \sum_{i=1}^{\min(N,J)} \delta_i \mathbf{u}_i \mathbf{v}_i'$$

where $\delta_i \mathbf{u}_i \mathbf{v}_i'$ is called the i^{th} Empirical Orthogonal Function (EOF) (Banerjee et al., 2004, section 8.1.1). If \mathbf{Y}^* can be approximated by its first EOF, $\delta_1 \mathbf{u}_1 \mathbf{v}_1'$, and \mathbf{u}_1 and \mathbf{v}_1 are independent, then \mathbf{Y}^* is separable (Banerjee et al., 2004). Therefore, the following three step procedure provides an easy and fast way to check how reasonable the separability assumption is for the observed data:

1. Assume a mean function form and use Least Squares to detrend the data such that $vec(\mathbf{Y}^*) = vec(\mathbf{Y}) - \mathbf{F}\hat{\boldsymbol{\beta}}$. The mean structure might include an

Table 5.7. True vs. estimated optimal solutions for the 9 model combinations each at 3 levels of the non-separability parameter γ . Numbers in parentheses are the standard errors associated with the conformance probability estimation. The relative error is calculated as $e_x = \|\mathbf{x}_{true}^* - \mathbf{x}_{estimated}^*\|_F / \|\mathbf{x}_{true}^*\|_F$. In case the optimizer returns multiple optimal solutions, only one of them is listed.

True Mean	γ	True optimal solution			Fitted Mean	Estimated optimal solution			e_x
		x_1^*	x_2^*	\hat{p}		x_1^*	x_2^*	\hat{p}	
Intercept	0.1	2.310	1.000	1(0.0)	Additive	2.310	1.000	1(0.0)	0.000
					Interaction	2.310	1.000	1(0.0)	0.000
					Intercept	2.310	1.000	1(0.0)	0.000
	0.5	2.310	1.000	1(0.0)	Additive	2.310	1.000	1(0.0)	0.000
					Interaction	2.310	1.000	1(0.0)	0.000
					Intercept	2.310	1.000	1(0.0)	0.000
	1	2.310	1.000	1(0.0)	Additive	2.310	1.000	1(0.0)	0.000
					Interaction	2.310	1.000	1(0.0)	0.000
					Intercept	2.310	1.000	1(0.0)	0.000
Additive	0.1	1.155	2.882	1(0.0)	Additive	1.405	2.757	1(0.0)	0.008
					Interaction	1.405	2.757	1(0.0)	0.008
					Intercept	1.280	2.507	0.05(0.0069)	0.016
	0.5	1.155	2.882	1(0.0)	Additive	1.405	2.757	1(0.0)	0.008
					Interaction	1.405	2.757	1(0.0)	0.008
					Intercept	1.815	2.507	0.08(0.0086)	0.060
	1	1.155	2.882	1(0.0)	Additive	1.405	2.757	1(0.0)	0.008
					Interaction	1.405	2.757	1(0.0)	0.008
					Intercept	4.201	4.868	0.02(0.0044)	1.372
Interaction	0.1	1.155	2.850	1(0.0)	Additive	1.405	2.757	0.1(0.0095)	0.008
					Interaction	1.405	2.757	1(0.0)	0.008
					Intercept	1.940	2.513	0.07(0.0081)	0.077
	0.5	1.155	2.850	0.999(0.0010)	Additive	1.468	2.757	0.06(0.0075)	0.011
					Interaction	1.405	2.757	1(0.0)	0.008
					Intercept	1.437	2.694	0.05(0.0069)	0.011
	1	1.187	2.835	0.999(0.0010)	Additive	1.155	2.757	0.03(0.0054)	0.001
					Interaction	1.405	2.757	1(0.0)	0.006
					Intercept	1.280	2.757	0.05(.0054)	0.002

intercept, main effects of design factors and locations, design factor location interactions, etc.

2. Construct the singular value decomposition of \mathbf{Y}^* and its approximation $\mathbf{Y}_{approx}^* = \delta_1 \mathbf{u}_1 \mathbf{v}_1'$ where δ_1 is the largest eigenvalue, and \mathbf{u}_1 and \mathbf{v}_1 are the left and right singular vectors, respectively, corresponding to the largest eigenvalue.
3. Graph both \mathbf{Y}^* and \mathbf{Y}_{approx}^* vs. \mathbf{s} . If the two plots look similar, then sep-

arability is a reasonable assumption. A more precise metric is the relative approximation error defined as:

$$e_{approx} = \frac{\|\mathbf{Y}^* - \delta_1 \mathbf{u}_1 \mathbf{v}_1'\|_F}{\|\mathbf{Y}^*\|_F}$$

where $\|\cdot\|$ is the Frobenius norm.

The construction of the normal probability plots depends on Σ , but Σ is approximated by a separable covariance structure. Hence, the normal probability plots would be misleading if the covariance structure is significantly non-separable. If the data is approximately normal and the predictions are acceptable, the model is useful for optimization purposes. Here, we suggest to use cross validation techniques (Hastie et al., 2009) for model assessment. In this chapter, we use two cross validations, one based on leave-one-out predictions and a second one based on leave-10%-out predictions. In leave-one-out predictions, one of the observed profiles is left out for testing and the remaining $N - 1$ profiles are used for model fitting. The fitted model is then used to predict the left out profile. In leave-10%-out, we instead leave 10% of the observed profiles out and use the remaining 90% for model fitting. Then the fitted model is used to predict the 10% profiles left out. If the fitted model provides acceptable predictions for testing profiles, then it would provide acceptable predictions at other locations in the x -space.

Example 1 (cont.). Recall the metal injection process example mentioned in the previous section. To check the observed elastic modulus data for separability, we assume three mean structures. The first is an intercept model only, that is $\mu(\mathbf{x}_i, s_j) = \mu$. The second assumes an additive mean structure in the design factors and locations, that is $\mu(\mathbf{x}_i, s_j) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 s_j$. The third mean structure assumes an interaction model, that is $\mu(\mathbf{x}_i, s_j) = \beta_0 + \beta_1 s_j + \beta_2 x_{i1} + \beta_3 x_{i1} s_j + \beta_4 x_{i2} + \beta_5 x_{i2} s_j$. Figure 5.11 shows the plot of \mathbf{Y}^* and \mathbf{Y}_{approx}^* for the three assumed models. The relative error statistics were 0.48887 for the intercept model, 0.28234 for the additive model and 0.28187 for the interaction model. Looking at the plots and the relative error values, it can be seen that an additive mean structure is acceptable, since it provides a simple model with a relative error almost equal to the most complicated (interaction) model.

The additive model fitted for this data was further checked by, leave-one-out and leave-10%-out cross validations were performed. Cross validation results are shown in Figures 5.12 and 5.13. Based on these figures it can be seen that the model provides acceptable predictions. Hence, the fitted model can be used for process optimization.

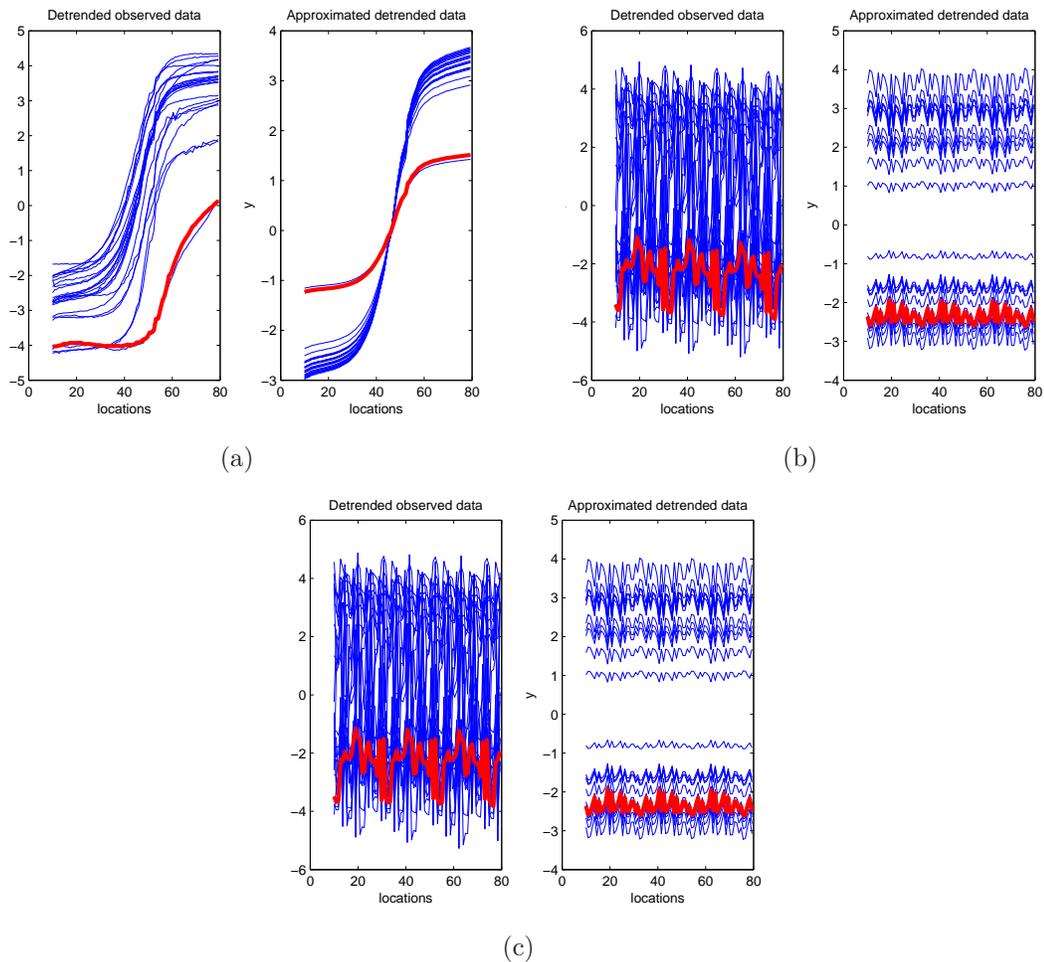


Figure 5.11. Checking the separability assumption, metal injection example: plots of \mathbf{Y}^* and \mathbf{Y}_{approx}^* for (a) intercept mean structure (b) additive mean structure and (c) interaction mean structure. The first row of the observed and approximated data matrix is shown with thick line.

Example 2 (cont.). Recall the electric alternator design example mentioned in section 5.1. To check the observed profile data of the electric current for separability, we assume three mean structures: an intercept model only, an additive

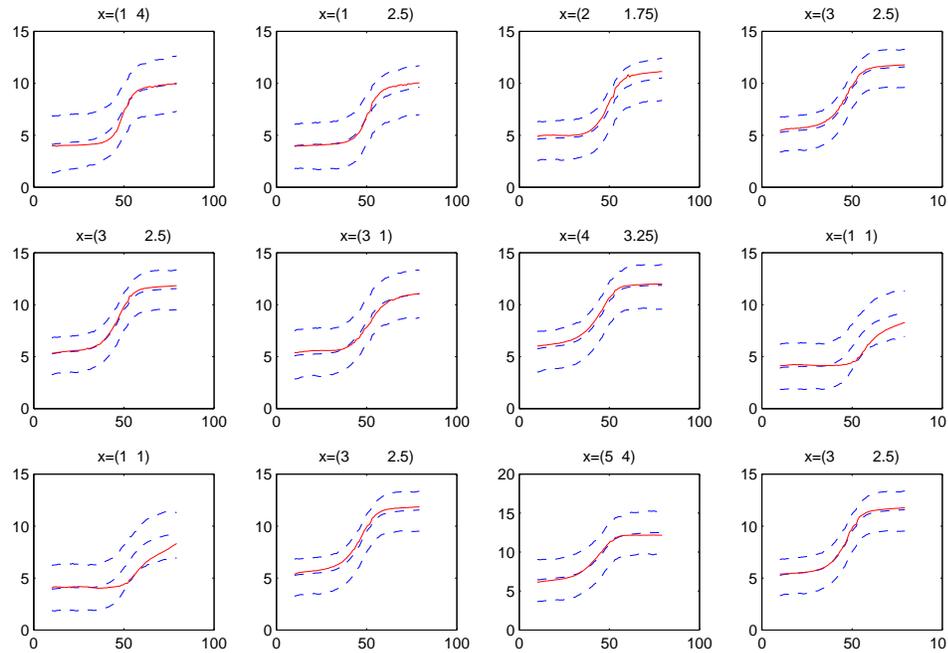


Figure 5.12. Leave-one-out cross validation results for the metal injection example. Solid lines are the observed profiles used for testing. The mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines. Typically, there would be 24 plots since there are 24 observed profiles, but we only show 12 of them here.

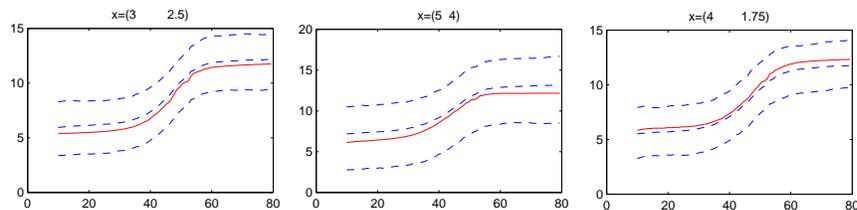


Figure 5.13. An instance of leave-10%-out cross validation results (10% \approx 3 out of 24 observed profiles) for the metal injection example. Solid lines are the observed profiles used for testing. The mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines.

mean structure in the design factors and locations, and an interaction model. The relative error statistics were 0.85403 for the intercept model, 0.64357 for the additive model and 0.63703 for the interaction model. These relative error values

suggest that this data is significantly nonseparable, but if we were to choose a model among these three, then an additive mean structure is acceptable, since it provides a simple model with a relative error almost equal to the most complicated (interaction) one. But since we have noise factors in this example, all two-factor control \times noise interaction terms are also need to be added to the mean structure model to allow for RPD optimization, i.e., the used model has an intercept, main effects and all control \times noise interactions but not control \times control or noise \times noise interactions.

To validate the fitted model, for the alternator data, in example 2 before, leave-one-out and leave-10%-out cross validations were performed. Cross validation results are shown in Figures 5.14 and 5.15. Based on cross validation results it can be seen that the model provides acceptable predictions, and hence it can be used for process optimization.

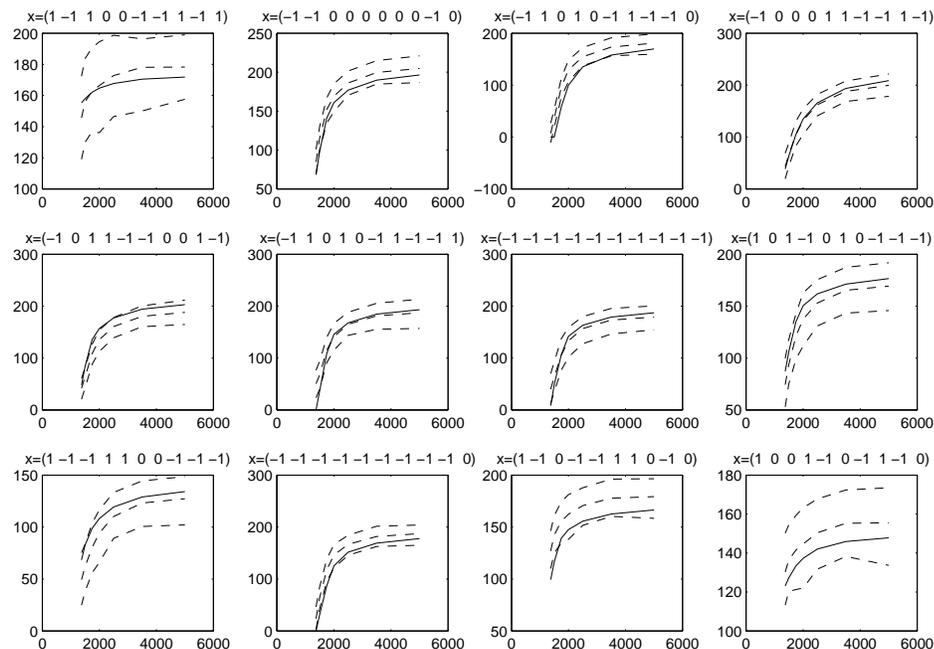


Figure 5.14. Leave-one-out cross validation results for the electric alternator example. Solid lines are the observed profiles used for testing. The mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines.

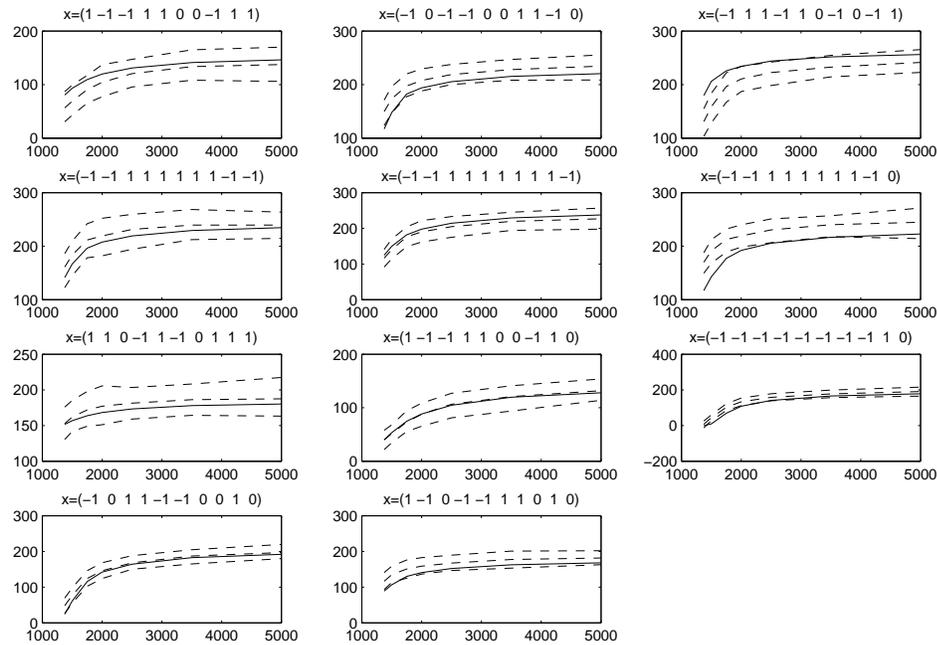


Figure 5.15. An instance of leave-10%-out cross validation results for the electric alternator example. Solid lines are the observed profiles used for testing. The mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines.

Example 3 (cont.). Recall the plastic injection process example mentioned in section 5.1. To check the observed profile data of the plastic injection process for separability, we assume three mean structures: an intercept model, an additive mean structure in the design factors and locations, and an interaction model. The relative error statistics were 0.45636 for the intercept model, 0.37021 for the additive model and 0.36061 for the interaction model. These relative error values suggest that an additive mean structure is acceptable. Again, since we have noise factors in this example, all two-factor control \times noise interaction terms are also need to be added to the mean structure model to allow for RPD optimization. Leave-one-out and leave-10%-out cross validations were performed to validate the fitted model for this data. Cross validation results are shown in Figures 5.16 and 5.17. These figures show that the model provides an excellent fit to the data, and hence it can be used for optimization purposes.

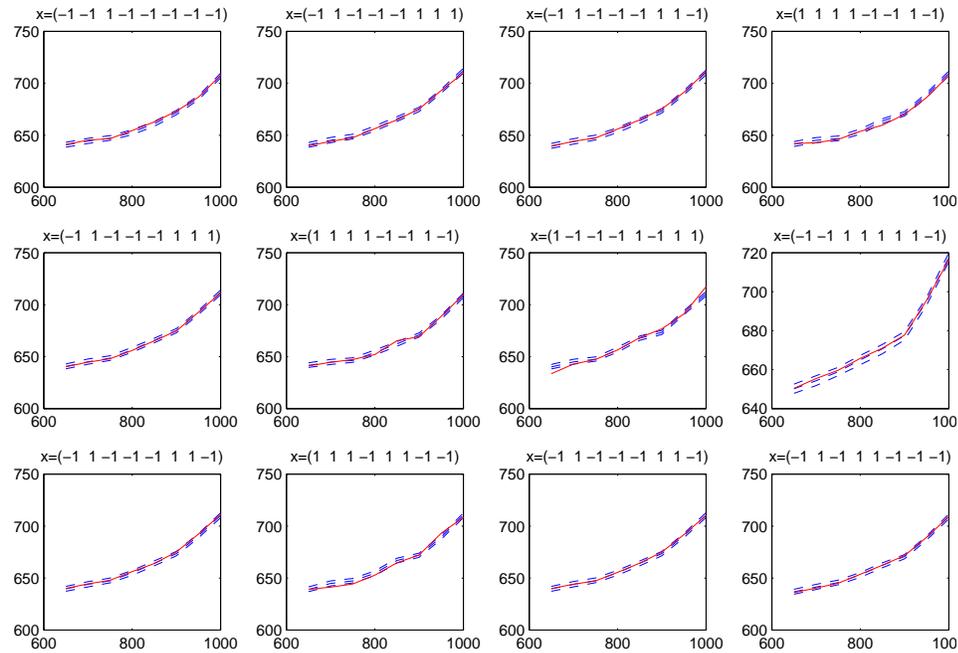


Figure 5.16. Leave-one-out cross validation results for the plastic injection process example. Solid lines are the observed profiles used for testing. The mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines.

5.3 Conclusions and Future Work of This Chapter

Profile responses arise in a variety of applications where a continuous response is sampled at some frequency or more generally over a specified set of locations. Current approaches to solve RPD problems of profile responses include the frequentist point of view modeling approach by Nair et al. (2002). This approach does not consider model parameter uncertainties and assumes small number of locations. A different approach is that by Del Castillo et al. (2011) where profiles are modeled in a two stage hierarchical regression model. Their approach may not be flexible enough since it requires a parametric model that provides a good fit for the mean. Furthermore, it neglects the profile-to-profile correlations.

This chapter introduces a new approach to solve the RPD problem of profile

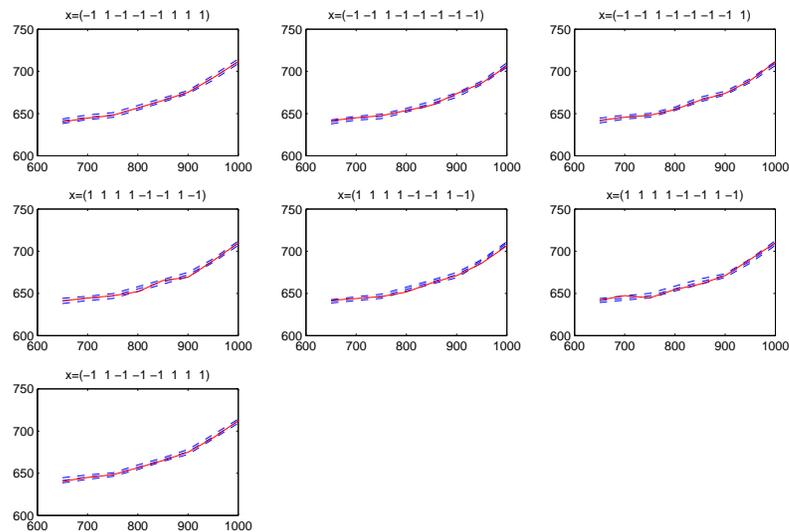


Figure 5.17. An instance of leave-10%-out cross validation results for the plastic injection process example. Solid lines are the observed profiles used for testing. The mean, 10th and 90th percentiles of the posterior predictive density are shown in dashed lines.

response systems based on a spatio-temporal Gaussian Random Function. The spatial and temporal spaces correspond to the design factor space and the location measurement space, respectively. In this approach, the observed data is assumed to have a multivariate normal distribution with mean structure μ and covariance structure Σ . Like in Universal Kriging practice, the main purpose of the mean structure is to detrend the data and obtain zero mean stationary residuals. However, to allow for a better Robust Parameter Design, we also suggest to include all control \times noise interaction terms (judged important by the experiment) in the mean structure model if noise factors are present. The covariance structure is assumed to be the Kronecker product of the between-profiles and the within-profile covariance structures (i.e., we assume covariance separability). As usually done in Spatial Statistics, a parametric model is assumed for both covariance structures.

The proposed model assumes data normality and separability of the covariance matrix. The first assumption is justified by the continuity of the data and large sample theory. Sensitivity to violations in the normality assumption was discussed in section 5.2.1. It was found that the model is robust and can provide good data

fit even for considerably non-normally distributed errors. Model performance to an instance of non-separable covariance structure was performed in section 5.2.2. It has been found that the model still provides a good data fit under that particular covariance structure. Non-separable covariance structures can have many forms, and hence it is impossible to study model performance for all of them. Instead, we suggest to use cross validations to assess the model performance regardless of covariance structure separability.

The proposed model has three main advantages over the currently available approaches. First, the modeling approach is fully Bayesian, and hence incorporates all uncertainty sources present. It also provides a probability measure for the process performance for a given factor settings. Second, no complicated parametric regression model is needed for the mean structure. The modeled covariance structure can modify the assumed mean to provide good predictions. Finally, the proposed model is more flexible, since no complicated mean structure form is required. Hence, it can model a wider range of profile shapes than previous approaches.

The proposed modeling approach assumes covariance structure separability to speed up computations, but this assumption might not hold for some profile response data. Hence, a non-separable covariance structure is an area where future research is needed. For example, the electric current data used in example 2 is clearly nonseparable. A non-separable covariance modeling approach that still has computational attractiveness is the covariance tapering approach proposed by Furrer et al. (2006). This approach has the advantage of generating sparse covariance matrices. Sparse matrices are easy to deal with from the computational point of view.

All GRF process modeling approaches assume a smooth stochastic process. If the observed process is not smooth, the GRF process assumption is not suitable. Hence, other modeling approaches (other parametric covariance models) should be considered.

Chapter 6

Research Contributions and Areas for Further Research

In this dissertation we have studied the analysis and optimization of shape and profile response systems. This chapter summarizes the results of the work presented in chapters 3 through 5. Section 6.1 gives a summary of this dissertation main contributions to the existing literature and section 6.2 discusses some possible areas for further research in the topics covered.

6.1 Research Contributions

We considered three research problems in this dissertation:

1. The performance of Statistical Shape Analysis (SSA) tests used to determine significance of factor effects in designed experiments where the response of interest is the 2D or 3D shape of an object.
2. Robust parameter optimization of shape responses, i.e, experiments where the response is the shape of an object but there are noise factors present in the experiment.
3. Robust parameter optimization of profile response systems.

The contributions of our work to each of these three problems are discussed in sections 6.1.1 through 6.1.3 respectively.

6.1.1 Statistical Performance of Shape Analysis Tests

The power performance of the statistical tests studied in chapter 3 depends on several factors, such as the response dimension, the sample size, the error assumptions, etc. Statistical tests to analyze experiments with shape responses have existed in fields other than manufacturing for the past two decades. Such tests have been recently introduced to analyze designed experiments for shape responses in manufacturing applications. There has been no study thus far that addresses the performance of these tests in manufacturing, in which, unlike other applications, a large number of measurements per shape is common.

Five tests found in the literature were considered in the statistical performance studies. These tests can be classified into two types based on the test statistic used:

1. Procrustes-based tests:

- The one-way ANOVA F-test (Goodall, 1991), and its two-way (interaction) extension (Del Castillo and Colosimo, 2011).
- The ANOVA Permutation test (Del Castillo and Colosimo, 2011).

2. Distance-based tests:

- EDMA-I (Lele and Richtsmeier, 1991).
- EDMA-II (Lele and Cole, 1996).
- ANOVA on the form error.

The latter test is the most common in manufacturing practice and applicable only when there exists a standard definition of form error for the geometrical feature of interest. The EDMA tests are available only for the one-way, two levels case.

The performance analyses conducted were classified according to the type of mean shapes utilized. First, the performance of tests for detecting differences in mean shape were studied for shapes with arbitrary geometry that are not commonly found in manufacturing, such as triangles and arbitrary polygons. These types of shapes were considered by previous SSA authors, and thus we included them to allow for comparisons. Second, we studied the performance of SSA tests

for shapes more commonly found in manufacturing, such as circles and cylinders. In the published data sets with arbitrary geometry, there is only a small number of landmarks per configuration, five at the maximum. The second type of shapes, those of interest in manufacturing, have not been studied from a SSA point of view before, and thus simulated data sets with considerably more landmarks per configuration were studied, imitating manufacturing applications where it is common to observe hundreds of measurements for each object.

We considered three variations of the error properties: isotropic and non-isotropic variances and non-normal errors. For the case of non-isotropic errors, three covariance structures that are found in the literature were utilized. Two non-normal error distributions were used: a t-student distribution and a uniform distribution.

The overall conclusion is that the most powerful and robust test is the ANOVA Permutation test in Del Castillo and Colosimo (2011). The ANOVA F-test is usually equally powerful, but suffers from higher than advertised type-I error rate when the error distributions are not normal or the variances not isotropic. The ANOVA on the circularity and cylindricity form errors showed little power overall, and the EDMA tests showed worst performance.

A detailed summary of the results found under the four categories of mean shapes and error properties considered is as follows:

1. *Mean shapes of arbitrary geometry under isotropic errors*: the ANOVA Permutation test and the ANOVA F-test showed the highest power performance. The ANOVA F-test showed a slightly lower than nominal type-I error rate.
2. *Circles (2D) and cylinders (3D) under isotropic errors*: both the ANOVA F-test and the ANOVA Permutation test showed the best power performance with type-I error rates close to nominal.
3. *Circles (2D) and cylinders (3D) under non-isotropic errors*: again, we found the ANOVA F-test and the ANOVA Permutation test to give the best power performance among all considered tests with higher power for the ANOVA F-test in some cases, but also with a higher type-I error rate than assumed. We observed that the power of these two tests improves as the number of landmarks per configuration increases.

4. *Circles under non-normal errors*: the ANOVA Permutation test showed the best power performance and type-I error rate control. The ANOVA F-test showed similar power to the ANOVA Permutation test but with a higher than nominal type-I error rate. In contrast, the ANOVA on the form errors provides good power performance only when the distribution of the errors is bounded since it is based on a range statistic.

6.1.2 Robust Parameter Optimization of Shape Responses

Chapter 4 of this dissertation discusses the robust parameter optimization problem for shape responses, i.e, the problem of achieving a desired shape in the presence of noise factors. Statistical tools for shape optimization in Del Castillo and Colosimo (2011) were extended to the Robust Parameter Design (RPD) case. Also, instead of testing for differences in shape and size separately (as these authors do) we study instead the *form* of the objects, defined as the geometrical information that remains when rotation and location effects (but *not* changes of scale) are discounted.

We proposed a statistical tool for geometric shape optimization that considers both controllable and noise factors using a mean-square error-like objective. The proposed approach uses the form of the object for optimization, and hence size information is preserved. The proposed approach was illustrated with several simulated and real manufacturing shape data. The results show that our method provides good solutions in terms of a mean shape close to the target with minimum variability.

6.1.3 Robust Parameter Optimization of Profile Responses

The third and last problem considered in this dissertation is the analysis and optimization of profile response systems, studied in chapter 5. A new approach based on a spatio-temporal Gaussian Random Function process was presented and an adaptive Markov Chain Monte Carlo (MCMC) algorithm with good convergence properties (for the assumed mode) was developed. The spatial and temporal spaces correspond to the design factor space and the location measurement space, respectively. In this methodology, the observed data is assumed to have a multivariate normal distribution with some mean and covariance structures. Similarly

to Universal Kriging practices, the mean structure is used for data detrending to obtain zero mean stationary residuals, but it is also used to allow the inclusion of control \times noise interaction effects, crucial in RPD. The covariance structure is assumed to be the Kronecker product of the between-profiles and the within-profile covariance structures. As usually done in Spatial Statistics, a parametric model is assumed for both covariance structures, to reduce the number of parameters to estimate.

The proposed model assumes data normality and covariance structure separability. The first assumption is justified by the continuity of the data and large sample theory. It was found through simulations that the model is robust and can provide good data fit even for non-normally distributed errors. We provided two model assessment techniques. The first is to compare the observed data and the reconstructed data assuming separability. The second is to assess model performance through cross validation.

The proposed model was illustrated through three real profile data examples taken from the literature. The three examples have different profile shapes, but the same model form was used for all of them. It was shown through cross validations that the model is able to provide good predictions for the three examples used.

The proposed model has three main advantages over the currently available methodology for this problem. First, the modeling is fully Bayesian, and hence incorporates all uncertainty sources present. It also provides a probability measure of the process performance. Second, no complicated parametric regression model is needed for the mean structure. The modeled covariance structure can modify the assumed mean to provide good predictions. Finally, the proposed model is more flexible, since no complicated mean structure form is required. Hence, it can model a wider range of profile shapes.

6.2 Areas for Further Research

Possible extensions and modifications of some of the topics covered in this dissertation include the following:

1. In chapter 3 a modified Generalized Procrustes Algorithm $GPA(\Sigma)$ was used to register shapes when the error structure is non-isotropic. Although the

method appears to converge to good solutions for the 2D cases studied in chapter 3, no formal proof of convergence exists for this algorithm. Hence, an area for further research is to study the convergence properties of $\text{GPA}(\Sigma)$.

2. In the derivation of the one- and two-way analysis of variance (ANOVA) tests it was assumed that the error variance is small. We stated in chapters 2 and 3 that the ANOVA identities break down when this assumption is not true. A recently proposed multivariate analysis of variance (MANOVA) test that does not require this assumption is that of Huckemann et al. (2010a). The performance of this test for manufacturing applications where small error assumption holds has not been studied yet, hence a future work possibility is to compare the performance of the existing ANOVA tests and the proposed MANOVA test by Huckemann et al. (2010a) for manufacturing applications. The assumption of small error variance holds for traditional manufacturing processes where advanced machines are used. For non-conventional manufacturing operations, such as nano- and micro-manufacturing, this assumption might not hold. Hence, statistical tests that do not require the small error variance assumption are required to study shape analysis for these applications. Given that Huckemann et al. (2010a) methods are based on the intrinsic geometry of shape space, rather than on its extrinsic properties as the methods in chapters 3 and 4 do, a first step in this direction is to adopt those methods to manufacturing applications. It is unclear how or if these methods can be applied with problems with many landmarks. Furthermore, the intrinsic methods have been developed only for 2D shapes.
3. The existing statistical shape tests consider the cases of one- and two-factor factorial designs only. An extension of the existing ANOVA tests to the case of K -factor factorial designs is indeed needed. An additional matter is that existing tests assume balanced designed experiments. The case of unbalanced designed experiments has not been considered yet and it might be of interest. Also, the performance of the tests considered for more shapes of interest in manufacturing, other than circles and cylinders, could be undertaken.
4. All SSA techniques considered in this dissertation assume a landmark repre-

sentation of shapes. Currently, other representation systems are present in the literature, for instance, contour-based shape representations (Zhang and Lu, 2004). A large unexplored area is the development of statistical tests for shape (or form) differences based on these alternative representations.

5. In solving the RPD problem of shape response experiments (chapter 4), the optimization is performed over the finite set of design factor treatments considered in the experimental design. To optimize over the whole design factors space, one possible future work idea is to fit a model to the calculated G_{Total} statistic in (4.6). The fitted model can then be used for interpolations at treatment locations other than the design treatments, and then an optimization algorithm can be used to solve the RPD problem over a continuous space.
6. If objects are aligned differently when using a Coordinate Measuring Machine, the measurements obtained will tend to have a fixture alignment error, i.e., measured points are not exactly the same in every object. This error increases the calculated error sum of squares in the ANOVA table in Table 2.1, hence it reduces the power of the F-test. A research opportunity is to decompose the total error term into an error term due to misalignment and an error term due to shape differences. Then the latter error term can be used to test for factor significance.
7. The spatio-temporal Gaussian Process model proposed in chapter 5 assumes an isotropic exponential covariance structure for the between profiles correlations, Σ_x . This assumption might not hold in practice (the computer experiments literature usually assumes non-isotropy), hence a non-isotropic covariance structure needs to be used. Development of methods for use of non-isotropic covariance structures such as the one in (3.14) is needed. This will require, first of all, development of new MCMC methods with reliable convergence properties, a hard task in itself.
8. If the interaction between the design factor space and the measurement locations space in a profile response system is significant (chapter 5), the separability assumption will not be appropriate. Other techniques that do

not require separability such as the variance tapering method of Furrer et al. (2006) could be employed. This method assumes the correlation between two points that distance more than d units to be zero. This will generate sparse covariance matrices where advanced algebraic methods can be used to speed up computations.

9. In chapter 5, we considered the case of one profile response observed at each experimental run. An extension of the proposed modeling approach to the case of multiple profile responses is an area for further research.
10. In chapter 5, it was assumed that the data is observed first and then the model is built based on the observed data. Jones et al. (1998) proposed a “sequential process optimization” approach where the observed data is used sequentially to fit a model and at each iteration the fitted model thus far is used to decide which design point to investigate next, until a local optimal solution is reached. This approach receives the name Efficient Global Optimization (EGO) in the literature (Jones et al., 1998). Their approach considers univariate responses and a frequentist point of view, although considerable discussion from a Bayesian point of view (although not a full Bayesian) is given by Santner et al. (2003). The EGO approach could be extended to the Bayesian case and applied to the profile response experiments considered in this dissertation to provide a sequential “profile to profile” optimizer.

Mathematical Notions Used in the Theory of Statistical Shape Analysis

This appendix reviews some of the key mathematical concepts used in the theory of Statistical Shape Analysis (SSA) as developed mainly by Kendall (1984). This appendix is based on a larger review by Del Castillo (2011) and is included here to help readers not familiar with Differential Geometry concepts to understand the mathematical notation and terminology used in SSA.

A.1 Relations, equivalence relations and equivalence classes

Definition 1. A *relation* on a set A is a subset, R , of $A^2 = A \times A$. Usually, relations are defined by providing a statement that singles out a collection of elements of $A \times A$ for membership in the relation. A relation R on a set A is:

- *reflexive* if for all $x \in A$, xRx .
- *symmetric* if, for all $x, y \in A$, xRy implies yRx .
- *transitive* if, for all $x, y, z \in A$, xRy and yRz imply xRz .
- an *equivalence relation* if R is reflexive, symmetric and transitive.

Equivalence relations are sometimes written with the symbol \sim ; thus, $x \sim y$ is read “ x is equivalent to y ”. Two elements of a set do not need to be equal to be equivalent, they need only to share a specified property.

Definition 2. Let A be a set and let \sim be an equivalence relation defined on this set. For each $a \in A$, the *equivalence class of a* is a subset, denoted $[a]_{\sim}$, consisting of all elements of A that are equivalent to a , i.e.,

$$[a]_{\sim} = \{x \in A : x \sim a\}$$

If there is no ambiguity about the equivalence relation one is talking about, the corresponding equivalence class is written $[a]$. Here the word “class” has been used historically to simply mean a set. Other names for equivalence class is an *orbit*, and, in case the underlying set is a manifold (see A.3 below) they are also called a *fibre*.

Theorem 1. Let \sim be an equivalence relation on A and let $x, y \in A$. Then 1) if $x \sim y$, then $[x] \sim [y]$; 2) if $x \not\sim y$, then $[x] \cap [y] = \emptyset$; 3) $A = \bigcup_{x \in A} [x]$.

Proof of 3): each equivalence class is a subset of A by definition. Each $x \in A$ is in the equivalence class $[x]$. Therefore, A is contained in the union of the equivalence classes of all the elements of A . Since from part 2) distinct equivalence classes do not intersect, this union is actually equal to set A .

Part 3) of the result above means that the set of all equivalence classes implied by an equivalence relation \sim forms a *partition* of A . Parts 1) and 2) say that if two equivalence classes have an element in common, then they are identical, or, in other words, that two distinct equivalence classes are always disjoint.

Definition 3. Let \sim be an equivalence relation in A . The set of all equivalence classes is called *A modulo \sim* or the *quotient of A by the equivalence relation \sim* , and is denoted A/\sim . The *projection map* $\pi : A \rightarrow A/\sim$ sends $x \in A$ to its equivalence class $[x]$. If the set A/\sim is closed under arbitrary unions and finite intersections (properties that define a *topology*), this set is called the *quotient space* of A by the

equivalence relation \sim .

Example 1. Quotient spaces A/\sim (and equivalence classes) are usually created by *identifying a subset of A to a point*. For instance, let $A = [0, 1]$ (unit interval on \mathbb{R}) and define the quotient space obtained from A by identifying the two endpoints $\{0, 1\}$ to be equivalent to the same point. Let S^1 be the unit circle on the complex plane \mathbb{C} . The function $f : A \rightarrow S^1$, $f = \exp(2\pi ix)$ equals the same value (1) at 0 and at 1, and hence it induces a function

$$g : A/\sim \rightarrow S^1.$$

Geometrically, points on the unit interval A are being mapped into the points of the unit circle on \mathbb{C} , with the two endpoints in A mapping into the same point on \mathbb{C} , namely the point $(1,0)$.

Definition 4. A *homeomorphism* is a mapping in Euclidean space from one object onto another that is continuous and one to one, i.e., it establishes a one to one correspondence between points in each figure. The inverse mapping has the same properties. For example, a sphere in \mathbb{R}^3 and a cube are homeomorphic.

The concept of homeomorphism is used to define the properties of objects (figures) that remain unchanged under continuous deformation (“rubber band deformations”). These properties are called the *topological* properties of the objects. Topological properties stand in contradistinction with *metrical* properties, which are associated with distances between points, angles between lines, and edges of a figure, properties that are preserved under rigid body transformations only.

Definition 5. An *n -sphere* S^n is a set of points in $(n + 1)$ -Euclidean space such that $S^n = \{x \in \mathbb{R}^{n+1} : \|x\| = r\}$ where the radius r is usually set to one (giving the unit n -sphere). A similar definition exists in case the base space is complex: a complex n -sphere is defined as $S^n = \{z \in \mathbb{C}^{n+1} : \|z\| = r\}$. The notation S^n refers to the dimension of the surface of the sphere. The n -sphere can be described as $S^n = \mathbb{R}^n \cup \{\infty\}$, which is n -dimensional Euclidean space plus a single point

representing infinity in all directions (this representation gives origin to the real projective space). Alternatively, if a single point is removed from an n -sphere, it becomes homeomorphic to \mathbb{R}^n .

A.2 Groups and transformations

Definition 6. A *group* is a set G with a binary operation $*$ (sometimes called “multiplication”) such that the operation: a) is associative, b) has an identity, and c) has an inverse operation. If in addition, d) $*$ is commutative, then the group is said to be *Abelian*, otherwise it is *non-Abelian*.

Example 2. An instance of an Abelian group is the integers with addition as the $*$ operation.

In Geometry, a *transformation* is a one-to-one correspondence $P \rightarrow P'$ among all the points in the plane (or space), i.e., a rule for associating pairs of points, where each pair has a first point belonging to P and a second point belonging to P' . The most trivial transformation is the *identity* transformation, which leaves each point unchanged. A set of transformations is said to form a group if it contains the inverse of each and the product of any two. For instance, the *symmetry operations*, which leave a figure unchanged while permuting its parts, forms a group, the so-called *symmetry group* (or group of symmetries) of the figure.

Example 3. An instance of a non-Abelian group of transformations is $SO(n)$, the *special orthogonal group* (also called rotation group), which consists of all n -dimensional rotation matrices (orthogonal matrices with determinant equal to one) under the “composition of rotations” operation. Performing a rotation defined by matrix R_1 in a given direction followed by a second one R_2 and a third one R_3 satisfies the associative condition since $(R_1 * R_2) * R_3 = (R_1 * R_2) * R_3$, we clearly have an inverse rotation for every rotation: $R_1^{-1} * R_1 (= I)$ leaves an object in its original position, and we have an identity matrix I which is the zero rotation $R_0 = I$, with $R_0 * R_1 = R_1$. However, rotations do not commute: $R_1 * R_2 \neq R_2 * R_1$

as a 3-dimensional example can demonstrate.

$SO(n)$ can be understood as the group of symmetries of a n -sphere S^n excluding reflections. Note this is a *continuous group*, in contrast to Example 2, where the group is clearly discrete. Continuous groups are called *Lie groups*, after Sofus Lie (1842-1899). Thus, for instance, if $n = 3$, $SO(3)$ is the set of all possible rotations of a 3-dimensional sphere. $SO(n)$ is a *subgroup* of $O(n)$, the orthogonal group. The elements of this set are all $n \times n$ orthogonal matrices, not only those that have determinant one. Hence, $SO(n)$ is a subgroup of $O(n)$. The orthogonal group contains not only the *non-reflective* symmetries $SO(n)$ but also the reflective ones. Both $SO(n)$ and $O(n)$ are in turn subgroups of $GL(n)$, the general linear space of all non-singular $n \times n$ matrices.

Quotient spaces (see Def. 3) can be defined by the *action* of a group on the elements of some manifold \mathcal{M} . If G is a group, then we define two points x, y in \mathcal{M} to be equivalent if there is a $g \in G$ such that $y = gx$ and this defines the quotient space \mathcal{M}/G . In this case, the left action of G on elements of \mathcal{M} define the equivalence relation and hence, the quotient space.

A.3 Projective geometry and Complex projective space

Contrary to the transformations in Euclidean space, Projective geometry deals with transformations that do *not* preserve angles and lengths, namely, projections. In addition, and as it was first known during the Renaissance with perspective painting, there exist points at infinity (“vanishing points”) where parallels meet. Thus, projective geometry allows infinity to be put on the same footing as the finite points of the plane (Stillwell, 2000). A natural question, first raised by Alberti during the Renaissance in his study on perspective, is this: if projections do not preserve angles and lengths, what is preserved? What is preserved under projections is the *cross ratio* of four points A,B,C,D on a line, defined by $\frac{CA}{CB} \cdot \frac{DA}{DB}$.

Homogeneous coordinates (invented by Möbius) give a natural extension of the Cartesian plane \mathbb{R}^2 by assigning new coordinates to the points already present and creating new points including points at infinity. They are the coordinates used in projective geometry.

Definition 7. The homogeneous coordinates of a point $(X, Y) \in \mathbb{R}^2$ are all the real triplets (Xz, Yz, z) with $z \neq 0$, i.e., all real triplets (x, y, z) with $x/z = X$, and $y/z = Y$. (Stillwell, 2000, page 134).

If we take X, Y to be the x, y coordinates in the plane $z = 1$, then the coordinates (Xz, Yz, z) are just the coordinates of points on the line in \mathbb{R}^3 from the origin to (X, Y) . Thus, homogeneous coordinates give a one-to-one correspondence between points $(X, Y) \in \mathbb{R}^2$ and nonhorizontal lines through the origin in \mathbb{R}^3 . The horizontal lines, those with coordinates $(x, y, 0)$, correspond to the points at infinity. In geometrical terms, we have enlarged the \mathbb{R}^2 Euclidean space to the *Real Projective Space* \mathbb{RP}^2 by “adding a point” to \mathbb{R}^2 to represent infinity.

One can consider either real projective spaces $\mathbb{P}^n = \mathbb{RP}^n$ or complex ones ($\mathbb{P}^n = \mathbb{CP}^n$).

Example 4. One example of a complex projective space is the so-called *Riemann sphere* (also called sometimes the Gauss sphere), which is \mathbb{CP}^1 . The Riemann sphere arises as the space of ratios of complex numbers (w, z) , not both zero, which is the space of complex lines through the origin in \mathbb{C}^2 . The Riemann sphere can be thought as a one-to-one correspondence established between the points on a sphere sitting on \mathbb{C} and the points in \mathbb{C} , obtained by *stereographic projection* of the plane into the sphere. This is achieved by drawing lines from the “north pole” N of the sphere into the plane \mathbb{C} below. Any such nonhorizontal line pierces the sphere and touches it in one point, which is then projected into the complex plane into a single point. As the line becomes more horizontal, the point on the sphere is closer to N and the point on the plane is farther away in \mathbb{C} , with a horizontal line at N not touching \mathbb{C} and corresponding to infinity on the plane. We thus have

enlarged \mathbb{C} to $\mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$. The projective completion of \mathbb{C} , \mathbb{CP}^1 , is therefore topologically (i.e., quantitatively) equivalent to a sphere.

More generally, any projective space can be assigned homogeneous coordinates like those illustrated in Definition 7 for the case \mathbb{P}^n . These are the n independent *ratios* of the coordinates z^0, z^1, \dots, z^n for the $n + 1$ -dimensional space from which \mathbb{P}^n arises:

$$z^0 : z^1 : z^2 : \dots : z^n$$

(where the z 's are not all zero) rather than the values of the individual z 's themselves. If the z 's are all real, then these coordinates describe \mathbb{RP}^n ; if they are all complex then they describe \mathbb{CP}^n . Kendall's coordinates (equation 3.1, chapter 3) are an instance of this for the case $n = 3$.

A.3.1 Manifolds, tangent space, submersions and immersions

Informally, a *manifold* is a space that can be thought as “curved” in various ways, but where, locally, (i.e., in the vicinity of each of its points) it can be approximated by ordinary Euclidean space. Manifolds can be thought of as a set of “points” tied together continuously and differentially, so that the points in any sufficient small region can be put into a one-to-one correspondence with an open set of \mathbb{R}^n . This correspondence furnishes a coordinate system for the neighborhood. The ideas of manifolds, their *charts* and *atlases*, were developed by Gauss when working in geodesy and cartography. In the same way that the curvilinear surface of the Earth is approximately represented by planar maps that describe small regions of the globe, which are then “glued” together to form a consistent *Atlas*, similar concepts explain the structure of a general manifold. A formal definition refers to the standard type of manifold, the *Hausdorff space*. A Hausdorff space has the defining property that, for two distinct points on the space, there are open sets containing each which do not intersect.

Example 5. The simplest example of a manifold is an open region in Euclidean space, for instance, that described by sets of solutions of systems of equations in

\mathbb{R}^n . A more interesting example is the space of all $n \times n$ real matrices, $GL(n)$, defined as

$$GL(n) = \{\mathbf{X} \in \mathbb{R}^{n \times n} : \det(\mathbf{X}) \neq 0\} = \det^{-1}(\mathbb{R} - \{0\}).$$

Since the determinant function,

$$\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$$

is continuous, $GL(n)$ is an open subset of $\mathbb{R}^{n \times n}$, and is therefore an n -dimensional manifold. Likewise, a subgroup of $GL(n)$ such as the rotation group $SO(n)$, whose “points” (elements) are the $n \times n$ matrices $\{\mathbf{X} : \mathbf{X}'\mathbf{X} = I, \det(\mathbf{X}) = 1\}$ also constitutes an n -dimensional manifold.

Definition 8. If in the neighborhood of a given point a coordinate system x^1, \dots, x^n is fixed (note the convention of indexing coordinates with superscripts), then at this point there naturally arise n linearly independent *tangent vectors* $e_i = \partial/\partial x^i$ that correspond to differentiations along the coordinate lines passing through the point x . The set of all tangent vectors to a point x in an n -dimensional manifold \mathcal{M} forms a linear space of dimension n . This space is called the *tangent space* to the manifold at x , and is denoted $T_x\mathcal{M}$.

Definition 9. Let f be a function whose domain is a set A . The function f is *injective* if for all a and b in A , if $f(a) = f(b)$, then $a = b$; that is, $f(a) = f(b)$ implies $a = b$. Equivalently, if $a \neq b$, then $f(a) \neq f(b)$. Thus, an injective function preserves distinctness; it never maps distinct elements of its domain to the same element of its codomain. A canonical injective function is the *inclusion function* $i : A \rightarrow B$ defined, for every $x \in A \subset B$, as $i(x) = x \in B$. That is, A is a subset of B and all elements of A are treated as elements of B as well.

Definition 10. A *surjective function* (or *onto function*) is a function whose image is equal to its codomain. Equivalently, a function f with domain X and codomain Y is surjective if for every $y \in Y$ there exists at least one $x \in X$ with $f(x) = y$. A surjective function is called a *surjection*. In a surjective function every point in the codomain is the value of $f(x)$ for at least one point x in the domain.

Recall that in vector calculus the *Jacobian matrix* is a matrix representation of the differential (or total derivative) of a smooth map ϕ at a point $x \in U \subset \mathbb{R}^m$ between subsets $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, i.e.,

$$d\phi_x : \mathbb{R}^m \rightarrow \mathbb{R}^n.$$

This idea can be generalized to the case ϕ is a smooth function between two manifolds \mathcal{M} and \mathcal{N} .

Definition 11. Let $\phi : \mathcal{M} \rightarrow \mathcal{N}$. For some $x \in \mathcal{M}$, the *differential* of ϕ at x is the map

$$\phi : T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$$

from the tangent space of \mathcal{M} at x to the tangent space of \mathcal{N} at $\phi(x)$. See Figure 1.

Definition 12. A smooth map between manifolds $f : \mathcal{M} \rightarrow \mathcal{N}$ is called an *immersion* if the differential $df : T_x\mathcal{M} \rightarrow T_{f(p)}\mathcal{N}$ is injective for every $p \in \mathcal{M}$. If an immersion is homeomorphic to its image it is said to be an *embedding*. The map f is called a *submersion* if df is surjective for every $p \in \mathcal{M}$.

Example 6. The prototype of an immersion is the inclusion of \mathbb{R}^m in a *higher* dimensional \mathbb{R}^n :

$$i(x^1, \dots, x^m) = (x^1, \dots, x^m, 0, 0, \dots, 0).$$

The prototype of a submersion is the projection of \mathbb{R}^m onto a *lower* dimensional \mathbb{R}^n :

$$\pi(x^1, \dots, x^n, x^{n+1}, \dots, x^m) \rightarrow (x^1, \dots, x^n).$$

Example 7. An important submersion in shape analysis is the *Hopf submersion* $S^3 \rightarrow S^2$ where each distinct point of a 2-sphere comes from a distinct circle (a fibre) on the 3-sphere. This can be explained in two different ways:

- Identify \mathbb{R}^4 with \mathbb{C}^2 and \mathbb{R}^3 with $\mathbb{C} \times \mathbb{R}$ by writing

$$(x_1, x_2, x_3, x_4) \text{ as } z_0 = x_1 + ix_2, \text{ and } z_1 = x_3 + ix_4$$

and

$$(x_1, x_2, x_3) \text{ as } z = x_1 + ix_2 \text{ and } x = x_3.$$

Thus,

S^3 is identified with the subset $(z_0, z_1) \in \mathbb{C}^2$ such that $|z_1|^2 + |z_2|^2 = 1$ and

S^2 is identified with the subset $(z, x) \in \mathbb{C} \times \mathbb{R}$ such that $|z|^2 + x^2 = 1$ (note: $|z|^2 = zz^*$). The *Hopf submersion* $p : S^3 \rightarrow S^2$ is then defined as

$$p(z_0, z_1) = (2z_0z_1^*, |z_0|^2 - |z_1|^2)$$

where the first entry on the right hand side is a complex number and the second one is real. Thus, $p(z_0, z_1) \in \mathbb{C} \times \mathbb{R}$, and since $p(z_0, z_1) = 1^1$, it actually lies on $S^2(1)$. Furthermore, since

$$p(z_0, z_1) = p(\lambda z_0, \lambda z_1)$$

for some $\lambda \in \mathbb{C}$ such that $|\lambda|^2 = 1$, then different points in S^3 map to the same point on the 2-sphere. Since $|\lambda|^2 = 1$ forms a circle on \mathbb{C} , it follows that for each point $w \in S^2$, $p^{-1}(w) = S^1$ (a circle) on S^3 . Thus, the 3-sphere is a disjoint union of circular fibres (for this reason this is also called the *Hopf fibration*).

- We can consider the complex projective space $\mathbb{C}\mathbb{P}^1$ as equal to the quotient space of $\mathbb{C}/\{0\}$ by the equivalence relation that identifies (z_0, z_1) with $(\lambda z_0, \lambda z_1)$ for z_0, z_1 and $\lambda (\neq 0) \in \mathbb{C}$ (set of equivalence classes under multiplication by a non-zero complex number). Then, on any complex line in \mathbb{C}^2 (a one dimensional complex subspace that replicates the entire complex space \mathbb{C}) there is a unit circle, so the quotient maps circles to points. Alternatively, (z_0, z_1) can be mapped to the point z_0/z_1 (using homogeneous coordinates) on the Riemann sphere $\mathbb{C} \cup \{\infty\}$.

¹Proof: $2z_0z_1^* \cdot 2z_0z_1^* + (|z_0|^2 - |z_1|^2)^2 = 4|z_0|^2|z_1|^2 + |z_0|^4 - 2|z_0|^2|z_1|^2 + |z_1|^4 = 2|z_0|^2|z_1|^2 + |z_0|^4 + |z_1|^4 = (|z_0|^2 + |z_1|^2)^2 = (1)^2 = 1$.

A.4 Intrinsic and extrinsic geometry and geodesics

The concept of intrinsic geometrical properties of an object originated from the work by Gauss, who, in 1827, conceived the idea of defining the curvature of a surface by measurements that take place entirely on the surface and not based on measurements on the ambient space where the surface is embedded, that is, he found a way to detect the curvature of a surface *intrinsically*. For instance, in the time of Gauss, the curvature of the earth was known on the basis of surveyors and explorers, not by viewing it from space (Stillwell, 2000).

Definition 13. A property of surfaces in \mathbb{R}^3 is called *intrinsic* if it is preserved by *local isometries*. Two surfaces S_1 and S_2 are locally isometric if any sufficiently small portion of S_1 can be mapped isometrically (i.e., preserving arc lengths) into any part of S_2 (thus the map takes any curve on S_1 to a curve in S_2 of equal length). Local isometries between S_1 and S_2 are obtained by a bending transformation that does not include stretching, compressing, or tearing.

Example 8. A plane can be bent into a cylinder, hence they are locally isometric, and hence they have the same Gaussian curvature (zero). A sphere and a plane are not locally isometric (and therefore have different Gaussian curvatures), a fact of great importance in cartography: any planar map of the Earth induces necessarily some distortion of distances.

Definition 14. A *geodesic* is a curve on a surface such that every sufficiently small portion of it is the shortest path on the surface connecting the end-points of the portion. It follows that the geodesic lines of a surface continue to be geodesic if the surface is subject to bending. Hence geodesics are fundamental in the intrinsic properties of a surface. In fact, all intrinsic properties of a surface (e.g., its Gaussian curvature) can be determined by drawing geodesics and measuring its arc lengths.

A.5 Kendall's Preshape and Shape Spaces

Let \mathbf{X} be a $k \times m$ matrix containing the k landmarks (coordinate pairs or triples) of an object in m (2 or 3) dimensions. \mathbf{X} is sometimes called a *configuration matrix* (since it is an element of the configuration space, the space of all possible arrangements of k landmarks in m dimensions). With this notation, the shape of a configuration \mathbf{X} is obtained, first, by removing location and scale effects by computing the so-called *pre-shape* \mathbf{Z} :

$$\mathbf{Z} = \frac{\mathbf{H}\mathbf{X}}{\|\mathbf{H}\mathbf{X}\|} \quad (\text{A.1})$$

where \mathbf{H} is a $(k-1) \times k$ Helmert submatrix (Dryden and Mardia, 1998) and $\|\cdot\|$ denotes the Frobenius norm of a matrix. If we define $h_j = -[j(j+1)]^{-1/2}$, then \mathbf{H} is a matrix whose j th row is: $(\underbrace{h_j, h_j, \dots, h_j}_{j \text{ times}}, \underbrace{0, \dots, 0}_{k-j-1 \text{ times}})$ for $j = 1, \dots, k-1$.

Note that $\mathbf{H}\mathbf{H}' = \mathbf{I}_{k-1}$ and that the rows of \mathbf{H} are contrasts. Alternatively, one could start with the *centered preshapes*, defined by $\mathbf{Z}_c = \mathbf{H}'\mathbf{Z}$ (these are $k \times m$ matrices), although the development below assumes Helmertized preshapes where one of the k coordinates is eliminated.

Transformation (A.1) removes location effects via the numerator, and re-scales the configurations to unit length via the denominator. Since we have not removed rotations from \mathbf{Z} it is not yet the shape of \mathbf{X} , hence the name preshape. The centered preshapes are equivalent to centering each coordinate of each configuration by its centroid and dividing each by its norm.

The *shape* of configuration \mathbf{X} , denoted $[\mathbf{X}]$, is defined as the geometrical information that is invariant to similarity transformations except reflections. In the work by Kendall (1984), reflections are *not* considered, thus, two objects, one the mirror image of the other are considered to have different shapes. Therefore, once location and scale effects are filtered as above, the shape is then defined as:

$$[\mathbf{X}] = \{\mathbf{Z}\mathbf{\Gamma} : \mathbf{\Gamma} \in SO(m)\} \quad (\text{A.2})$$

where \mathbf{Z} is the preshape of \mathbf{X} , $\mathbf{\Gamma}$ is a rotation matrix (i.e., a matrix such that $\mathbf{\Gamma}'\mathbf{\Gamma} = \mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}_m$ with $\det(\mathbf{\Gamma}) = +1$) and $SO(m)$ is the space of all $m \times m$ rotation

matrices that exclude reflections, the special (or non-reflective) orthogonal group. Multiplication by a suitable matrix $\mathbf{\Gamma}$ reorients (rotates) the object. Note that a shape is therefore defined as a set.

The following geometrical interpretation of these transformations is due to Kendall (1984 and 1989). Given that preshapes are scaled and centered objects, they can be represented by vectors from the center to the surface of a unit sphere of dimension $(k - 1)m$, because the numerator in (A.1) removes m degrees of freedom for location parameters and the denominator removes one additional degree of freedom for the change of scale. The preshapes, having unit length, form a space (denoted S_m^k), which has $(k - 1)m - 1$ dimensions by virtue of being on the surface. As one rotates a pre-shape \mathbf{Z} via (A.2), the vectors $\mathbf{Z}\mathbf{\Gamma}$ describe an *orbit* on S_m^k . All the vectors on an orbit correspond to the same shape, since by definition the shape of an object is invariant to rotations. Thus, the orbits (also called *fibers*) of the preshape space are mapped one to one into single points in the *shape space* (denoted Σ_m^k), the space of all possible shapes of k landmarks in m dimensions. This space in general will be a non-Euclidean M -dimensional manifold. Two objects have the same shape if and only if their preshapes lie on the same fiber. The shape space has dimension $M = (k - 1)m - 1 - m(m - 1)/2$ since in addition to losing location and dilation degrees of freedom we also lose $m(m - 1)/2$ degrees of freedom in the specification of the (symmetric) $m \times m$ rotation matrix $\mathbf{\Gamma}$.

Example 9. Preshape space and shape space for lines. In order to explain these ideas, consider one of the simplest possible cases, where we have 2 lines in \mathbb{R}^2 (see Figure A.1). Thus, we have that $m = 2$ and $k = 2$, where the obvious landmarks are the endpoints of the lines. After centering and scaling the two lines using (A.1), one obtains the preshapes with matrices \mathbf{Z}_1 and \mathbf{Z}_2 . Since the original objects evidently have the same shape (that of a line in Euclidean space) these two preshapes lie on the same fiber or orbit, generated as the preshapes are rotated using (A.2). The preshape space S_2^2 is of dimension $(k - 1)m - 1 = 1$, namely, the circumference of a unit circle. As the preshapes rotate (they can rotate clockwise or counterclockwise) they will eventually coincide, which corresponds to the centered and scaled lines coinciding. Finally, since there is a single shape,

the shape space Σ_2^2 is the simplest possible, namely, a single point (dimension is $M = (k - 1)m - 1 - m(m - 1)/2 = 0$, i.e., a 0-manifold).

In general, the shape space Σ_m^k will be a nonlinear space, the Riemannian M -manifold formed by the landmarks modulo similarity transformations, of reduced dimension than the always spherical preshape space. That is, the shape space is defined as a quotient space, i.e., $\Sigma_m^k = \mathbb{R}^{km}/\mathcal{G} = S_m^k/SO(m)$, where \mathcal{G} is the group of similarity transformations that exclude reflections. While the step of going from configuration space (the km -manifold of all possible arrangements of the landmarks) to preshape space is easy to understand, going from preshape space to shape space is a non-trivial step. For instance, for planar shapes Kendall (1984) showed that $\Sigma_2^k = \mathbb{C}P^{k-2}(4)$, the complex projective space of sectional curvature 4 (thus in the previous example, $\Sigma_2^2 = \mathbb{C}P^0(4)$, a one-point space). See Kendall et al. (1999) for a detailed discussion of the geometry of shape spaces.

Example 10. Preshape and shape space for triangles. The map $S_2^3 \rightarrow \Sigma_2^3$ is the *Hopf submersion* of Example 7, which is a map from each non-overlapping circular fibre (the preshapes) to the points in shape space. Each fibre $[x]$ in S_2^3 (point in Σ_2^3) corresponds to a particular triangular shape, the equivalence class generated by the quotient space $S_2^3/SO(3)$.

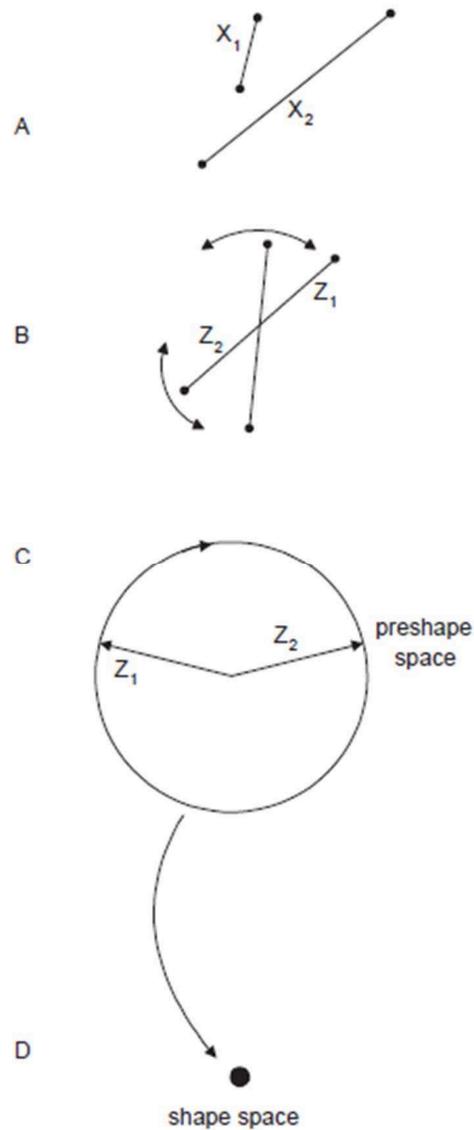


Figure A.1. One of the simplest illustrations of preshape and shape space. A) two lines in the original 2-dimensional space; B), preshapes on 2-dimensional Euclidean space, after centering and scaling; C) the corresponding pre-shape space is the (one-dimensional) circumference of a unit circle. The two pre-shapes lie on the single fiber or orbit generated as the preshapes are rotated, hence there is a single shape; D) the shape space for the two lines (Σ_2^2) is zero dimensional (a single point) and corresponds to the only shape that exists in this example.

The Landmark Matching Problem

The Statistical Shape Analysis (SSA) techniques considered in this dissertation assume corresponding landmark data. It might happen that similar parts measured with a Coordinate Measuring Machine (CMM) do not contain corresponding or labeled landmarks. This can be due to the difficulty of orienting the part when mounting on the CMM. If the orientation is different between parts, the CMM measurements will not correspond to each other. Therefore, a first problem that needs to be addressed is how to “match” the landmarks between two shapes so that we obtain corresponding shape data.

The landmark matching problem has been studied by Computer Vision researchers where it receives the name “point matching problem”. The problem under consideration is as follows: two objects are measured at the same locations, but the labels of each of the measurements differ from one object to the other. The goal is to relabel the landmarks of one of them so they match (correspond) between objects.

The work by Chui and Ranngarajan (2000) is based on solving a highly non-linear optimization problem where the objective is to minimize the sum of the Euclidean distance between the points $\{i\}$ in configuration 1 and the transformed points $\{j\}$ in configuration 2. A transformation is used since the second object may be oriented differently than the first. Jointly determining the matching correspondences and the transformation necessary for registering configuration 2 to configuration 1 results in a hard optimization problem.

A completely different approach is that of Belongie et al. (2002), who proposed

a method for matching 2D shapes. Their method separates the matching problem from the problem of registering the configurations, that is, their matching method is in principle invariant with respect to location, scaling and orientation of the two configurations. The main idea is to measure the amount of data in the neighborhood of each point of each shape using 2D histograms and use a distance metric between histograms that corresponds to points in different objects as costs to be minimized in a classical weighted matching problem, solvable via Linear Programming. The histograms measure the angle and distance between each landmark and all other landmarks in the object. If a similar 2D histogram can be found in another object, this is an indication that the two landmarks should be matched. The results of this method depend on the number of bins used in building the histograms, thus the number of bins should be chosen carefully.

Due to the advances in manufacturing processes, manufactured parts tend to have small variations from each other. Let \mathbf{X} and \mathbf{Y} be two $k \times m$ configuration matrices representing two objects. Assume both \mathbf{X} and \mathbf{Y} have the same mean shape and are located at the origin, i.e., there no need to estimate and filter a translation effect (see section 2.1.2). Following a common model in SSA with no translation, we have

$$\mathbf{X} = \beta_1(\boldsymbol{\mu} + \mathbf{E}_1)\boldsymbol{\Gamma}_1$$

and

$$\mathbf{Y} = \beta_2(\boldsymbol{\mu} + \mathbf{E}_2)\boldsymbol{\Gamma}_2.$$

We then have

$$\begin{aligned} \mathbf{X}\mathbf{X}' &= \beta_1(\boldsymbol{\mu} + \mathbf{E}_1)\boldsymbol{\Gamma}_1 \cdot \boldsymbol{\Gamma}_1'(\boldsymbol{\mu}' + \mathbf{E}_1')\beta_1 \\ &= \beta_1^2(\boldsymbol{\mu}\boldsymbol{\mu}' + 2\boldsymbol{\mu}\mathbf{E}_1' + \mathbf{E}_1\mathbf{E}_1') \end{aligned}$$

since $\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1' = \mathbf{I}$. Similarly, we have

$$\mathbf{Y}\mathbf{Y}' = \beta_2^2(\boldsymbol{\mu}\boldsymbol{\mu}' + 2\boldsymbol{\mu}\mathbf{E}_2' + \mathbf{E}_2\mathbf{E}_2')$$

When the errors variance σ^2 is small, the two matrices $\mathbf{X}\mathbf{X}'$ and $\mathbf{Y}\mathbf{Y}'$ tend to have similar eigenvectors since the matrices $(2\boldsymbol{\mu}\mathbf{E}_1' + \mathbf{E}_1\mathbf{E}_1')$ and $(2\boldsymbol{\mu}\mathbf{E}_2' + \mathbf{E}_2\mathbf{E}_2')$ will be

close element-wise when σ^2 is small (the scale parameters β_1 and β_2 do not matter since eigenvectors are scale invariant). Hence, a simple three step procedure to solve for the landmark matching problem for two planar configurations is to:

1. Align the first object, \mathbf{X} , to the eigenvectors of $\mathbf{X}\mathbf{X}'$.
2. Align the second object, \mathbf{Y} , to the eigenvectors of $\mathbf{Y}\mathbf{Y}'$.
3. Following Belongie et al. (2002), we solve the following optimization problem

$$\min \sum_{i=1}^k \sum_{j=1}^k D_{ij} a_{ij} \quad (\text{B.1})$$

subject to:

$$\begin{aligned} \sum_{j=1}^k a_{ij} &= 1, & i &= 1, 2, \dots, k \\ \sum_{i=1}^k a_{ij} &= 1, & j &= 1, 2, \dots, k \\ a_{ij} &= \{0, 1\} & i &= 1, 2, \dots, k; j = 1, 2, \dots, k \end{aligned}$$

where a_{ij} is a binary variable equal to 1 if point j of object \mathbf{Y} is assigned to point i of object \mathbf{X} and D_{ij} is the Euclidean distance between landmarks i of \mathbf{X} and j of \mathbf{Y} . The rationale is that the aligned objects after applying steps 1 and 2 above result in small distances between corresponding landmarks, and hence there is a high chance that these landmarks will be matched together. This optimization problem is in the form of the famous “set partitioning” problem in the Operations Research literature. The solution of this problem can be easily obtained via Linear Programming (Diaby, 2010).

Example: Suppose we have the two objects shown in Figure B.1. These are two handwritten digit 3’s, each with 13 landmarks. Suppose the landmarks are labeled as shown in the figure. We keep the same labels of the first shape and we try to match the labels of the second shape to those of the first. Running the proposed algorithm for these two objects, we get the “matched” objects shown in Figure B.2 with matched labels shown in Table B.1.

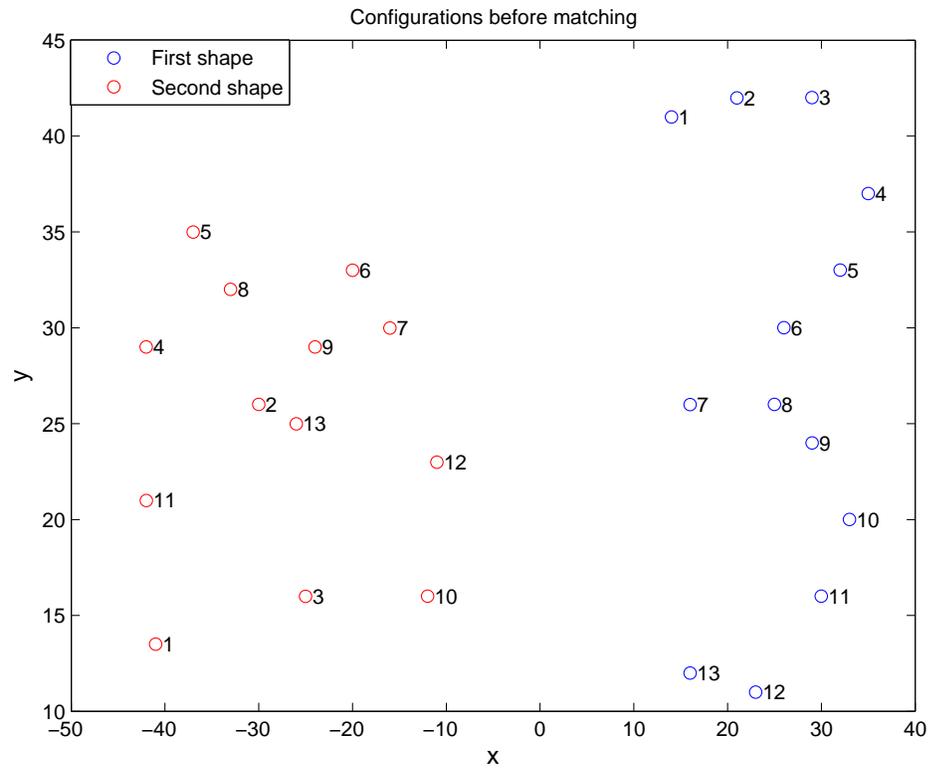


Figure B.1. Two handwritten digit 3's, each with $k = 13$ landmarks. Note that the labels of the landmarks do not correspond between the two shapes.

Table B.1. Matched landmark labels of Figure B.1

Shape 1 landmark Label	Shape 2 landmark Label
1	1
2	11
3	4
4	5
5	8
6	2
7	3
8	13
9	9
10	6
11	7
12	12
13	10

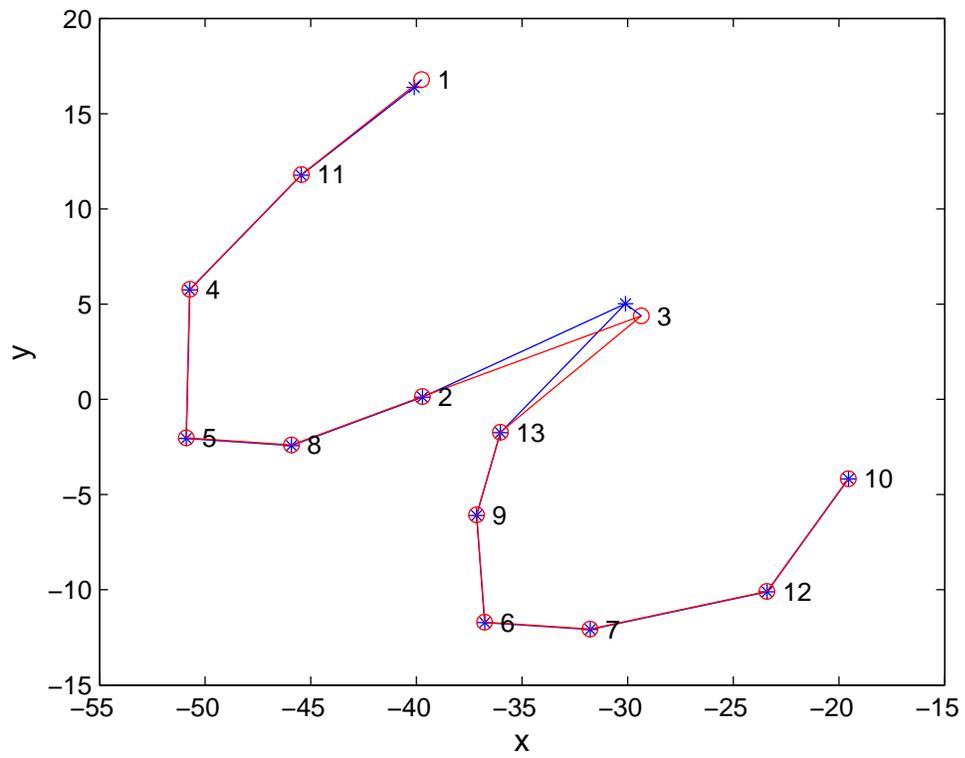


Figure B.2. Matched shapes of Figure B.1 using the proposed matching algorithm. Matched landmark labels are shown in Table B.1.

Kronecker Product Decomposition of Matrices

It was mentioned in chapter 5 that a separable covariance matrix was assumed to speed up computations in spatio-temporal data modeling. This appendix reviews some of the theoretical aspects related to the Kronecker product decomposition of matrices. A characterization of the error incurred when modeling a non-separable covariance matrix with a separable one is provided.

Let

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1J} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{N1} & \sigma_{N1} & \cdots & \sigma_{NJ} \end{bmatrix}$$

and let Σ_{ij} be the ij^{th} block of Σ such that (in MATLAB notation)

$$\Sigma_{ij} = \Sigma((i-1)J+1:iJ, (j-1)J+1:jJ), \quad i, j = 1, 2, \dots, N$$

Define $\mathbf{R}(\boldsymbol{\Sigma})$ of size $N^2 \times J^2$ to be

$$\mathbf{R}(\boldsymbol{\Sigma}) = \begin{pmatrix} \text{vec}(\boldsymbol{\Sigma}'_{11}) \\ \text{vec}(\boldsymbol{\Sigma}'_{12}) \\ \vdots \\ \text{vec}(\boldsymbol{\Sigma}'_{NN}) \end{pmatrix}$$

as defined in Genton (2007). For example for a 6×6 covariance matrix $\boldsymbol{\Sigma}$ with $N = 3$ and $J = 2$ we have

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} & \sigma_{56} \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_{66} \end{pmatrix}$$

and

$$\mathbf{R}(\boldsymbol{\Sigma}) = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{12} & \sigma_{22} \\ \sigma_{31} & \sigma_{41} & \sigma_{32} & \sigma_{42} \\ \sigma_{51} & \sigma_{61} & \sigma_{52} & \sigma_{62} \\ \sigma_{13} & \sigma_{23} & \sigma_{14} & \sigma_{24} \\ \sigma_{33} & \sigma_{43} & \sigma_{34} & \sigma_{44} \\ \sigma_{53} & \sigma_{63} & \sigma_{54} & \sigma_{64} \\ \sigma_{15} & \sigma_{25} & \sigma_{16} & \sigma_{26} \\ \sigma_{35} & \sigma_{45} & \sigma_{36} & \sigma_{46} \\ \sigma_{55} & \sigma_{65} & \sigma_{56} & \sigma_{66} \end{pmatrix}$$

In order to find the nearest decomposition of $\boldsymbol{\Sigma}$ into the Kronecker product of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_s$, we need to solve the following optimization problem:

$$\min_{\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_s} \|\mathbf{R}(\boldsymbol{\Sigma}) - \boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s\|_F$$

$$= \min_{\Sigma_x, \Sigma_s} \|\mathbf{R}(\Sigma) - \text{vec}(\Sigma_x)\text{vec}(\Sigma_s)'\|_F \quad (\text{C.1})$$

where $\|\cdot\|_F$ is the Frobenius norm. This problem has been address form linear algebra point of view and the optimal solution is well defined to be

$$\begin{aligned} \text{vec}(\Sigma_x) &= \sqrt{\delta_1} \mathbf{u}_1 \\ \text{vec}(\Sigma_s) &= \sqrt{\delta_1} \mathbf{v}_1 \end{aligned} \quad (\text{C.2})$$

where δ_1 is the largest eigenvalue of $\mathbf{R}(\Sigma)$ and \mathbf{u}_1 and \mathbf{v}_1 are the vectors that correspond to δ_1 in the SVD decomposition such that $\mathbf{R}(\Sigma) = \mathbf{U}\Delta\mathbf{V}'$. If $\mathbf{R}(\Sigma)$ is not decomposable, there will be an error associated with this approximation. This error is usually reported (Genton, 2007) as the relative error which equals

$$e = \frac{\|\mathbf{R}(\Sigma) - \text{vec}(\Sigma_x)\text{vec}(\Sigma_s)'\|_F}{\|\mathbf{R}(\Sigma)\|_F}. \quad (\text{C.3})$$

Proposition: *The maximum value of the relative error e can have is $\sqrt{1 - 1/r}$, where r is the rank of $\mathbf{R}(\Sigma)$.*

Proof: Assume $\mathbf{R}(\Sigma)$ has rank r , then $\|\mathbf{R}(\Sigma)\|_F = \sqrt{\sum_{i=1}^r \delta_i^2}$ and $\mathbf{R}(\Sigma) = \sum_{i=1}^r \delta_i \mathbf{u}_i \mathbf{v}_i'$ where δ_i 's are the sorted singular values of $\mathbf{R}(\Sigma)$ such that $\delta_1 \geq \delta_2 \geq \dots \geq \delta_r > 0$ and, \mathbf{u}_i and \mathbf{v}_i are the orthonormal vectors of the SVD decomposition as above. Then

$$\begin{aligned} e &= \frac{\left\| \sum_{i=1}^r \delta_i \mathbf{u}_i \mathbf{v}_i' - \sqrt{\delta_1} \mathbf{u}_1 \sqrt{\delta_1} \mathbf{v}_1' \right\|_F}{\sqrt{\sum_{i=1}^r \delta_i^2}} \\ &= \frac{\left\| \sum_{i=2}^r \delta_i \mathbf{u}_i \mathbf{v}_i' \right\|_F}{\sqrt{\sum_{i=1}^r \delta_i^2}} = \frac{\text{trace} \left[\left(\sum_{i=2}^r \delta_i \mathbf{u}_i \mathbf{v}_i' \right)' \left(\sum_{i=2}^r \delta_i \mathbf{u}_i \mathbf{v}_i' \right) \right]}{\sqrt{\sum_{i=1}^r \delta_i^2}} \\ &= \sqrt{\frac{\sum_{i=2}^r \delta_i^2}{\sum_{i=1}^r \delta_i^2}} = \sqrt{1 - \frac{\delta_1^2}{\sum_{i=1}^r \delta_i^2}} \end{aligned}$$

since the \mathbf{u}_i 's are orthonormal vectors. To maximize e , we would need to minimize $\frac{\delta_1^2}{\sum_{i=1}^r \delta_i^2}$ under the constraints that $\sum_{i=1}^r \delta_i = r$ and $\delta_1 \geq \delta_2 \geq \dots \geq \delta_r > 0$. It is obvious that the minimum will be achieved when $\delta_1 = \delta_2 = \dots = \delta_r = 1$, hence, $\frac{\delta_1^2}{\sum_{i=1}^r \delta_i^2} = 1/r$ and the upper bound of e is $\sqrt{1 - 1/r}$.

In order to have $e = 0$, we require that $r = 1$ (since $\sqrt{1 - 1/r} = 0$ if and only if $r = 1$) so we have

$$\begin{aligned} \mathbf{R}(\boldsymbol{\Sigma}) &= \sqrt{\delta_1} \mathbf{u}_1 \sqrt{\delta_1} \mathbf{v}'_1 \\ &= \text{vec}(\boldsymbol{\Sigma}_x) \text{vec}(\boldsymbol{\Sigma}_s)' \\ &= \boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s \end{aligned}$$

A rank 1 matrix $\mathbf{R}(\boldsymbol{\Sigma})$ can be generated if and only if all rows/columns of $\mathbf{R}(\boldsymbol{\Sigma})$ are linearly dependent which can happen only if:

$$\boldsymbol{\Sigma}_{ij} = C_{ij} \boldsymbol{\Sigma}_{11}, \quad i, j = 2, 3, \dots, N$$

where C_{ij} is a constant.

Adaptive MCMC Sampling Algorithm

The posterior distribution for model (5.4-5.8) is:

$$\pi(\phi_s, \psi_x, \phi_x, \kappa, \boldsymbol{\beta} \mid \mathbf{Y}, \mathbf{F}) \propto \pi(\phi_s)\pi(\psi_x)\pi(\phi_x)\pi(\kappa)\pi(\boldsymbol{\beta})|\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_s)^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

Assume the following noninformative priors for model parameters:

$$\pi(\boldsymbol{\beta}) \sim \text{constant} \leftarrow (\text{flat prior}) \tag{D.1}$$

$$\pi(\psi_x) \sim \log N(\mu = 7, \sigma^2 = 1) \tag{D.2}$$

$$\pi(\phi_s) \sim \log N(\mu = 7, \sigma^2 = 1) \tag{D.3}$$

$$\pi(\phi_x) \sim \log N(\mu = 7, \sigma^2 = 1) \tag{D.4}$$

$$\pi(\kappa) \sim \log N(\mu = 7, \sigma^2 = 1) \tag{D.5}$$

and the following proposal distributions

$$q(\psi_x^{new}) \sim \text{truncated } N(\mu = \psi_x^{current}, \sigma_{\psi_x}^2, 0, \text{inf})$$

$$q(\phi_s^{new}) \sim \text{truncated } N(\mu = \phi_s^{current}, \sigma_{\phi_s}^2, 0, \text{inf})$$

$$q(\phi_x^{new}) \sim \text{truncated } N(\mu = \phi_x^{current}, \sigma_{\phi_x}^2, 0, \text{inf})$$

$$q(\kappa^{new}) \sim \text{truncated } N(\mu = \kappa^{current}, \sigma_{\kappa}^2, 0, \text{inf})$$

The following function is used to update the proposal distribution variances, $\{\sigma_{\psi_x}^2, \sigma_{\phi_s}^2, \sigma_{\phi_x}^2, \sigma_{\kappa}^2\}$:

$$\sigma^2 = \begin{cases} \sigma_0^2, & k \leq k_0; \\ s_d \text{var}(s^0, s^1, \dots, s^{k-1}) + s_d \epsilon, & k > k_0. \end{cases} \quad (\text{D.6})$$

where s^k is the k^{th} sample of that parameter and $\sigma_0^2, \kappa_0, s_d$ (the variance multiplier) and ϵ are constants (see Haario et al., 2001).

Let $(\phi_s^k, \psi_x^k, \phi_x^k, \kappa^k, \beta^k)$ be the k^{th} sample from the posterior distribution. To get the $k + 1$ sample follow the steps:

1. Using the k^{th} sample, calculate Σ_x^k and Σ_s^k such that

$$\begin{aligned} \Sigma_s^k &= \exp\{-\mathbf{D}_s/\phi_s^k\} \\ \Sigma_x^k &= \kappa \exp\{-\mathbf{D}_x/\phi_x^k\} + \psi_x^k I \end{aligned}$$

2. **Update ϕ_s :** a. Evaluate the conditional distribution of ϕ_s and call it $f_{\phi_s}^k$ such that

$$\begin{aligned} f_{\phi_s}^k &= \pi(\phi_s^k) |\Sigma_x^k \otimes \Sigma_s^k|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\beta)'(\Sigma_x^k \otimes \Sigma_s^k)^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\beta)\right\} \end{aligned}$$

- b. Update $\sigma_{\phi_s}^2$ using (D.6), propose ϕ_s^{new} from its proposal distribution, and then update Σ_s such that

$$\Sigma_s^{\text{new}} = \exp\{-\mathbf{D}_s/\phi_s^{\text{new}}\}$$

- c. Evaluate $f_{\phi_s}^{\text{new}}$ where

$$\begin{aligned} f_{\phi_s}^{\text{new}} &= \pi(\phi_s^{\text{new}}) |\Sigma_x^k \otimes \Sigma_s^{\text{new}}|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\beta)'(\Sigma_x^k \otimes \Sigma_s^{\text{new}})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\beta)\right\} \end{aligned}$$

- d. If $u \sim \text{uniform}(0,1) \leq \frac{f_{\phi_s}^{\text{new}} q(\phi_s^k, \phi_s^{\text{new}})}{f_{\phi_s}^k q(\phi_s^{\text{new}}, \phi_s^k)}$, then $\phi_s^{k+1} = \phi_s^{\text{new}}$ and $\Sigma_s^{k+1} = \Sigma_s^{\text{new}}$

otherwise $\phi_s^{k+1} = \phi_s^k$ and $\Sigma_s^{k+1} = \Sigma_s^k$.

- 3. Update ψ_x :** a. Evaluate the conditional distribution of ψ_x and call it $f_{\psi_x}^k$ such that

$$f_{\psi_x}^k = \pi(\psi_x^k) |\Sigma_x^k \otimes \Sigma_s^{k+1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\Sigma_x^k \otimes \Sigma_s^{k+1})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

- b. Update $\sigma_{\psi_x}^2$ using (D.6), propose ψ_x^{new} from its proposal distribution, and then update Σ_x such that

$$\Sigma_x^{new} = \kappa^k \exp\{-\mathbf{D}_x/\phi_x^k\} + \psi_x^{new} I$$

- c. Evaluate $f_{\psi_x}^{new}$ where

$$f_{\psi_x}^{new} = \pi(\psi_x^{new}) |\Sigma_x^{new} \otimes \Sigma_s^{k+1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\Sigma_x^{new} \otimes \Sigma_s^{k+1})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

- d. If $u \sim \text{uniform}(0,1) \leq \frac{f_{\psi_x}^{new} q(\psi_x^k, \psi_x^{new})}{f_{\psi_x}^k q(\psi_x^{new}, \psi_x^k)}$, then $\psi_x^{k+1} = \psi_x^{new}$ and $\Sigma_x^k = \Sigma_x^{new}$ otherwise $\psi_x^{k+1} = \psi_x^k$.

- 4. Update ϕ_x :** a. Evaluate the conditional distribution of ϕ_x and call it $f_{\phi_x}^k$ such that

$$f_{\phi_x}^k = \pi(\phi_x^k) |\Sigma_x^k \otimes \Sigma_s^{k+1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\Sigma_x^k \otimes \Sigma_s^{k+1})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

- b. Update $\sigma_{\phi_x}^2$ using (D.6), propose ϕ_x^{new} from its proposal distribution, and then update Σ_x such that

$$\Sigma_x^{new} = \kappa^k \exp\{-\mathbf{D}_x/\phi_x^{new}\} + \psi_x^{k+1} I$$

c. Evaluate $f_{\phi_x}^{new}$ where

$$f_{\phi_x}^{new} = \pi(\phi_x^{new}) |\Sigma_x^{new} \otimes \Sigma_s^{k+1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\Sigma_x^{new} \otimes \Sigma_s^{k+1})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

d. If $u \sim \text{uniform}(0,1) \leq \frac{f_{\phi_x}^{new} q(\phi_x^k, \phi_x^{new})}{f_{\phi_x}^k q(\phi_x^{new}, \phi_x^k)}$, then $\phi_x^{k+1} = \phi_x^{new}$ and $\Sigma_x^k = \Sigma_x^{new}$ otherwise $\phi_x^{k+1} = \phi_x^k$.

5. Update κ : a. Evaluate the conditional distribution of κ and call it f_{κ}^k such that

$$f_{\kappa}^k = \pi(\kappa^k) |\Sigma_x^k \otimes \Sigma_s^{k+1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\Sigma_x^k \otimes \Sigma_s^{k+1})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

b. Update σ_{κ}^2 using (D.6), propose κ^{new} from its proposal distribution, and then update Σ_x such that

$$\Sigma_x^{new} = \kappa^{new} \exp\{-\mathbf{D}_x / \phi_x^{k+1}\} + \psi_x^{k+1} I$$

c. Evaluate f_{κ}^{new} where

$$f_{\kappa}^{new} = \pi(\kappa^{new}) |\Sigma_x^{new} \otimes \Sigma_s^{k+1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})'(\Sigma_x^{new} \otimes \Sigma_s^{k+1})^{-1}(\text{vec}(\mathbf{Y}') - \mathbf{F}\boldsymbol{\beta})\right\}$$

d. If $u \sim \text{uniform}(0,1) \leq \frac{f_{\kappa}^{new} q(\kappa^k, \kappa^{new})}{f_{\kappa}^k q(\kappa^{new}, \kappa^k)}$, then $\kappa^{k+1} = \kappa^{new}$ and $\Sigma_x^{k+1} = \Sigma_x^{new}$ otherwise $\kappa^{k+1} = \kappa^k$ and $\Sigma_x^{k+1} = \Sigma_x^k$.

6. Update $\boldsymbol{\beta}$: Given Σ_x^{k+1} and Σ_s^{k+1} then

$$\Sigma^{k+1} = \Sigma_x^{k+1} \otimes \Sigma_s^{k+1}$$

Sample $\boldsymbol{\beta}^{k+1}$ from $\pi(\boldsymbol{\beta} | \Sigma^{k+1}, \mathbf{Y}, \mathbf{F})$, where the full conditional of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta} | \Sigma^{k+1}, \mathbf{Y}, \mathbf{F}) \sim N\left(\left(\mathbf{F}'(\Sigma^{k+1})^{-1}\mathbf{F}\right)^{-1}\mathbf{F}'(\Sigma^{k+1})^{-1}\text{vec}(\mathbf{Y}'), \left(\mathbf{F}'(\Sigma^{k+1})^{-1}\mathbf{F}\right)^{-1}\right)$$

Computer implementation

E.1 Robust parameter optimization of shape responses

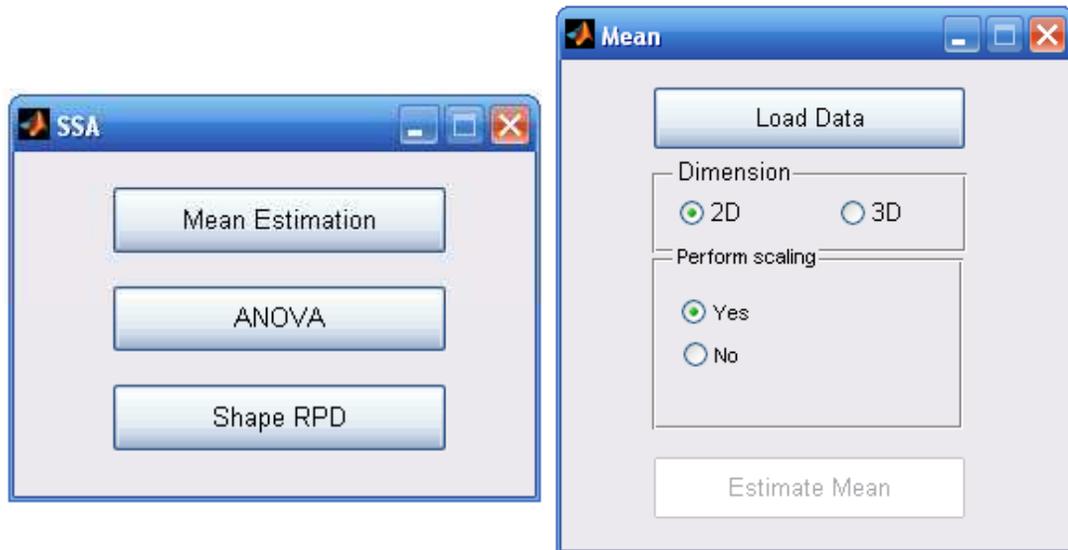
A MATLAB graphical user interface that performs several SSA tasks including robust parameter optimization was programmed. Figure E.1(a) shows the graphical user interface main window. Three main tasks can be performed:

1. *Mean shape estimation*: it uses the GPA and the modified GPA algorithms to estimate the mean shape of a given n objects. Users can choose to perform scaling or not. Mean shape estimation graphical user interface is shown in Figure E.1(b).
2. *Analysis of Variance (ANOVA)*: One-way and two-way ANOVA of shape responses can be performed. This graphical user interface uses previously written MATLAB code provided by Del Castillo and Colosimo (2011). ANOVA graphical user interface is shown in Figure E.1(c).
3. *Geometric shape RPD*: G_{Total} statistic is used in this graphical user interface, shown in Figure E.1(d). Users need to choose the number of noise factors in their experiment. This graphical user interface can also be used when all factors are controllable. All provided examples in chapter 4 have been analyzed by this graphical user interface.

Practitioners can analyze their data by a few clicks only, they do not need to know many statistical details behind SSA.

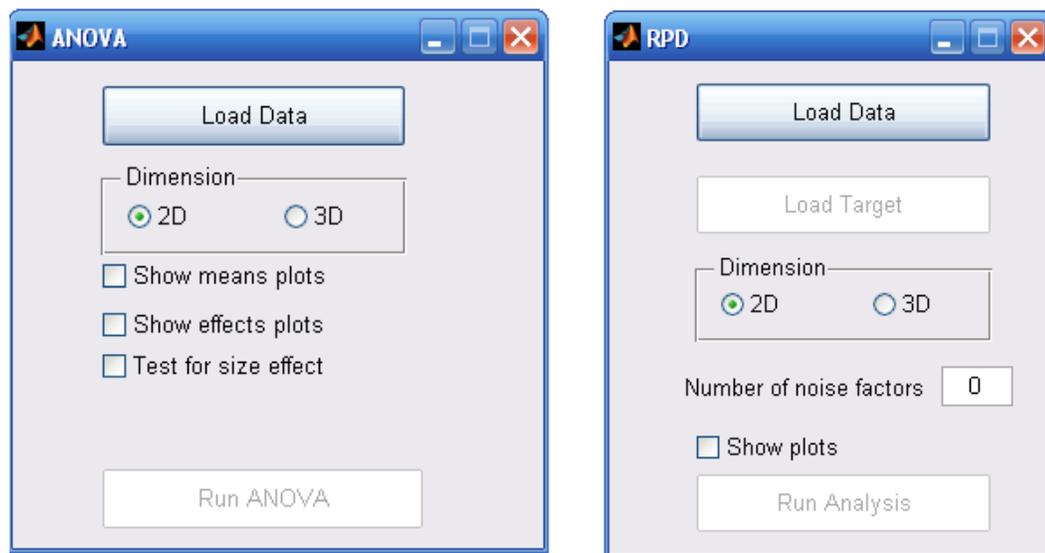
E.2 Robust parameter optimization of profile responses

All computations reported in chapter 5 were implemented in MATLAB (version 2010a). A graphical user interface was developed for model building and optimization. The program requires the Statistics and Global Optimization toolboxes. E.2 shows the interface of the build graphical user interface.



(a)

(b)



(c)

(d)

Figure E.1. (a) The programmed SSA graphical user interface. (b) Mean shape estimation graphical user interface. (c) ANOVA of shape responses graphical user interface. (d) Geometric shape RPD graphical user interface.

The screenshot shows a MATLAB window titled "STRPD" with a light beige background. The window is divided into several sections:

- Workspace data:** This section contains five input fields for variable names: "Profile response" (Y), "Design matrix" (D), "Sampling locations" (S), "Upper target limits" (U), and "Lower target limits" (L).
- Mean structure form:** A dropdown menu is set to "Intercept".
- MCMC and optimizer parameters:** This section contains six input fields: "Number of noise factors" (0), "Number of MCMC samples" (1000), "Number of samples to warmup" (100), "Number of profiles to be used for probability estimation" (100), and "Variance multiplier" (3).
- Action buttons:** A vertical stack of buttons on the right side includes "Load data" (highlighted in blue), "Check for Separability", "Run MCMC", "Plot sample summaries", "Run Cross validation", "Run optimizer", "Save results", and "Start over".

A red text message in the center of the workspace data section reads: "Program is ready, specify variable names as in MATLAB workspace and press 'Load data' to start the analysis."

Figure E.2. Model building and optimization of profile response systems MATLAB interface.

Bibliography

- Adams, D.C., Rohlf, F.J., and Slice, D.E. (2004), “Geometric morphometrics: ten years of progress following the “revolution””, *Italian Journal of Zoology*, 71, 5-16.
- Andrieu, C. and Thoms, J. (2008), “A Tutorial on Adaptive MCMC”, *Statistical Computing*, 18, 343-373.
- Atchad'e, Y. F. and Rosenthal, J. S (2005), “On Adaptive Markov Chain Monte Carlo Algorithms”, *Bernoulli*, 11, 815-828.
- Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, Chapman & Hall/CRC.
- Barcenas C.C. and Griffin, P.M. (2001), “Geometric Tolerance Verification using Superquadrics”. *IIE Transactions*, 33, 1109-1120.
- Bastos, L. S., and O'Hagan, A. (2009), “Diagnostics for Gaussian Process Emulators”, *Technometrics*, 51(4), 425-438.
- Belongie, S., Malik, J. and Puzicha, J. (2002), “Shape Matching and Object Recognition using Shape Contexts”. *IEEE Transactions on Pattern Recognition and Machine Intelligence* 24(24), 509-522.
- Besag, J., and Green, P. J. (1993), “Spatial Statistics and Bayesian Computation”, *Journal of the Royal Statistical Society, Series B (Methodological)*, 55(1), 25-37.
- Bookstein, F.L. (1986), “Size and Shape Spaces for Landmark Data in Two Dimensions”, *Statistical Science*, 1, 181-242.
- Box, G.E.P. (1988), “Signal-to-noise ratios, performance criteria, and transformations”, *Technometrics*, 30, 1-40.

- Box, G.E.P., and Wilson, K.B. (1951), "On the Experimental Attainment of Optimum Conditions", *Journal of the Royal Statistical Society. Series B*, 13(1), 1-45.
- Carlin, B. P. and Louis, T. A. (2008), *Bayesian Methods for Data Analysis*, Boca Raton, Chapman & Hall/CRC.
- Cressie, N. A. (1993), *Statistics for Spatial Data*, New York, Wiley.
- Chib, S., and Carlin, B.P. (1999), "On MCMC Sampling in Hierarchical Longitudinal Models", *Statistics and Computing*, 9, 17-26.
- Chui, H. and Ranngarajan, A. (2000), "A new Algorithm for Non-grid Point Matching". *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2, 44-51.
- Colosimo, B.M., Mammarella, F. and Petro, S. (2007), "Quality Control of Manufactured Surfaces". IX International workshop on Intelligent Statistical Quality Control, Tsinghua University, Beijing, September 12-14.
- Coward, W.M. and McConathy, D. (1996), "A Monte Carlo Study of the Inferential Properties of Three Methods of Shape Comparison". *American Journal of Physical Anthropology*, 99, 369-377.
- Del Castillo, E. (2011), "Mathematical Notions Used in the Theory of Statistical Shape Analysis". Working paper, Engineering Statistics Laboratory, Penn State University.
- Del Castillo, E. (2007), *Process Optimization, A Statistical Approach*, New York, Springer.
- Del Castillo, E., Colosimo, B.M., and Alshraideh, H. (2011), "Bayesian Modeling and Robust Optimization of Functional Responses Affected by Noise Factors". Submitted paper, Engineering Statistics Laboratory, Penn State University.
- Del Castillo, E. and Colosimo, B.M. (2011), "Statistical Shape Analysis of Experiments for Manufacturing Processes". *Technometrics*, 53(1), 1-15.
- Diaby, M. (2010), "Linear Programming Formulation of the Set Partitioning Problem", *International Journal of Operational Research*, 8(4), 399-427.
- Dryden, I.L. and Mardia, K.V. (1998), *Statistical Shape Analysis*, Chichester, John Wiley & Sons.
- Ecker, M.D. and Gelfand, A.E. (1999), "Bayesian Modeling and Inference for Geometrically Anisotropic Spatial Data". *Mathematical Geology*, 31, 67-83.
- Edgington, E.S. (1995) *Randomization Tests*, Marcel Dekker, NY.

- Fang, K.T., Li, R. and Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, Boca Raton, Chapman & Hall/CRC.
- Flegal, J., Haran, M. and Jones, G. L. (2008), "Markov Chain Monte Carlo: Can We Trust the Third Significant Figure?". *Statistical Science*, 23(2), 250-260.
- Furrer, R., Genton, M. G. and Nychka, D. (2006), "Covariance Tapering for Interpolation of Large Spatial Datasets". *Journal of Computational and Graphical Statistics*, 15, 502-523.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004), *Bayesian Data Analysis*. 2nd ed., Chapman and Hall CRC, Boca Raton, FL.
- Genton, M. G. (2007), "Separable Approximations of Space-time Covariance Matrices", *Environmetrics*, 18, 681-695.
- Gneiting, T. (2002), "Nonseparable, Stationary Covariance Functions for Space-Time Data", *Journal of the American Statistical Association*, 97(458), 590-600.
- Goodall, C. (1991), "Procrustes Methods in the Statistical Analysis of Shape". *J. of the Royal Statistical Society, B*, 53(2), 285-339.
- Gower, J.C. (1975), "Generalized Procrustes Analysis". *Psychometrika*, 40(1), 33-51.
- Govaerts, B., and Noel, J. (2005), "Analysing the Results of a Designed Experiment when the Response is a Curve: Methodology and Application in Metal Injection Moulding", *Quality and Reliability Engineering International*, 21, 509-520.
- Gupta, S., Kulahci, M., Montgomery, D.C. and Borrer, C.M. (2010), "Analysis of SignalResponse Systems Using Generalized Linear Mixed Models". *Quality and Reliability Engineering International*, 26, 375-385.
- Haario, H., Saksman, E. and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm", *Bernoulli*, 7(2), 223-242.
- Hastie, T., Tibshirani, R. and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York.
- Henderson, H.V., and Searle, S.R. (1979), "Vec and Vech Operators for Matrices, With Some Uses in Jacobians and Multivariate Statistics", *The Canadian Journal of Statistics*, 7(1), 65-81.
- Henzold, G. (2006), *Geometrical Dimensioning and Tolerancing for Design, Manufacturing, and Inspection*, 2nd ed., Butterworth-Heinemann, Oxford UK.

- Horn, B.K.P., Hilden, H.M., and Negahdaripour, S. (1988), "Closed-form Solution of Absolute Orientation using Orthonormal Matrices", *Journal of the Optical Society of America A*, 5, 1127-1135.
- Huckemann, S., Hotz, T., and Munk, A. (2010a), "Intrinsic MANOVA for Riemannian Manifolds With an Application to Kendalls Space of Planar Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4), 593-603.
- Huckemann, S., Hotz, T., and Munk, A. (2010b), "Intrinsic Shape Analysis: Geodesic PCA for Riemannian Manifolds Modulo Isometric Lie Group Actions", *Statistica Sinica* preprint SS-07-173.
- Jackson, J.E. (2003), *A Users Guide to Principal Components*, New York, John Wiley and Sons.
- Johnson, R.A. and Wichern, D.W. (1998), *Applied Multivariate Statistical Analysis*, 4th ed., Prentice Hall, New Jersey.
- Jones, D.R., Schonlau, M. and Welch, W.J. (1998), "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization*, 13(4), 455-492.
- Jung, B.C., Jhun, M. and Song, S.H. (2006), "A New Random Permutation Test in ANOVA Models". *Statistical Papers*, 48, 47-62.
- Kang, L., and Albin, S. L. (2000), "On-line Monitoring When the Process Yields a Linear Profile". *Journal of Quality Technology* 32, 418-426.
- Kendall, D.G. (1984), "Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces". *Bulletin of the London Mathematical Society*, 16, 81-121.
- Kendall, D.G. (1989), "A Survey of the Statistical Theory of Shape", *Statistical Science*, 4(2), 87-89.
- Kendall, D.G., Barden, D., Carne, T.K., and Le, H., (1999), *Shape and shape theory*. New York, NY: John Wiley & sons.
- Kendall, W.S., and Molchanov, I., (eds.), (2010). *New Perspectives in Stochastic Geometry*, Oxford University Press, NY (chapter 12).
- Kim, K., Mahmoud, M.A., and Woodall, W.H. (2003), "On the Monitoring of Linear Profiles", *Journal of Quality Technology*, 317-326.
- Klingenberg, C.P. and McIntyre, G.S. (1998), "Geometric Morphometrics of Developmental Instability: Analyzing Patterns of Fluctuating Asymmetry with Procrustes Methods". *Evolution*, 52(5), 1363-1375.

- Krulikowski, A. (1996), *Geometric Dimensioning & Tolerancing*, (A companion to the ASME Y14.5M-1994 dimensioning & tolerancing standard), Effective Training Inc., Westland, MI.
- Laird, N. M. and Ware J. H. (1982), "Random-Effects Models for Longitudinal Data". *Biometrics*, 38, 963-974.
- Lange, N., Carlin, B.P., and Gelfand, A.E. (1992), "Hierarchical Bayes Models for the progression of HIV infection using longitudinal CD4 T-cell numbers". *Journal of the American Statistical Association*, 87, 615-626.
- Lele, S.R. and Richtsmeier, J.T. (1991), "Euclidean Distance Matrix Analysis: A Coordinate-Free Approach for Comparing Biological Shapes Using Landmark Data". *American Journal of Physical Anthropology*, 86, 415-427.
- Lele, S.R. (1993), "Euclidean Distance Matrix Analysis(EDMA): Estimation of Mean Form and Mean Form Difference". *Mathematical Geology*, 25, 573-601.
- Lele, S.R. and Cole, T.M. (1996), "A New Test for Shape Differences when Variance-Covariance Matrices are Unequal". *Journal of Human Evolution*, 31, 193-212.
- Lele, S.R. and Richtsmeier, J.T. (2001), *An invariant approach to statistical analysis of shapes*, Boca Raton, Chapman & Hall/CRC.
- Lele, S.R. and McCulloch, C.E. (2002), "Invariance, Identifiability, and Morphometrics". *J. of the American Statistical Association*, 97(459), 796-806.
- Miller, A. (2002), "Analysis of Parameter Design Experiments for Signal-Response Systems", *Journal of Quality Technology*, 34, 139-151.
- Miller, A., and Wu. C.F.J. (1996), "Parameter Design for Signal-Response systems: A Different Look at Taguchi's Dynamic Parmater Design", *Statistical Science*, 11, 2, 122-136.
- Miro, G., Del Castillo, E., and Peterson, J.J. (2004), "Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables", *Journal of Applied Statistics*, 31 (3), 251-270.
- Montgomery, D.C. (2009), *Design and Analysis of Experiments*, 7th ed., John Wiley & Sons.
- Myers, R.H., and Montgomery, D.C. (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons.

- Nair, V.N., Taam, W. and Ye, K.Q. (2002), "Analysis of Functional Responses from Robust Design Studies", *Journal of Quality Technology*, 34, 4, 355-370.
- Pan, J. X. and Fang, K.T. (2002), *Growth Curve Models With Statistical Diagnostics*, New York, Springer.
- Peterson, J.P. (2004), "Bayesian Reliability Approach to Multiple Response Surface Optimization", *Journal of Quality Technology*, 36(2), 139-153.
- Qian, P. Z. G., and Wu, C. F. J. (2008), "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments", *Technometrics*, 50(2), 192-204.
- Ramsay, J. O. and Silverman, B.W. (2005), *Functional Data Analysis*. 2nd ed., New York, Springer.
- Rao, R.C. (2000), "A Note on Statistical Analysis of Shape Through Triangulation of Landmarks". *Proceedings of the National Academy of Sciences*, 97(7), 2995-2998.
- Rohlf, F.J. (1999), "Shape Statistics: Procrustes Superimposition and Tangent Spaces". *Journal of Classification*, 16, 197-223.
- Rohlf, F.J. (2000), "Statistical Power Comparisons Among Alternative Morphometric Methods". *American Journal of Physical Anthropology*, 111, 463-479.
- Santner, T., Williams, B., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer Verlag, New York.
- Shen, Q., and Xu, H. (2007), "Diagnostics for Linear Models with Functional Responses", *Technometrics*, 49(1), 26-33.
- Snee, R.D., and Andrews, H.P. (1971), "Statistical Design and Analysis of Shape Studies", *Applied Statistics*, 20(3), 250-258.
- Stillwell, J. (2000), *Mathematics and its History*, Springer, New York.
- Taguchi, G., and Wu, Y. (1980), "Introduction to off-line Quality Control", Technical report, Central Japan Quality Control Association, Nagoya, Japan.
- Taguchi, G. (1987), *System of Experimental Design*, NY: Unipub/Kraus International Publications.
- Ten Berge, J.M.F. (1977), "Orthogonal Procrustes Rotation for Two or More Matrices". *Psychometrika*, 42(2), 267-276.

- Traband, M., Medeiros, D. and Chandra, M. (2004), "A Statistical Approach to Tolerance Evaluation for Circles and Cylinders". *IIE Transactions*, 36(8), 777-785.
- Ware, J. H. (1985), "Linear Models for the Analysis of Longitudinal Studies". *Journal of the American Statistician* 39, 95-101.
- Wikle, C. K., and Berliner, L. M. (2005), "Combining Information Across Spatial Scales", *Technometrics*, 47(1), 80-91.
- Wu, C.F.J and Hamada, M. (2000), *Experiments, Planning, Analysis and Parameter Design Optimization*. John Wiley and Sons, New York, NY.
- Zhang, D., and Lu, G. (2004), "Review of Shape Representation and Description Techniques", *Pattern Recognition*, 37, 1-19.
- Ziezold, H. (1994), "Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane", *Biometrical Journal*, 36(4), 491-510.

Vita

Hussam Alshraideh

Hussam Alshraideh was born in Tobneh, Jordan on September 17, 1981. He graduated from Dair Abi-Saeed high school in 1999, attended Jordan University of Science and Technology, and graduated with a B.S. in Industrial Engineering. Being ranked first of his graduation class, Hussam was awarded a scholarship to pursue his M.S. and Ph.D degrees. In December 2007, Hussam received his Masters degree in Industrial Engineering/Quality Engineering option from Arizona State University where he was supervised by Professor George Runger. In August, 2008 he entered the Department of Industrial and Manufacturing Engineering at The Pennsylvania State University as a Ph.D in Industrial Engineering and Operations Research with a minor in Statistics under the supervision of Professor Enrique Del Castillo, and will graduate in August, 2011. Soon after graduation, he will start working as an assistant professor in the Industrial Engineering Department at Jordan University of Science and Technology, Jordan. His dissertation focusses on the robust parameter optimization of profile and geometric shape response systems.