

MULTIPLE SURROGATES AND ERROR MODELING IN OPTIMIZATION OF LIQUID
ROCKET PROPULSION COMPONENTS

By

TUSHAR GOEL

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2007

© 2007 Tushar Goel

To my parents Sushil and Ramesh, sister Manjari, and brother Arun.

ACKNOWLEDGMENTS

This dissertation could not be completed without enormous help from my teachers, family, and friends. While I feel that words would never be sufficient to adequately reflect their contributions, I give a try. I am incredibly grateful to my advisors Prof. Raphael Haftka and Prof. Wei Shyy for their continuous encouragement, very generous support, patience, and peerless guidance. Both Prof. Shyy and Prof. Haftka provided me numerous opportunities to develop and to hone my research and personal skills. I am amazed by their never-ending enthusiasm and depth of knowledge, and feel extremely fortunate to have been taught by them.

I would like to especially thank my advisory committee members, Prof. Nam-Ho Kim, Prof. Jacob N. Chung, and Prof. André I. Khuri, for their willingness to serve on my committee, for evaluating my dissertation, and for offering constructive criticism that has helped improved this work. I particularly thank Dr Kim, for many discussions during our weekly group meetings and afterwards.

I feel deeply indebted to Prof. Nestor V. Queipo for a very fruitful collaboration. Not only did he significantly contribute to my work but also he was very supportive and helpful during the entire course of my graduate studies. I express my sincere gratitude to Dr Layne T Watson and Dr Daniel J Dorney for their collaboration and help in my research. I thank Dr Siddharth Thakur, who provided a huge assistance with the STREAM code and suggestions related to numerical aspects in my research work. I also thank Prof. Peretz P. Friedmann and Prof. Kwang-Yong Kim for the opportunities to test some of our ideas. I thank my collaborators Dr Raj Vaidyanathan, Ms Yolanda Mack, Dr Melih Papila, Dr Yogen Utturkar, Dr Jiongyang Wu, Mr Abdus Samad, Mr Bryan Glaz, and Dr Li Liu. I learnt a lot from you and I feel sincerely indebted for your help, both personally and academically.

I thank the staff of Mechanical Engineering department, particularly Jan, Pam, and David, for their help with administrative and technical support. I also am thankful to the staff at International Center, library, CIRCA, and ETD for their help with this thesis and other administrative details. I sincerely acknowledge the financial support provided by NASA Constellation University Institute Program (CUIP).

I duly thank my colleagues in the Structural and Multi-disciplinary Optimization Group and Computational Thermo-fluids Laboratory for their assistance and many fruitful discussions about all the worldly issues related to academics and beyond. I am highly obliged to Prof. Kalyanmoy Deb and Prof. Prashant Kumar at IIT Kanpur, who gave me very sage advice at different times in my life. They indeed have played a big role in shaping my career.

I also thank my colleagues Emre, Eric, Pat, Nick, and Amor, who made my stay at the University of Michigan a memorable one. I am grateful to have true friends in Ashish, Jaco, Erdem, Murali, Siva, Ashwin, Girish, Saurabh, Ved, Priyank, Satish, Tandon, Sudhir, Kale, Dragos, Christian, Victor, Ben, and Palani for lending me a shoulder, when I had a bad day and for sharing with me the happy moments. These memories will remain etched for life.

Finally, but not at all the least, I must say that I would never have completed this work, had it not been the unconditional love, appreciation, and understanding of my family. Despite the fact that we missed each other very much, they always motivated me to take one more step forward throughout my life and rejoiced all my achievements. To you, I dedicate this dissertation!

TABLE OF CONTENTS

| | <u>page</u> |
|--|-------------|
| ACKNOWLEDGMENTS | 4 |
| LIST OF TABLES | 12 |
| LIST OF FIGURES | 16 |
| LIST OF ABBREVIATIONS..... | 20 |
| ABSTRACT..... | 29 |
| CHAPTER | |
| 1 INTRODUCTION AND SCOPE | 31 |
| Space Propulsion Systems | 31 |
| Design Requirements of Propulsions Systems | 32 |
| System Identification and Optimization: Case Studies | 33 |
| Sensitivity Evaluation and Model Validation for a Cryogenic Cavitation Model | 34 |
| Shape Optimization of Diffuser Vanes..... | 34 |
| Surrogate Modeling | 35 |
| Issues with Surrogate Modeling | 37 |
| Sampling Strategies | 37 |
| Type of Surrogate Model..... | 38 |
| Estimation of Errors in Surrogate Predictions..... | 38 |
| Scope of Current Research | 39 |
| 2 ELEMENTS OF SURROGATE MODELING | 43 |
| Steps in Surrogate Modeling | 44 |
| Design of Experiments | 45 |
| Numerical Simulations at Selected Locations | 45 |
| Construction of Surrogate Model | 45 |
| Model Validation..... | 45 |
| Mathematical Formulation of Surrogate Modeling Problem..... | 46 |
| Design of Experiments | 48 |
| Factorial Designs | 49 |
| Central Composite Designs | 50 |
| Variance Optimal DOEs for Polynomial Response Surface Approximations | 51 |
| Latin Hypercube Sampling | 51 |
| Orthogonal Arrays | 52 |
| Optimal LHS, OA-based LHS, Optimal OA-based LHS..... | 53 |
| Construction of Surrogate Model | 54 |
| Polynomial Response Surface Approximation | 54 |
| Kriging Modeling | 56 |

| | | |
|---|---|-----|
| | Radial Basis Functions | 58 |
| | Kernel-based Regression | 61 |
| | Model Selection and Validation | 62 |
| | Split Sample | 62 |
| | Cross Validation | 63 |
| | Bootstrapping | 64 |
| 3 | PITFALLS OF USING A SINGLE CRITERION FOR SELECTING EXPERIMENTAL DESIGNS..... | 72 |
| | Introduction..... | 72 |
| | Error Measures for Experimental Designs | 76 |
| | Test Problems and Results..... | 80 |
| | Comparison of Different Experimental Designs | 81 |
| | Space filling characteristics of D-optimal and LHS designs..... | 81 |
| | Tradeoffs among various experimental designs..... | 82 |
| | Extreme example of risks in single criterion based design: Min-max RMS bias CCD | 87 |
| | Strategies to Address Multiple Criteria for Experimental Designs..... | 89 |
| | Combination of model-based D-optimality criterion with geometry based LHS criterion..... | 90 |
| | Multiple experimental designs combined with pointwise error-based filtering..... | 92 |
| | Concluding Remarks | 94 |
| 4 | ENSEMBLE OF SURROGATES..... | 105 |
| | Conceptual Framework..... | 107 |
| | Identification of Region of Large Uncertainty | 107 |
| | Weighted Average Surrogate Model Concept..... | 107 |
| | Non-parametric surrogate filter | 109 |
| | Best PRESS for exclusive assignments..... | 109 |
| | Parametric surrogate filter | 109 |
| | Test Problems, Numerical Procedure, and Prediction Metrics..... | 112 |
| | Test Problems | 112 |
| | Branin-Hoo function | 112 |
| | Camelback function..... | 112 |
| | Goldstein-Price function | 112 |
| | Hartman functions..... | 112 |
| | Radial turbine design for space launch | 113 |
| | Numerical Procedure | 114 |
| | Prediction Metrics | 115 |
| | Correlation coefficient..... | 115 |
| | RMS error..... | 116 |
| | Maximum error | 116 |
| | Results and Discussion | 117 |
| | Identification of Zones of High Uncertainty | 117 |
| | Robust Approximation <i>via</i> Ensemble of Surrogates | 119 |

| | | |
|---|--|-----|
| | Correlations | 120 |
| | RMS errors | 121 |
| | Maximum absolute errors..... | 123 |
| | Studying the role of generalized cross-validation errors..... | 123 |
| | Effect of sampling density..... | 124 |
| | Sensitivity analysis of PWS parameters..... | 126 |
| | Conclusions..... | 126 |
| 5 | ACCURACY OF ERROR ESTIMATES FOR SURROGATE APPROXIMATION OF NOISE-FREE FUNCTIONS | 144 |
| | Introduction..... | 144 |
| | Error Estimation Measures | 145 |
| | Error Measures for Polynomial Response Surface Approximation | 146 |
| | Estimated standard error..... | 148 |
| | Root mean square bias error..... | 148 |
| | Error Measures for Kriging | 149 |
| | Model Independent Error Estimation Models | 151 |
| | Generalized cross-validation error | 151 |
| | Standard deviation of responses | 152 |
| | Ensemble of Error Estimation Measures..... | 154 |
| | Averaging of multiple error measures..... | 154 |
| | Identification of best error measure | 154 |
| | Simultaneous application of multiple error measures | 155 |
| | Global Prediction Metrics..... | 155 |
| | Root mean square error | 155 |
| | Correlation between predicted and actual errors..... | 156 |
| | Maximum absolute error | 157 |
| | Test Problems and Testing Procedure | 157 |
| | Test Problems | 157 |
| | Branin-Hoo function | 158 |
| | Camelback function..... | 158 |
| | Goldstein-Price function | 158 |
| | Hartman functions..... | 158 |
| | Radial turbine design problem | 159 |
| | Cantilever beam design problem | 159 |
| | Testing Procedure..... | 160 |
| | Design of experiments | 160 |
| | Test points | 161 |
| | Surrogate construction..... | 161 |
| | Error estimation..... | 162 |
| | Results: Accuracy of Error Estimates..... | 162 |
| | Global Error Measures | 163 |
| | Pointwise Error Measures..... | 165 |
| | Root mean square errors..... | 165 |
| | Correlations between actual and predicted errors | 168 |
| | Maximum absolute errors..... | 169 |

| | |
|---|-----|
| Ensemble of Multiple Error Estimators | 170 |
| Averaging of Errors | 170 |
| Identification of Suitable Error Estimator for Kriging | 171 |
| Detection of High Error Regions using Multiple Errors Estimators | 173 |
| Conclusions..... | 175 |
| Global Error Estimators..... | 175 |
| Pointwise Error Estimation Models..... | 176 |
| Simultaneous Application of Multiple Error Measures..... | 177 |
| | |
| 6 CRYOGENIC CAVITATION MODEL VALIDATION AND SENSITIVITY EVALUATION | 199 |
| | |
| Introduction..... | 199 |
| Cavitating Flows: Significance and Previous Computational Efforts | 199 |
| Influence of Thermal Environment on Cavitation Modeling | 201 |
| Experimental and Numerical Modeling of Cryogenic Cavitation..... | 203 |
| Surrogate Modeling Framework..... | 204 |
| Scope and Organization..... | 205 |
| Governing Equations and Numerical Approach..... | 206 |
| Transport-based Cavitation Model | 207 |
| Thermodynamic Effects | 208 |
| Speed of Sound Model | 210 |
| Turbulence Model | 211 |
| Numerical Approach | 212 |
| Results and Discussion | 213 |
| Test Geometry, Boundary Conditions, and Performance Indicators..... | 213 |
| Surrogates-based Global Sensitivity Assessment and Calibration | 214 |
| Global Sensitivity Assessment | 215 |
| Surrogate construction..... | 216 |
| Main and interaction effects of different variables | 217 |
| Validation of global sensitivity analysis | 218 |
| Calibration of Cryogenic Cavitation Model | 219 |
| Surrogate modeling of responses | 220 |
| Multi-objective optimization..... | 220 |
| Optimization outcome for hydrogen | 222 |
| Validation of the calibrated cavitation model | 222 |
| Investigation of Thermal Effects and Boundary Conditions | 223 |
| Influence of Thermo-sensitive Material Properties | 223 |
| Impact of Boundary Conditions | 225 |
| Conclusions..... | 226 |
| Influence of Turbulence Modeling on Predictions | 243 |
| | |
| 7 IMPROVING HYDRODYNAMIC PERFORMANCE OF DIFFUSER VIA SHAPE OPTIMIZATION..... | 245 |
| | |
| Introduction..... | 245 |
| Problem Description | 247 |

| | |
|--|-----|
| Vane Shape Definition | 248 |
| Mesh Generation, Boundary Conditions, and Numerical Simulation | 250 |
| Surrogate-Based Design and Optimization | 251 |
| Surrogate Modeling | 251 |
| Global Sensitivity Assessment | 255 |
| Optimization of Diffuser Vane Performance | 257 |
| Design Space Refinement – Dimensionality Reduction | 259 |
| Final Optimization | 260 |
| Analysis of Optimal Diffuser Vane Shape | 260 |
| Flow Structure | 260 |
| Vane Loadings | 261 |
| Empirical Considerations | 261 |
| Summary and Conclusions | 262 |
| | |
| 8 SUMMARY AND FUTURE WORK | 277 |
| | |
| Pitfalls of Using a Single Criterion for Experimental Designs | 277 |
| Summary and Learnings | 277 |
| Future Work | 278 |
| Ensemble of Surrogates | 278 |
| Summary and Learnings | 278 |
| Future Work | 279 |
| Accuracy of Error Estimates for Noise-free Functions | 279 |
| Summary and Learnings | 279 |
| Future Work | 280 |
| System Identification of Cryogenic Cavitation Model | 280 |
| Summary and Learnings | 280 |
| Future Work | 281 |
| Shape Optimization of Diffuser Vanes | 282 |
| Summary and Learnings | 282 |
| Future Work | 282 |
| | |
| APPENDIX | |
| | |
| A THEORETICAL MODELS FOR ESTIMATING POINTWISE BIAS ERRORS | 283 |
| | |
| Data-Independent Error Measures | 285 |
| Data-Independent Bias Error Bounds | 286 |
| Data-Independent RMS Bias Error | 286 |
| Data-Dependent Error Measures | 287 |
| Bias Error Bound Formulation | 288 |
| Root Mean Square Bias Error Formulation | 289 |
| Determining the Distribution of Coefficient Vector β | 289 |
| Analytical Expression for Pointwise Bias Error Bound | 291 |
| Analytical Estimate of Root Mean Square Bias Error | 292 |
| | |
| B APPLICATIONS OF DATA-INDEPENDENT RMS BIAS ERROR MEASURES | 294 |

| | |
|---|------------|
| Construction of Experimental Designs | 294 |
| Why Min-max RMS Bias Designs Place Points near Center for Four-dimensional Space? | 295 |
| Verification of Experimental Designs | 296 |
| Comparison of Experimental Designs | 297 |
| RMS Bias Error Estimates for Trigonometric Example | 298 |
| C GLOBAL SENSITIVITY ANALYSIS | 304 |
| D LACK-OF-FIT TEST WITH NON-REPLICATE DATA FOR POLYNOMIAL RESPONSE SURFACE APPROXIMATION | 307 |
| LIST OF REFERENCES | 310 |
| BIOGRAPHICAL SKETCH | 329 |

LIST OF TABLES

| <u>Table</u> | <u>page</u> |
|---|-------------|
| 2-1. Summary of main characteristics of different DOEs..... | 70 |
| 2-2. Examples of kernel functions and related estimation schemes. | 70 |
| 2-3. Summary of main characteristics of different surrogate models. | 71 |
| 3-1. D-optimal design (25 points, 4-dimensional space) obtained using JMP. | 102 |
| 3-2. LHS designs (25 points, 4-dimensional space) obtained using MATLAB. | 102 |
| 3-3. Comparison of RMS bias CCD, FCCD, D-optimal, and LHS designs for 4-dimensional space (all designs have 25 points). | 102 |
| 3-4. Prediction performance of different 25-point experimental designs in approximation of example functions F_1 , F_2 , and F_3 in four-dimensional spaces. | 103 |
| 3-5. Min-max RMS bias central composite designs for 2-5 dimensional spaces and corresponding design metrics. | 103 |
| 3-6. Mean and coefficient of variation (based on 100 instances) of different error metrics for various experimental designs in four-dimensional space (30 points)..... | 104 |
| 3-7. Reduction in errors by considering multiple experimental designs and picking one experimental design using appropriate criterion (filtering). | 104 |
| 4-1. Parameters used in Hartman function with three variables. | 137 |
| 4-2. Parameters used in Hartman function with six variables. | 137 |
| 4-3. Mean, coefficient of variation (COV), and median of different analytical functions. | 137 |
| 4-4. Range of variables for radial turbine design problem. | 137 |
| 4-5. Numerical setup for the test problems. | 138 |
| 4-6. Median, 1 st , and 3 rd quartile of the maximum standard deviation and actual errors in predictions of different surrogates at the location corresponding to maximum standard deviation over 1000 DOEs for different test problems. | 138 |
| 4-7. Median, 1 st , and 3 rd quartile of the minimum standard deviation and actual errors in predictions of different surrogates at the location corresponding to minimum standard deviation over 1000 DOEs for different test problems. | 139 |

| | |
|--|-----|
| 4-8. Median, 1 st , and 3 rd quartile of the maximum standard deviation and maximum actual errors in predictions of different surrogates over 1000 DOEs for different test problems..... | 139 |
| 4-9. Effect of design of experiment: Number of cases when an individual surrogate model yielded the least PRESS error. | 140 |
| 4-10. Opportunities of improvement <i>via</i> PWS: Number of points when individual surrogates yield errors of opposite signs. | 140 |
| 4-11. Mean and coefficient of variation of correlation coefficient between actual and predicted response for different surrogate models..... | 140 |
| 4-12. Mean and coefficient of variation of RMS errors in design space for different surrogate models..... | 141 |
| 4-13. Mean and coefficient of variation of maximum absolute error in design space..... | 141 |
| 4-14. Mean and coefficient of variation of the ratio of RMS error and PRESS over 1000 DOEs..... | 142 |
| 4-15. The impact of sampling density in approximation of Branin-Hoo function. | 142 |
| 4-16. The impact of sampling density in approximation of Camelback function. | 143 |
| 4-17. Effect of parameters in parametric surrogate filter used for PWS. | 143 |
| 5-1. Summary of different error measures used in this study. | 187 |
| 5-2. Parameters used in Hartman function with three variables. | 187 |
| 5-3. Parameters used in Hartman function with six variables. | 187 |
| 5-4. Range of variables for radial turbine design problem. | 188 |
| 5-5. Ranges of variables for cantilever beam design problem. | 188 |
| 5-6. Numerical setup for different test problems. | 188 |
| 5-7. Mean and coefficient of variation of normalized actual RMS error in the entire design space. | 189 |
| 5-8. Mean and coefficient of variation of ratio of global error measures and corresponding actual RMS error in design space. | 190 |
| 5-9. Mean and COV of ratio of root mean squared predicted and actual errors for different test problems. | 191 |

| | |
|--|-----|
| 5-10. Mean and COV of correlations between actual and predicted errors for different test problems. | 192 |
| 5-11. Mean and COV of ratio of maximum predicted and actual errors for different test problems..... | 193 |
| 5-12. Mean and COV of ratio of root mean square average error and actual RMS errors for different test problems. | 194 |
| 5-13. Comparison of performance of individual error measures and GCV chosen error measure for kriging. | 194 |
| 5-14. Number of cases out of 1000 for which error estimators failed to detect high error regions. | 195 |
| 5-15. Number of cases out of 1000 for which error estimators failed to detect maximum error regions. | 196 |
| 5-16. Number of cases out of 1000 for which different error estimators wrongly marked low error regions as high error. | 197 |
| 5-17. Number of cases out of 1000 for which different error estimators wrongly marked low error regions as the maximum error region. | 198 |
| 5-18. High level summary of performance of different pointwise error estimators. | 198 |
| 6-1. Summary of a few relevant numerical studies on cryogenic cavitation. | 239 |
| 6-2. Ranges of variables for global sensitivity analyses. | 240 |
| 6-3. Performance indicators and corresponding weights in surrogate approximations of prediction metrics P_{diff} and T_{diff} | 240 |
| 6-4. Performance indicators and corresponding weights in surrogate approximations of prediction metrics P_{diff} and T_{diff} in model-parameter space. | 241 |
| 6-5. Predicted and actual P_{diff} and T_{diff} at best-compromise model parameter for liquid N2 (Case 290C). | 241 |
| 6-6. Description of flow cases chosen for the validation of the calibrated cryogenic cavitation model..... | 242 |
| 6-7. Model parameters in Launder-Spalding and non-equilibrium $k-\varepsilon$ turbulence models. | 243 |
| 7-1. Design variables and corresponding ranges. | 273 |
| 7-2. Summary of pressure ratio on data points and performance metrics for different surrogate models fitted to Set A. | 273 |

| | |
|--|-----|
| 7-3. Range of data, quality indicators for different surrogate models, and weights associated with the components of PWS for Set B and Set C data, respectively..... | 274 |
| 7-4. Optimal design variables and pressure ratio obtained using different surrogates constructed using Set C data. | 274 |
| 7-5. Comparison of actual and predicted pressure ratio of optimal designs obtained from multiple surrogate models (Set C). | 275 |
| 7-6. Modified ranges of design variables and fixed parameters in refined design space. | 275 |
| 7-7. Range of data, summary of performance indicators, and weighted associated with different surrogate models in the refined design space. | 276 |
| 7-8. Design variables and pressure ratio at the optimal designs predicted by different surrogates. | 276 |
| 7-9. Actual and empirical ratios of gaps between adjacent diffuser vanes. | 276 |
| B-1. Design variables and maximum RMS bias errors for min-max RMS bias central composite designs in $N_v=2-5$ dimensional spaces. | 303 |
| B-2. Comparison of different experimental designs for two dimensions. | 303 |
| B-3. Comparison of actual and predicted RMS bias errors for min-max RMS bias central composite experimental designs in four-dimensional space. | 303 |

LIST OF FIGURES

| <u>Figure</u> | <u>page</u> |
|---|-------------|
| 1-1. Schematic of liquid fuel rocket propulsion system. | 41 |
| 1-2. Classification of propulsion systems according to power cycles. | 42 |
| 2-1. Key stages of the surrogate-based modeling approach. | 66 |
| 2-2. Anatomy of surrogate modeling: model estimation + model appraisal. | 66 |
| 2-3. A surrogate modeling scheme provides the expected value of the prediction and the uncertainty associated with that prediction. | 67 |
| 2-4. Alternative loss functions for the construction of surrogate models. | 67 |
| 2-5. A two-level full factorial design of experiment for three variables. | 68 |
| 2-6. A central composite design for three-dimensional design space. | 68 |
| 2-7. A representative Latin hypercube sampling design with $N_s = 6, N_v = 2$ for uniformly distributed variables in the unit square. | 69 |
| 2-8. LHS designs with significant differences in terms of uniformity. | 69 |
| 3-1. Boxplots of radius of the largest unoccupied sphere inside the design space $[-1, 1]^{N_v}$ | 97 |
| 3-2. Illustration of the largest spherical empty space inside the 3D design space $[-1, 1]^3$ (20 points). | 97 |
| 3-3. Tradeoffs between different error metrics. | 98 |
| 3-4. Comparison of 100 D-optimal, LHS, and combination (D-optimality + LHS) experimental designs in four-dimensional space (30 points) using different metrics. | 99 |
| 3-5. Simultaneous use of multiple experimental designs concept, where one out of three experimental designs is selected using appropriate criterion (filtering). | 100 |
| 4-1. Boxplots of weights for 1000 DOE instances (Camelback function). | 128 |
| 4-2. Contour plots of two variable test functions. | 129 |
| 4-3. Boxplots of function values of different analytical functions. | 130 |
| 4-4. Contour plots of errors and standard deviation of predictions considering PRS, KRG, and RBNN surrogate models for Branin-Hoo function. | 130 |

| | |
|--|-----|
| 4-5. Standard deviation of responses and actual errors in prediction of different surrogates at corresponding locations (boxplots of 1000 DOEs using Branin-Hoo function)..... | 131 |
| 4-6. Correlations between actual and predicted response for different test problems. | 132 |
| 4-7. Normal distribution approximation of the sample mean correlation coefficient data obtained using 1000 bootstrap samples (kriging, Branin-Hoo function)..... | 133 |
| 4-8. RMS errors in design space for different surrogate models. | 134 |
| 4-9. Maximum absolute error in design space for different surrogate models. | 135 |
| 4-10. Boxplots of ratio of RMS error and PRESS over 1000 DOEs for different problems. | 136 |
| 5-1. Contour plots of two variable analytical functions. | 178 |
| 5-2. Cantilever beam subjected to horizontal and vertical random loads. | 178 |
| 5-3. Ratio of global error measures and relevant actual RMS error. | 180 |
| 5-4. Ratio of root mean square values of pointwise predicted and actual errors for different problems, as denoted by predicted error measure. | 182 |
| 5-5. Correlation between actual and predicted error measures for different problems. | 184 |
| 5-6. Ratio of maximum predicted and actual absolute errors in design space for different problems. | 186 |
| 6-1. Variation of physical properties for liquid nitrogen and liquid hydrogen with temperature. | 228 |
| 6-2. Experimental setup and computational geometries. | 229 |
| 6-3. Sensitivity indices of main effects using multiple surrogates of prediction metric (liquid N2, Case 290C). | 230 |
| 6-4. Influence of different variables on performance metrics quantified using sensitivity indices of main and total effects (liquid N2, Case 290C). | 231 |
| 6-5. Validation of global sensitivity analysis results for main effects of different variables (liquid N2, Case 290C). | 232 |
| 6-6. Surface pressure and temperature predictions using the model parameters for liquid N2 that minimized P_{diff} and T_{diff} , respectively (Case 290C). | 232 |
| 6-7. Location of points (C_{dest}) and corresponding responses used for calibration of the cryogenic cavitation model (liquid N2, Case 290C). | 233 |

| | |
|---|-----|
| 6-8. Pareto optimal front and corresponding optimal points for liquid N2 (Case 290C) using different surrogates. | 233 |
| 6-9. Surface pressure and temperature predictions on benchmark test cases using the model parameters corresponding to original and best-compromise values for different fluids. | 234 |
| 6-10. Surface pressure and temperature predictions using the original parameters and best-compromise parameters for a variety of geometries and operating conditions. | 236 |
| 6-11. Surface pressure and temperature profile on 2D hydrofoil for Case 290C where the cavitation is controlled by, (1) temperature-dependent vapor pressure, and (2) zero latent heat, and hence isothermal flow field. | 237 |
| 6-12. Impact of different boundary conditions on surface pressure and temperature profile on 2D hydrofoil (Case 290C, liquid N2) and predictions on first computational point next to boundary. | 238 |
| 6-13. Influence of turbulence modeling on surface pressure and temperature predictions in cryogenic cavitating conditions. | 244 |
| 7-1. A representative expander cycle used in the upper stage engine. | 264 |
| 7-2. Schematic of a pump. | 264 |
| 7-3. Meanline pump flow path. | 264 |
| 7-4. Baseline diffuser vane shape and time-averaged flow. | 265 |
| 7-5. Definition of the geometry of the diffuser vane. | 265 |
| 7-6. Parametric Bezier curve..... | 265 |
| 7-7. A combination of H- and O-grids to analyze diffuser vane. | 266 |
| 7-8. Surrogate based design and optimization procedure. | 266 |
| 7-9. Surrogate modeling..... | 267 |
| 7-10. Sensitivity indices of main effect using various surrogate models. | 267 |
| 7-11. Sensitivity indices of main and total effects of different variables using PWS. | 268 |
| 7-12. Actual partial variance of different design variables. | 268 |
| 7-13. Baseline and optimal diffuser vane shape obtained using different surrogate models. | 269 |
| 7-14. Comparison of instantaneous and time-averaged flow fields of intermediate optimal (PRS) and baseline designs. | 270 |

| | |
|---|-----|
| 7-15. Instantaneous and time-averaged pressure for the final optimal diffuser vane shape. | 271 |
| 7-16. Pressure loadings on different vanes. | 271 |
| 7-17. Gaps between adjacent vanes. | 272 |
| B-1. Two-dimensional illustration of central composite experimental design constructed using two parameters α_1 and α_2 | 301 |
| B-2. Contours of scaled predicted RMS bias error and actual RMS error when assumed true model to compute bias error was quintic while the true model was trigonometric. | 301 |
| Figure B-3. Contours of scaled predicted RMS bias error and actual RMS error when different distributions of $\beta^{(2)}$ were specified. | 302 |

LIST OF ABBREVIATIONS

| | |
|--|--|
| A | Alias matrix |
| a_{ij} | Constants in Hartman functions |
| b_i | Estimated coefficients associated with i^{th} basis function |
| \mathbf{b} | Vector of estimated coefficients of basis functions |
| cov | Covariance matrix in kriging |
| \mathbf{c} | Bounds on coefficient vectors |
| C_{dest}, C_{prod} | Cavitation model parameters |
| c_i | Constants in Hartman functions |
| C_p | Specific heat at constant pressure |
| $C_{\varepsilon_1}, C_{\varepsilon_2}, \sigma_k, \sigma_\varepsilon, C_\mu, C$ | Turbulence model parameters |
| D_{eff} | D-efficiency |
| $E(x)$ | Expected value of random variable \mathbf{x} |
| E_{avg} | Average of surrogate models |
| E_i | Error associated with i^{th} surrogate model |
| $e(\mathbf{x})$ | Approximation error at design point \mathbf{x} |
| $e_b(\mathbf{x})$ | Bias error at design point \mathbf{x} |
| $e_b^{beb}(\mathbf{x})$ | Bias error bound at design point \mathbf{x} |
| $e_b^I(\mathbf{x})$ | Data-independent bias error bound at design point \mathbf{x} |
| $e_b^{rms}(\mathbf{x})$ | Root mean square bias error at design point \mathbf{x} |

| | |
|---------------------------------|--|
| $e_{es}(\mathbf{x})$ | Standard error at design point \mathbf{x} |
| $f(\mathbf{x}), f, f_i, f_{ij}$ | Function of design variables and decomposed functions |
| $\mathbf{f}(\mathbf{x})$ | Vector of basis functions in polynomial response surface model |
| f_v | Vapor mass fraction |
| h | Sensible enthalpy |
| $h(\mathbf{x})$ | Radial basis function |
| \mathbf{h} | Vector of radial basis functions |
| I | Identity matrix |
| K | Thermal conductivity |
| k | Turbulent kinetic energy |
| L | Loss function, Latent heat |
| M | Moment matrix |
| \dot{m}^+, \dot{m}^- | Cavitation source terms |
| N_{DOE} | Number of design of experiments |
| N_e | Number of eigenvectors |
| N_{lhs} | Number of Latin hypercube samples |
| N_{RBF} | Number of radial basis functions |
| N_q | Number of symbols for orthogonal arrays |
| N_s | Number of sampled data points |
| N_{SM} | Number of surrogate models |
| N_{test} | Number of test points |

| | |
|-----------------------------|--|
| N_v | Number of variables |
| N_{β_1} | Number of basis functions in approximation model |
| N_{β_2} | Number of basis functions missing from the approximation model |
| P_{diff} | L_2 norm of the difference between predicted and benchmark experimental pressure data |
| P_t | Turbulence production term |
| P_y, P_z | Location of mid-point on the lower side of diffuser vane |
| p | Pressure |
| p_{ij} | Constants in Hartman functions |
| R | Correlation matrix |
| R_{adj}^2 | Adjusted coefficient of multiple determination |
| r | Strength of orthogonal array |
| r_{max} | Radius of largest unoccupied sphere |
| $\mathbf{r}(\mathbf{x})$ | Vector of correlation between prediction and data points (kriging) |
| S_i, S_{ij}, S_i^{total} | Sensitivity indices, main, interaction, and total effects |
| s_{resp} | Standard deviation of responses |
| T | Temperature |
| T_{diff} | L_2 norm of the difference between predicted and benchmark experimental temperature data |
| t | Time, magnitude of tangents of Bezier curves |
| $u, v, w, u_i, u_j, u_k, U$ | Velocity components |

| | |
|--|--|
| \mathbf{V} | Null eigenvector |
| V, V_i, V_{ij} | Variance and its components (partial variances), volume |
| w_i | Weight associated with i^{th} surrogate model |
| X | Gramian design matrix |
| \mathbf{x} | Vector of design variables |
| x_i, x_j, x_k | Space variables (coordinates) |
| \mathbf{y} | Vector of responses |
| $y, y(\mathbf{x})$ | Function or response value at a design point \mathbf{x} |
| $\hat{y}, \hat{y}(\mathbf{x})$ | Predicted response at a design point \mathbf{x} |
| $Z(\mathbf{x})$ | Systematic departure term in kriging |
| Z | Subset of design variables for global sensitivity analysis |
| α | Volume fraction, Thermal diffusivity, Parameter in weighted average model |
| α_1, α_2 | Parameters used to define vertex and axial point locations in a central composite design |
| $\boldsymbol{\beta}$ | Vector of coefficients of true basis functions |
| $\boldsymbol{\beta}^*$ | Estimated coefficient vector in kriging |
| β | Coefficient associated with a basis function in polynomial response surface approximation, coefficient of thermal expansion, parameter in weighted average model |
| γ | Constant used to estimate root mean square bias error |
| δ_{ij} | Kronecker-delta |
| $\varepsilon, \varepsilon(\mathbf{x})$ | Turbulent dissipation, error in surrogate model |
| $\eta(\mathbf{x})$ | True function or response |

| | |
|-------------------|---|
| θ | Probability density function, variable defining diffuser vane shape |
| $\mathbf{\theta}$ | Vector of parameters in the Gaussian correlation function |
| λ | Regularization parameter |
| μ | Mean of responses at sampled points, dynamic viscosity |
| ξ_i | Weights used in numerical integration |
| ρ | Density |
| σ^2 | Variance of noise, estimated process variance in kriging |
| σ_a | Adjusted root mean square error |
| ϕ | Degree of correlation, flow variable |
| $\mathbf{1}$ | Vector of ones |

Subscripts

| | |
|-----------|--|
| c | Cavity, candidate true polynomial |
| $dopt$ | D-optimal design |
| i, j, k | Indices |
| $krig$ | Kriging |
| l | Liquid phase |
| lhs | Latin hypercube sampling design |
| m | Mixture of liquid and vapor |
| max | Maximum of the quantity |
| min | Minimum of the quantity |
| prs | Polynomial response surface |
| pws | PRESS-based weighted average surrogate |

| | |
|---------------|---|
| <i>rbnn</i> | Radial basis neural network |
| <i>RMS</i> | Root mean square value |
| <i>t</i> | Turbulent |
| <i>wta</i> | Weighted average surrogate |
| ε | Solution with least deviation from the data |
| <i>v</i> | Vapor phase |
| ∞ | Reference conditions |

Superscripts/Overhead Symbols

| | |
|---------------------|--|
| <i>I</i> | Data independent error measure |
| <i>R</i> | Reynolds stress |
| <i>(i)</i> | i^{th} design point |
| <i>(-i)</i> | All points except i^{th} design point |
| <i>(l)</i> | Lower bound |
| <i>T</i> | Transpose |
| <i>(u)</i> | Upper bound |
| <i>(1)</i> | Terms in the response surface model |
| <i>(2)</i> | Terms missing from the response surface model |
| $\hat{}$ | Predicted value |
| $\bar{}$ | Average of responses in surrogate models |

Non-dimensional Numbers

| | |
|------------|-------------------------------------|
| <i>CFL</i> | Courant, Freidricks and Levy number |
| <i>Pr</i> | Prandtl number |
| <i>Re</i> | Reynolds number |

σ Cavitation number

Acronyms

| | |
|------|--|
| ADS | All possible data sets |
| AIAA | American Institute of Aeronautics and Astronautics |
| ASME | American Society of Mechanical Engineers |
| ASP | Alkaline-surfactant-polymer |
| CCD | Central composite design |
| CFD | Computational fluid dynamics |
| COV | Coefficient of variation |
| CV | Cross validation |
| DOE | Design of experiment |
| ED | Experimental design |
| EGO | Efficient global optimization |
| ESE | Estimated standard error |
| FCCD | Face-centered central composite design |
| GCV | Generalized cross validation error |
| GMSE | Generalized mean squared error |
| GSA | Global sensitivity analysis |
| KRG | Kriging |
| IGV | Inlet guide vane |
| LHS | Latin hypercube sampling |
| LH2 | Liquid Hydrogen |
| LOX | Liquid Oxygen |

| | |
|------------------|--|
| MSE | Mean squared error |
| NASA | National Aeronautics and Space Administration |
| N-S | Navier-Stokes |
| NIST | National Institute of Standards and Technology |
| NPSF | Non-parametric surrogate filter |
| OA | Orthogonal array |
| POF | Pareto optimal front |
| PRESS | Predicted residual sum of squares |
| PRS | Polynomial response surface |
| PSF | Parametric surrogate filter |
| PWS | PRESS-based weighted average surrogate |
| RBF | Radial basis function |
| RBNN | Radial basis neural network |
| RMS | Root mean square |
| RMSBE | Root mean square bias error |
| RMSE | Root mean squared error |
| RP1 | Refined petroleum |
| SoS | Speed of sound |
| SQP | Sequential quadratic programming |
| SS | Split sample |
| SVR | Support vector regression |
| <u>Operators</u> | |
| $E(a)$ | Expected value of the quantity a |

| | |
|-----------------------------|---|
| $\max(a, b)$ | Maximum of a and b |
| $\min(a, b)$ | Minimum of a and b |
| $r(\mathbf{a}, \mathbf{b})$ | Correlation between vectors \mathbf{a} and \mathbf{b} |
| $V(a)$ | Variance of the quantity a |
| $\sigma(a)$ | Standard deviation of the quantity a |
| $(a)_{\text{avg}}$ | Space-averaged value of the quantity a |
| $(a)_{\text{max}}$ | Maximum of the quantity a |
| $ a $ | Absolute value of the quantity a |
| $\ \mathbf{a}\ $ | (L_2) Norm of the vector \mathbf{a} |

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MULTIPLE SURROGATES AND ERROR MODELING IN OPTIMIZATION OF LIQUID
ROCKET PROPULSION COMPONENTS

By

Tushar Goel

May 2007

Chair: Raphael T. Haftka
Cochair: Wei Shyy
Major: Mechanical Engineering

Design of space propulsion components is extremely complex, expensive, and involves harsh environments. Coupling of computational fluid dynamics (CFD) and surrogate modeling to optimize performance of space propulsion components is becoming popular due to reduction in computational expense. However, there are uncertainties in predictions using this approach, like empiricism in computational models and surrogate model errors. We develop methods to estimate and to reduce such uncertainties.

We demonstrate the need to obtain experimental designs using multiple criteria by showing that using a single-criterion may lead to high errors. We propose using an ensemble of surrogates to reduce uncertainties in selecting the best surrogate and sampling strategy. We also develop an averaging technique for multiple surrogates that protects against poor surrogates and performed at par with best surrogate for many problems.

We assess the accuracy of different error estimation models, including an error estimation model based on multiple surrogates, used to quantify prediction errors. While no single error model performs well for all problems, we show possible advantage of combining multiple error models.

We apply these techniques to two problems relevant to space propulsion systems. First, we employ surrogate-based strategy to understand the role of empirical model parameters and uncertainties in material properties in a cryogenic cavitation model, and to calibrate the model. We also study the influence of thermal effects on predictions in cryogenic environment in detail. Second, we use surrogate models to improve the hydrodynamic performance of a diffuser by optimizing the shape of diffuser vanes. For both problems, we observed improvements using multiple surrogate models.

While we have demonstrated the approach using space propulsion components, the proposed techniques can be applied to any large-scale problem.

CHAPTER 1 INTRODUCTION AND SCOPE

Liquid rocket propulsion systems are the most popular form of space propulsion systems for high thrust and specific impulse applications as required for space applications (Humble et al., 1995, Chapter 5). Unlike propulsion systems used in aircraft, space propulsion systems carry both fuel and oxidizer with the vehicle. This poses additional requirements on the selection of suitable propellants and design of propulsion systems.

Apart from high energy density, the choice of propellants is also affected by the ease of storage and handling, mass or volume of propellant, and nature of products of combustion. Typical propellants used for liquid space propulsion are refined petroleum (RP1) with liquid oxygen (LOX), hypergolic propellants (mono-methyl hydrazine with nitrogen tetroxide) and cryogenics (liquid hydrogen LH2 and LOX). Cryogenics (LH2 and LOX) are most popular due to higher power/gallon ratio and specific thrust, lower weight of LH2, and cleaner combustion products (water). Despite difficulties in storage (due to the tendency of cryogenics to return to gaseous state unless super-cooled, the boiling point of LH2 and LOX at standard conditions is -423°F and -298°F , respectively) and safety considerations, the rewards of using cryogenics as space propellants are significant (NASA online facts, 1991).

Space Propulsion Systems

A conceptual schematic of a typical bi-propellant space propulsion system is shown in Figure 1-1. There are five major components of the propulsion system: fuel and oxidizer storage tanks, fuel and oxidizer pumps, gas turbine, combustion chamber, and nozzle. Based on the type of power cycle, different propulsion systems are classified as follows:

- Gas-Generator Cycle (Figure 1-2(A))

A small amount of fuel and oxidizer is fed to gas generator where the fuel is burnt at less than optimal ratio to keep the temperature in turbine low. The hot gases produced in gas

generator drive the turbine to produce power required to run the fuel and oxidizer pumps. Thrust of the engine is regulated by controlling the amount of propellants through gas generator. This is an open cycle since the hot gas from turbines is either dumped overboard or sent into the main nozzle downstream. This configuration is useful for moderate power requirements but not good for the applications which require high power.

- Staged Combustion Cycle (Figure 1-2(B))

This is a closed cycle system in which an enriched mixture of fuel and oxidizer is generated in pre-burner, and after passing through the turbine this vaporized mixture is fed to the main combustion chamber. No fuel or oxidizer is wasted in this cycle as complete combustion takes place in the main combustion chamber. This cycle is used for high power applications but the engine development cost is high and components of propulsion system are subjected to harsh conditions.

- Expander Cycle (Figure 1-2(C))

In this closed cycle, the main combustion chamber is cooled by passing liquid fuel that in turn gets vaporized by exchange of heat. This fuel vapor runs the turbine and is fed back to the main combustion chamber, where complete combustion takes place. The limitation on heat transfer to the fuel limits the power available in the turbine. This configuration is more suitable to small/mid size engines.

- Pressure-Fed Cycle (Figure 1-2(D))

This is the simplest configuration which does not require any pump or turbine. Instead, the fuel and the oxidizer are fed to the combustion chamber by the tank pressure. This cycle can be applied only to relatively low-chamber-pressure applications because higher pressure requires bulky storage tanks.

Design Requirements of Propulsions Systems

The design of an optimal propulsion system requires efficient performance of all components. As can be seen from Figure 1-2, pumps, turbine, combustion chamber, and nozzle are integral part of almost all space propulsion systems. The system level goal of designing a space propulsion system is to obtain the highest possible thrust with the lowest weight (Griffin and French, 1991), but the requirements of individual components are also governed by corresponding operating conditions. Storage tanks are required to withstand high pressure in cryogenic environments while keeping the weight of tank low. Nozzles impart high velocities to high temperature combustion products to produce maximum possible thrust. The requirements

for the design of the turbine and the combustion chamber are high efficiency, compact design, and ability to withstand high pressures and temperatures. Similarly, pumps are required to supply the propellants to the combustion chamber at a desired flow rate and injection pressure. Major issues in the design of pumps include harsh environments, compact design requirements, and cavitation under cryogenic conditions. Each subsystem has numerous design options for example, number of stages, number and geometry of blades in turbines and pumps, different configurations and geometry of injectors and combustion chambers etc., that relate to the subsystem level requirements.

System Identification and Optimization: Case Studies

The design of a propulsion system is extremely complex due to the often conflicting requirements posed on individual components and interaction of different subsystems. Experimental design of propulsion systems is very expensive, time consuming, and involves harsh environments. With improvements in numerical algorithms and increase in computational power, the role of computational fluid dynamics (CFD) for design of complex systems with various levels of complexities has grown many-fold. Computer simulation of design problems has not only reduced the cost of developing the designs but also has reduced risks and design cycle time. The flexibility of trying alternative design options also has increased extensively compared to the experiments. With improvements in computer hardware and CFD algorithms, the complexity of the simulation model is increasing in an effort to capture physical phenomenon more accurately.

Our current efforts in the design of liquid rocket propulsion systems are focused on two distinct classes of problems. First, we try to understand the roles of thermal effects and model parameters used for numerical modeling of cryogenic cavitation; and second, we aim to carry out shape optimization of diffuser vanes to maximize diffuser efficiency. While our interests in the

design of diffuser are motivated by the ongoing interests in the Moon and Mars exploration endeavors, the study of cryogenic cavitation model validation and sensitivity analysis is relatively more generic, but very relevant to the design of liquid rocket propulsion systems. We discuss each problem in more detail as follows.

Sensitivity Evaluation and Model Validation for a Cryogenic Cavitation Model

Though a lot of improvements have been made over the years in the numerical modeling of fluid physics, areas like cavitation, both in normal as well as cryogenic environment, are still continuously developing. Complex physics, phase changes and resulting variations in material properties, interaction of convection, viscous effects and pressure, time dependence, multiple time scales, interaction between different phases of fluids and many fluids, temperature dependent material properties, and turbulence make these problems difficult (Utturkar, 2005). Numerous algorithms (Singhal et al. 1997, Merkle et al. 1998, Kunz et al. 2000, Senocak and Shyy 2004a-b, Utturkar et al. 2005a-b) have been developed to capture the complex behavior of cavitating flows, which have serious implications on the performance of the propulsion components.

We study one cryogenic cavitation model in detail to assess the role of thermal boundary conditions, thermal environment, uncertainties in material properties, and empirical model parameters on the prediction of pressure and temperature field. Finally, we calibrate the cryogenic cavitation model parameters and validate the outcome using several benchmark data.

Shape Optimization of Diffuser Vanes

The diffuser is a critical component in liquid rocket turbomachinery. The high velocity fluid from turbo-pump is passed through the diffuser to partially convert kinetic energy of the fluid into pressure. The efficiency of a diffuser is determined by its ability to induce pressure recovery that is characterized by the ratio of outlet to inlet pressure. While exploring different

concepts for diffuser design, Dorney et al. (2006a) observed that diffusers with vanes are more effective than vaneless diffusers. Consequently, we seek further improvements in diffuser efficiency through shape optimization of vanes.

Since computational cost of simulations is high, we adopt a surrogate model-based framework for optimization and sensitivity analysis that is briefly introduced in next section and discussed in detail in a following chapter. Together, the present effort demonstrates that the same technical framework can be equally adapted to treat hardware and computational modeling development, with information to allow one to inspect the overall design space characteristics, to facilitate quantitative trade-off considerations between of multiple and competing objectives, and to rank order the importance of the various design variables.

Surrogate Modeling

High computational cost of simulations involved in performance evaluation of propulsion components makes direct coupling of optimization tools and simulations infeasible for most practical problems. To alleviate the problems associated with the optimization of such complex and computationally expensive components, surrogate models based on a limited amount of data have been frequently used. Surrogate models offer a low-cost alternative to evaluate a large number of designs and are amenable to the optimization process. Surrogate models also can be used to assess the trends in the design space as well as help identify the problems associated with numerical simulations.

A sample of recent application of surrogate models for the design of space-propulsion system is given as follows. Lepsch et al. (1995) used polynomial response surface approximations to minimize the empty-weight of the dual fuel vehicles by considering propulsion systems and vehicle design parameters. Rowell et al. (1999) and the references within discuss the application of regression techniques for the design of single-stage-to-orbit vehicles.

Madsen et al. (2000) used polynomial response surface models for the design of diffusers. Gupta et al. (2000) used response surface methodology to improve the lift-to-drag ratio of artificially blunted leading edge spherical cone that is a representative geometry for reentry vehicle while constraining the heat transfer rates. Chung and Alonso (2000) estimated boom and drag for low boom supersonic business jet design problem *via* kriging.

Shyy et al. (2001a-b) employed global design optimization techniques for the design of rocket propulsion components; Papila et al. (2000, 2001, 2002), Papila 2001, approximated the design objectives for supersonic turbines using polynomial response surface approximations and neural networks; Vaidyanathan et al. (2000, 2004a-b), Vaidyanathan (2004) modeled performance indicators of liquid rocket injectors using polynomial response surface approximations; Simpson et al. (2001b) used kriging and polynomial response surface approximations for the design of aerospike nozzles; Steffen (2002a) developed an optimal design of a scramjet injector to simultaneously improve efficiency and pressure loss characteristics using polynomial response surfaces. In a follow-up work, Steffen et al. (2002b) used polynomial response surface approximations to analyze the parameter space for the design of a combined cycle propulsion components, namely, mixed-compression inlet, hydrogen fueled scramjet combustor, a ducted-rocket nozzle.

Charania et al. (2002) used response surface methodology to reduce computational cost in system-level uncertainty assessment for reusable launch vehicle design. Huque and Jahingir (2002) used neural networks in optimization of integrated inlet/ejector system of an axisymmetric rocket-based combined cycle engine. Qu et al. (2003) applied different polynomial response surface approximations for the structural design of hydrogen tanks; Keane (2003) employed response surfaces for the optimal design of wing shapes; Umakant et al. (2004) used

kriging surrogate models to compute probability density functions for robust configuration design of air-breathing hypersonic vehicle. Baker et al. (2004a-b) used response surface methodology for system level optimization of booster and ramjet combustor while considering multiple constraints. Levy et al. (2005) used support vector machines and neural networks to approximate objectives temperature field and pressure loss while searching Pareto optimal solutions for the design of a combustor.

Issues with Surrogate Modeling

The accuracy of surrogate models is an important factor for its effective application in design and optimization process. Some of the issues that influence the accuracy of surrogate models are: (1) number and location of sampled data points, (2) numerical simulations, and (3) choice of the surrogate model. The characterization of uncertainty in predictions *via* different error estimation measures is also useful for optimization algorithms such as, EGO (Jones et al., 1998) etc. These issues are widely discussed in literature (Li and Padula 2005, Queipo et al. 2005). While a detailed review of different issues and surrogate models is provided in next chapter, we briefly discuss the intent of current research in context of addressing the issues related to surrogate prediction accuracy.

Sampling Strategies

Typically the amount of data used for surrogate model construction is limited by the availability of computational resources. Many design of experiments (DOE) techniques are developed to sample the locations for conducting simulations such that the errors in approximation are reduced. However, these strategies mostly optimize a single criterion that caters to an assumed source of error in approximation. The dominant sources of errors in approximation are rarely known *a priori* for practical engineering problems. This renders the selection of appropriate DOE technique a very difficult task because an unsuitable DOE may

lead to poor approximation. We demonstrate this issue of high errors in approximation due to a single-criterion based DOE with the help of simple examples, and highlight the need to simultaneously consider multiple criteria.

Type of Surrogate Model

The influence of the choice of surrogate model on prediction accuracy is widely explored in literature. Many researchers have compared different surrogates for various problems and documented their recommendations. The general conclusion is that no single surrogate model may prove adequate for all problems, and the selection of a surrogate for a given problem is often influenced by the past experience. However, as we demonstrate in this work, the suitability of any surrogate model also depends on the sampling density and sampling scheme. Then, the selection of an appropriate surrogate model becomes more complicated exercise. Here, we present methods to exploit available information by simultaneous use of multiple surrogate models. Specifically, we use multiple surrogate models to assess the regions of high uncertainty in response predictions. Further we develop a weighted average surrogate model, which is demonstrated to be more robust than the individual surrogate models for a wide variety of problems and sampling schemes.

Estimation of Errors in Surrogate Predictions

Since surrogate models only approximate the response, there are errors in predictions. An estimate of prediction errors is beneficial in determining the sampling locations in many surrogate-based optimization methods such as EGO (Jones et al., 1998) and for adaptive sampling. The effectiveness of these methods depends on the accuracy of error estimation measures. Mostly, these error estimation measures are based on statistical assumptions. For example, prediction variance for polynomial response surface approximation is developed assuming that errors are exclusively due to noise that follows a normal distribution of zero mean

and σ^2 variance independent of locations. When these assumptions are not satisfied, the accuracy of error estimation measures is questionable. We compare different error estimation measures using a variety of test problems and give recommendations. We also explore the idea of simultaneously using multiple error estimation measures.

While we use the presented surrogate-based framework for design and optimization of diffuser design and cryogenic cavitation model, we employ analytical examples, which are primarily used to test optimization algorithms, to exhibit the key concepts relating to different issues with surrogate modeling.

Scope of Current Research

In short, the goal of present work is to develop methodologies for designing optimal propulsion systems while addressing issues related to numerical uncertainties in surrogate modeling. The scope of the current work can be summarized as follows:

- 1) To illustrate risks in using a single criterion based experimental design for approximation and the need to consider multiple criteria.
- 2) To explore the use of an ensemble of surrogates to help identify regions of high uncertainties in predictions and to possibly provide a robust prediction method.
- 3) To appraise different error estimation measures and to present methods to enhance the error detection capability by combining multiple error measures.
- 4) To demonstrate the proposed surrogate-model based approach to liquid rocket propulsion problems dealing with (a) cryogenic-cavitation model validation and sensitivity study to appraise the influence of different parameters on performance, and (b) shape optimization of diffuser vanes to maximize diffuser efficiency.

The organization of this work is as follows. We review different surrogate models and relevant issues associated with surrogate modeling in Chapter 2. In Chapter 3, we evidence risks in using a single criterion for constructing design of experiments. Methods to harness the potential of multiple surrogate models are presented in Chapter 4. The performance of different error estimation measures is compared in Chapter 5 and we propose methods to simultaneously use multiple error measures. This surrogate-based analysis and optimization framework is applied to two problems related to liquid rocket propulsion, model validation and sensitivity analysis of a cryogenic cavitation model in Chapter 6, and shape optimization of diffuser vanes in Chapter 7. We recapitulate major conclusions of the current work and delineate the scope of future work in Chapter 8.

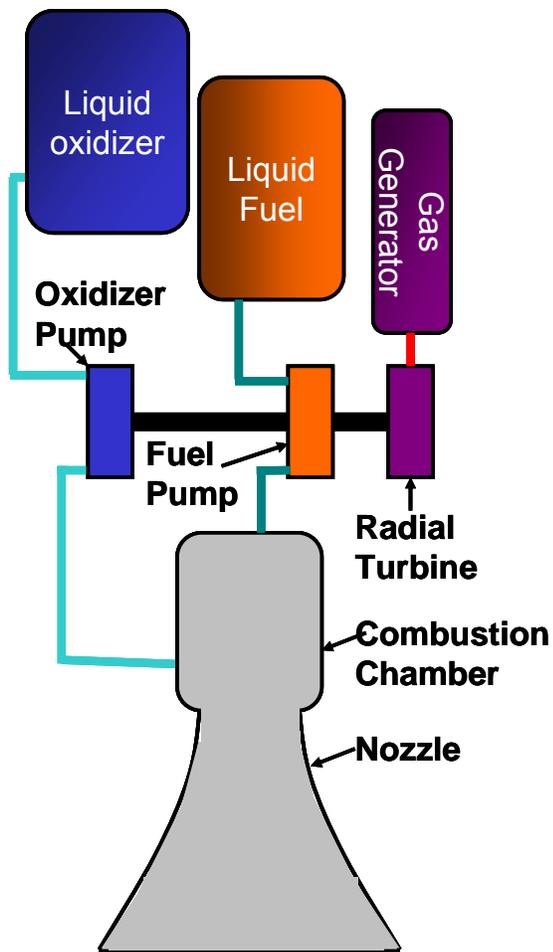


Figure 1-1. Schematic of liquid fuel rocket propulsion system.

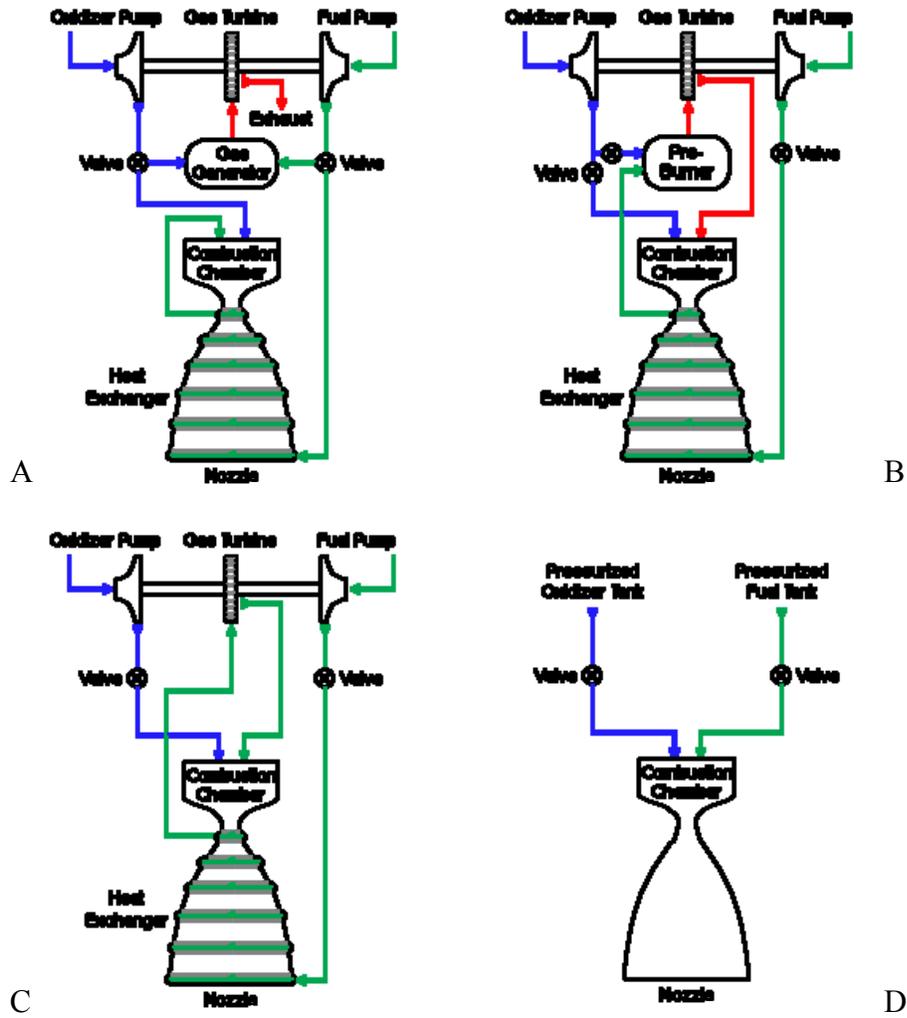


Figure 1-2. Classification of propulsion systems according to power cycles. A) Gas-generator cycle. B) Staged combustion cycle. C) Expander cycle. D) Pressure-fed cycle.

CHAPTER 2 ELEMENTS OF SURROGATE MODELING

Surrogate models are widely accepted for the design and optimization of components with high computational or experimental cost (typically encountered in CFD simulations based designs), as they offer a computationally less expensive way of evaluating designs. Surrogate models are constructed using the limited data generated from the analysis of carefully selected designs. Numerous successful applications of surrogate models for design and optimization of aerospace systems, automotive components, electromagnetic applications and chemical processes etc. are available in literature. A few examples are given as follows.

Kaufman et al. (1996), Balabanov et al. (1996, 1999), Papila and Haftka (1999, 2000), and Hosder et al. (2001) constructed polynomial response surface approximations (PRS) for structural weight based on structural optimizations of high speed civil transport. Hill and Olson (2004) applied PRS to approximate noise models in their effort to reduce the noise in the conceptual design of transport aircrafts. Madsen et al. (2000), Papila et al. (2000, 2001), Shyy et al. (2001a-b), Vaidyanathan et al. (2000, 2004a-b), Goel et al. (2004), and Mack et al. (2005b, 2006) used polynomial- and neural networks-based surrogate models as design evaluators for the optimization of propulsion components including turbulent flow diffuser, supersonic turbine, swirl coaxial injector element, liquid rocket injector, and radial turbine designs. Burman et al. (2002), Goel et al. (2005), and Mack et al. (2005a) used different surrogates to maximize the mixing efficiency facilitated by a trapezoidal-shaped bluff body in the time dependent Navier-Stokes flow while minimizing the resulting drag coefficient.

Knill et al. (1999), Rai and Madavan (2000, 2001) and, Madavan et al. (2001) used surrogate models for airfoil shape optimization. Dornberger et al. (2000) used neural networks and polynomial response surfaces for design of turbine blades. Kim et al. (2002), Keane (2003),

Jun et al. (2004) applied PRS to optimize wing designs. Ong et al. (2003) used radial basis functions to approximate the objective function and constraints of an aircraft wing design.

Bramantia et al. (2001) used neural network-based models to approximate the design objectives in electro-magnetic problems. Farina et al. (2001) used multiquadrics interpolation-based response surface approximations to optimize the shape of electromagnetic components like, C-core and magnetizer. Wilson et al. (2001) used response surface approximations and kriging for approximating the objectives while designing piezomorph actuators.

Redhe et al. (2002a-b), Craig et al. (2002), and Stander et al. (2004) used PRS, kriging and neural networks in design of vehicles for crashworthiness. Rais-Rohani and Singh (2002) used PRS to approximate limit state functions for estimating the reliability of composite structures. Rikards and Auzins (2002) developed PRS to model buckling and axial stiffness constraints while minimizing the weight of composite stiffened panels. Vittal and Hajela (2002) proposed using PRS to estimate the statistical confidence intervals on the reliability estimates. Zerpa et al. (2005) used kriging, radial basis functions, and PRS to optimize the cumulative oil recovery from a heterogeneous, multi-phase reservoir subject to an ASP (alkaline-surfactant-polymer) flooding.

Steps in Surrogate Modeling

Li and Padula (2005) and Queipo et al. (2005) have given a comprehensive review of the relevant issues in surrogate modeling. In this chapter, we discuss the key steps in the surrogate modeling as explained with the help of Figure 2-1. Discussion of major issues and the most prominent approaches followed in each step of the surrogate modeling process, briefly described below, lays the outline of this chapter.

Design of Experiments (DOEs)

The design of experiment is the sampling plan in design variable space. Other common names for DOEs are experimental designs or sampling strategies. The key question in this step is how we assess the goodness of such designs, considering the number of samples is severely limited by the computational expense of each sample. We discuss the most prominent approaches related to DOE in a subsequent section. Later in Chapter 3, we demonstrate some practical issues with the construction of DOEs.

Numerical Simulations at Selected Locations

Here, the computationally expensive model is executed for all the designs selected using the DOE specified in the previous step. In context of the present work, the details of relevant numerical simulation tools used to evaluate designs were briefly discussed in appropriate chapters.

Construction of Surrogate Model

Two questions are of interest in this step: 1) what surrogate model(s) should we use (model selection) and, 2) how do we find the corresponding parameters (model identification)? A formal description of the problem of interest is discussed in next section. A framework for the discussion and mathematical formulation of alternative surrogate-based modeling approaches is outlined in next section and the section on construction of surrogate models.

Model Validation

The purpose of this step is to establish the predictive capabilities of the surrogate model away from the available data (generalization error). Different schemes to estimate generalization error for model validation are discussed in this chapter. We compare different error estimation measures specific to a few popular surrogates in a following chapter.

Mathematical Formulation of Surrogate Modeling Problem

With reference to Figure 2-2, surrogate modeling can be seen as a non-linear inverse problem for which one aims to determine a continuous function ($y(\mathbf{x})$) of a set of design variables from a limited amount of available data (\mathbf{y}). The available data \mathbf{y} while deterministic in nature can represent exact evaluations of the function $y(\mathbf{x})$ or noisy observations; and in general cannot carry sufficient information to uniquely identify $y(\mathbf{x})$. Thus, surrogate modeling deals with the twin problems of: 1) constructing a model $\hat{y}(\mathbf{x})$ from the available data \mathbf{y} (model estimation), and 2) assessing the errors ε attached to it (model appraisal). A general description of the anatomy of inverse problems can be found in Snieder (1998).

Using the surrogate modeling approach the prediction of the simulation-based model output is formulated as $y(\mathbf{x}) = \hat{y}(\mathbf{x}) + \varepsilon(\mathbf{x})$. The prediction expected value and its variance $V(y)$ are illustrated in Figure 2-3, with θ being a probability density function.

Different model estimation and model appraisal components of the prediction have been shown to be effective in the context of surrogate based analysis and optimization (see for example, McDonald et al., 2000; Chung and Alonso, 2000; Simpson et al., 2001a; Jin et al., 2001), namely polynomial response surface approximation (PRS), Gaussian radial basis functions (GRF) (also referred as radial basis neural networks RBNN), and (ordinary) kriging (KRG) as described by Sacks et al. (1989). Model estimation and appraisal components of these methods are presented in a following section.

A good paradigm to illustrate how particular solutions (\hat{y}) to the model estimation problem can be obtained is provided by regularization theory (see for example, Tikhonov and Arsenin (1977), and Morozov (1984)), which imposes additional constraints on the estimation.

More precisely, \hat{y} can be selected as the solution to the following Tikhonov regularization problem:

$$\min_{\hat{y} \in S} \square(\hat{y}) = \frac{1}{N_s} \sum_{i=1}^{N_s} L(y_i - \hat{y}(\mathbf{x}^{(i)})) + \lambda \int \|D^m \hat{y}\|_S dx, \quad (2.1)$$

where S is the family of surrogate models under consideration, $L(x)$ is a loss or cost function used to quantify the so called empirical error (e.g., $L(x) = x^2$), λ is a regularization parameter, and $D^m \hat{y}$ represents the value of the m -derivative of the proposed model at location \mathbf{x} . Note that $D^m \hat{y}$ represents a penalty term; for example, if m is equal to two, it penalizes high local curvature. Hence, the first term enforces closeness to the data (goodness of fit), while the second term addresses the smoothness of the solution with λ (a real positive number) establishing the tradeoff between the two. Increasing values of λ provide smoother solutions. The purpose of the regularization parameter λ is, hence, to help implement Occam's razor principle (Ariew, 1976), which favors parsimony or simplicity in model construction. A good discussion on statistical regularization of inverse problems can be found in Tenorio (2001).

The quadratic loss function (i.e., L_2 norm) is most commonly used in part because it typically allows easy estimation of the parameters associated with the surrogate model; however, it is very sensitive to outliers. The linear (also called Laplace) loss function takes the absolute value of its argument (i.e., L_1 norm); on the other hand, the Huber loss function is defined as quadratic for small values of its argument and linear otherwise. The so called ε -loss function has received considerable attention in the context of the support vector regression surrogate (Vapnik, 1998; Girosi, 1998), and assigns an error equal to zero if the true and estimated values are within an ε distance. Figure 2-4 illustrates the cited loss functions.

Design of Experiments

As stated earlier, the design of experiment is the sampling plan in design variable space and the key question in this step is how we assess the goodness of such designs. In this context, of particular interest are sampling plans that provide a unique value (in contrast to random values) for the input variables at each point in the input space, and are model-independent; that is, they can be efficiently used for fitting a variety of models.

Typically, the primary interest in surrogate modeling is minimizing the error and a DOE is selected accordingly. Two major components of the empirical error and corresponding expressions for average error formulation considering quadratic loss functions are given as follows.

- **Variance:** measures the extent to which the surrogate model $\hat{y}(\mathbf{x})$ is sensitive to particular data set D . Each data set D corresponds to a random sample of the function of interest. This characterizes the noise in the data,

$$E_{\text{var}}(\mathbf{x}) = E_{ADS} [\hat{y} - E_{ADS}[\hat{y}(\mathbf{x})]]^2. \quad (2.2)$$

- **Bias:** quantifies the extent to which the surrogate model outputs (i.e., $\hat{y}(\mathbf{x})$) differ from the true values (i.e., $y(\mathbf{x})$) calculated as an average over all possible data sets D (ADS),

$$E_{\text{bias}^2}(\mathbf{x}) = \{E_{ADS} [\hat{y}(\mathbf{x})] - y(\mathbf{x})\}^2. \quad (2.3)$$

In both expressions, E_{ADS} denotes the expected value considering all possible data sets.

There is a natural tradeoff between bias and variance. A surrogate model that closely fits a particular data set (lower bias) will tend to provide a larger variance and vice-versa. We can decrease the variance by smoothing the surrogate model but if the idea is taken too far then the bias error becomes significantly higher. In principle, we can reduce both bias (can choose more complex models) and variance (each model more heavily constrained by the data) by increasing the number of points, provided the latter increases more rapidly than the model complexity.

In practice, the number of points in the data set is severely limited (e.g., due to computational expense) and often during the construction of surrogate model, a balance between bias and variance errors is sought. This balance can be achieved, for example, by reducing the bias error while imposing penalties on the model complexity (e.g., Tikhonov regularization).

With reference to most applications, where the actual model is unknown (see previous and next sections), and data is collected from deterministic computer simulations, *bias error* is the dominant source of error because the numerical noise is small, and a DOE is selected accordingly. When response surface approximation is used, there is good theory for obtaining minimum bias designs (Myers and Montgomery, 1995, Section 9.2) as well as some implementations in low dimensional spaces (Qu et al., 2004). For the more general case, the bias error can be reduced through a DOE that distributes the sample points uniformly in the design space (Box and Draper, 1959; Sacks and Ylvisaker, 1984; as referenced in Tang, 1993). However, fractional factorial designs replace dense full factorial designs for computationally expensive problems. The uniformity property in designs is sought by, for example, maximizing the minimum distances among design points (Johnson et al., 1990), or by minimizing correlation measures among the sample data (Iman and Conover, 1982; Owen, 1994). Practical implementations of a few most commonly used DOEs are discussed next.

Factorial Designs

Factorial designs are one of the simplest DOEs to investigate the effect of main effects and interactions of variables on the response for box-shaped design domain. 2^{N_v} -factorial design where each design variable is allowed to take two extreme levels is often used as a screening DOE to eliminate the unimportant variables. Qualitative and binary variables can also be used for this DOE. A typical two-level full factorial DOE for three variables is shown in Figure 2-5.

These designs can be used to create a linear polynomial response surface approximation. For higher order approximations, the number of levels of each variable is increased for example a quadratic polynomial response surface approximation can be fitted by a three-level full factorial design (3^{N_v} -designs).

Some times the number of experiments is reduced by using 2^{N_v-p} (or $1/2^p$) fractional factorial designs, which require the selection of p independent design generators (least influential interactions). Typically, these designs are classified according to resolution number:

- **Resolution III:** Main effects are not aliased with other main effects, but are confounded with one or more two-way interactions.
- **Resolution IV:** Main effects are not aliased with other main effects or two-way interactions. Two factor interactions are confounded with other two-way interactions.
- **Resolution V:** Main effects and two-way interactions are not confounded with one another.

More details can be obtained from Myers and Montgomery (1995, Chapter 4, pp. 156-179). Factorial designs produce orthogonal designs for polynomial response surface approximation. However, for higher dimensions ($N_v > 6$), factorial designs requires a large number of experiments making them particularly unattractive for computationally expensive problems.

Central Composite Designs

These include designs on 2^{N_v} vertices, $2N_v$ axial points, and N_c repetitions of central point. The distance of axial point α is varied to generate, face-centered, spherical design, or orthogonal design. A typical central composite design for three-variable problem is shown in Figure 2-6. These designs reduce variance component of the error. The repetitions at the center reduce the variance, improve stability (defined as the ratio of maximum variance to minimum variance in the entire design space), and give an idea of magnitude of the noise, but are not

useful for the computer simulations. These designs are also not practical for higher dimension spaces ($N_v > 8$) as the number of simulations becomes very high. For $N_v > 3$, when the designs on the vertices of the design spaces are not feasible, Box-Behnken designs can be used for quadratic polynomial response surface approximation. These are spherical designs (sampling designs at these locations enables us to exactly determine the function value on the points equidistant from the center) but these designs introduce higher uncertainty near the vertices.

Variance Optimal DOEs for Polynomial Response Surface Approximations

Moment matrix $M = X^T X / N_s$ ¹ for PRS (see next section) is very important quantity as this affects the prediction variance, and the confidence on the coefficients, hence, used to develop different variance optimal DOEs. *D*-optimal design maximizes the determinant of the moment matrix M to maximize the confidence in the coefficients of polynomial response surface approximation. *A*-optimal design minimizes the trace of inverse of M to minimize the sum of variances of the coefficients. *G*-optimal design minimizes the maximum of prediction variance. *I*-optimal design minimizes the integral of the prediction variance over the design domain. All these DOEs require the solution of difficult optimization problem, which is solved heuristically in higher dimensional spaces.

Latin Hypercube Sampling (LHS)

Stratified sampling ensures that all portions of a given partition are sampled. LHS (McKay et al., 1979) is a stratified sampling approach with the restriction that each of the input variables (x_k) has all portions of its distribution represented by input values. A sample of size N_s can be constructed by dividing the range of each input variable into N_s strata of equal marginal

¹ An example of matrix X is given in Equation (2.6).

probability $1/N_s$ and sampling once from each stratum. Let us denote this sample by

$x_k^{(j)}, j = 1, 2, \dots, N_s; k = 1, 2, \dots, N_v$. The sample is made of components of each of the x_{N_v} 's

matched at random. Figure 2-7 illustrates a LHS design for two variables, when six designs are selected.

While LHS represents an improvement over unrestricted stratified sampling (Stein, 1987), it can provide sampling plans with very different performance in terms of uniformity measured by, for example, maximum minimum-distance among design points, or by correlation among the sample data. Figure 2-8 illustrates this shortcoming; the LHS plan in Figure 2-8(B) is significantly better than that in Figure 2-8(A).

Orthogonal Arrays (OA)

These arrays were introduced by C. R. Rao in the late 40's (Rao, 1947), and can be defined as follows. An OA of strength r is a matrix of N_s rows and N_v columns with elements taken from a set of N_q symbols, such that in any $N_s \times r$ submatrices each of the $(N_q)^r$ possible rows occur the same number γ (index) of times. The number of rows (N_s) and columns (N_v) in the OA definition represents the number of samples and variables or factors under consideration, respectively. The N_q symbols are related to the levels defined for the variables of interest, and the strength r is an indication of how many effects can be accounted for (to be discussed later in this section) with values typically between two and four for real-life applications. Such an array is denoted by $OA(N_s, N_v, N_q, r)$. Note that, by definition, a LHS is an OA of strength one, $OA(N_s, N_v, N_q, 1)$. There are two limitations on the use of OA for DOE:

- **Lack of flexibility:** Given desired values for the number of rows, columns, levels, and strength, the OA may not exist. For a list of available orthogonal arrays, theory and applications, see, for example, Owen (1992), Hedayat et al. (1999), and references therein.

- **Point replicates:** OA designs projected onto the subspace spanned by the effective factors (most influential design variables) can result in replication of points. This is undesirable for deterministic computer experiments where the bias of the proposed model is the main concern.

Optimal LHS, OA-based LHS, Optimal OA-based LHS

Different approaches have been proposed to overcome the potential lack of uniformity of LHS. Among those, most of them adjust the original LHS by optimizing a spread measure (e.g., minimum distance or correlation) of the sample points. The resulting designs have been shown to be relatively insensitive to the optimal design criterion (Ye et al., 2000). Examples of this strategy can be found in the works of Iman and Conover (1982), Johnson et al. (1990), and Owen (1994). Tang (1993) and Ye (1998) presented the construction of strength r OA-based LHS which stratify each r -dimensional margin, and showed that they offer a substantial improvement over standard LHS. Another strategy optimizes a spread measure of the sample points, but restricts the search of LHS designs, which are orthogonal arrays, resulting in so called optimal OA-based LHS (Leary et al., 2003). Adaptive DOE, in which model appraisal information is used to place additional samples, have also been proposed (Jones et al., 1998, Sasena et al., 2000, Williams et al., 2000).

A summary of main characteristics and limitations of different DOE techniques is listed in Table 2-1. If feasible, two sets of DOE are generated, one (so called training data set) for the construction of the surrogate (next section), and second for assessing its quality (validation as discussed in a later section). Given the choice of surrogate, the DOE can be optimized to suit a particular surrogate. This has been done extensively for minimizing variance in polynomial RSA (e.g., D - and A - optimal designs, Myers and Montgomery, 1995, Chapter 8) and to some extent for minimizing bias (e.g., Qu et al., 2004). However, for non-polynomial models, the cost of the optimization of a surrogate-specific DOE is usually prohibitive, and so is rarely attempted.

Construction of Surrogate Model

There are both parametric (e.g., polynomial response surface approximation, kriging) and non-parametric (e.g., projection-pursuit regression, radial basis functions) alternatives for constructing surrogate models. The parametric approaches presume the global functional form of the relationship between the response variable and the design variables is known, while the non-parametric ones use different types of simple, local models in different regions of the data to build up an overall model. This section discusses the estimation and appraisal components of the prediction of a sample of both parametric and non-parametric approaches.

Specifically, the model estimation and appraisal components corresponding to polynomial response surface approximation (PRS), kriging (KRG), and radial basis functions (RBF) surrogate models are discussed next, followed by a discussion of a more general non-parametric approach called *kernel-based regression*. Throughout this section a square loss function is assumed unless otherwise specified, and given the stochastic nature of the surrogates, the available data is considered a sample of a population.

Polynomial Response Surface Approximation (PRS)

The regression analysis is a methodology to study the quantitative relation between a function of interest y , and N_{β_1} basis functions f_j , using N_s sample values of the response y_i , for a set of basis functions $f_j^{(i)}$ (Draper and Smith, 1998, Section 5.1). Monomials are the most preferred basis functions by practitioners. For each observation i , a linear equation is formulated:

$$y_i(\mathbf{f}(\mathbf{x})) = \sum_{j=1}^{N_{\beta_1}} \beta_j f_j^{(i)} + \varepsilon_i; \quad E(\varepsilon_i) = 0, \quad V(\varepsilon_i) = \sigma^2, \quad (2.4)$$

where the errors ε_i are considered independents with expected value equal to zero and variance σ^2 , and $\boldsymbol{\beta}$ represents the quantitative relation between basis functions. The set of equations specified in Equation (2.4) can be expressed in matrix form as:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad E(\boldsymbol{\varepsilon}) = 0, \quad V(\boldsymbol{\varepsilon}) = \sigma^2 I, \quad (2.5)$$

where X is a $N_s \times N_{\beta_1}$ matrix of basis functions, also known as Gramian design matrix, with the design variable values for sampled points. A Gramian design matrix for a quadratic polynomial in two variables ($N_v = 2; N_{\beta_1} = 6$) is shown in Equation (2.6)

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_1^{(1)2} & x_1^{(1)} x_2^{(1)} & x_2^{(1)2} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_1^{(2)2} & x_1^{(2)} x_2^{(2)} & x_2^{(2)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(i)} & x_2^{(i)} & x_1^{(i)2} & x_1^{(i)} x_2^{(i)} & x_2^{(i)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(N_s)} & x_2^{(N_s)} & x_1^{(N_s)2} & x_1^{(N_s)} x_2^{(N_s)} & x_2^{(N_s)2} \end{bmatrix}. \quad (2.6)$$

The polynomial response surface approximation of the observed response $y(\mathbf{x})$ is,

$$\hat{y}(\mathbf{x}) = \sum_j b_j f_j(\mathbf{x}), \quad (2.7)$$

where b_j is the estimated value of the coefficient associated with the j^{th} basis function $f_j(\mathbf{x})$.

Then, the error in approximation at a design point \mathbf{x} is given as $e(\mathbf{x}) = y(\mathbf{x}) - \hat{y}(\mathbf{x})$. The

coefficient vector \mathbf{b} can be obtained by minimizing a loss function L , defined as

$$L = \sum_{i=1}^{N_s} |e_i|^p, \quad (2.8)$$

where e_i is the error at i^{th} data point, p is the order of loss function, and N_s is the number of sampled design points. A quadratic loss function ($p = 2$), that minimizes the variance of the error

in approximation, is mostly used because we can estimate coefficient vector \mathbf{b} using an analytical expression, as follows,

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}. \quad (2.9)$$

The estimated parameters \mathbf{b} (by least squares) are unbiased estimates of $\boldsymbol{\beta}$ that minimize variance. At a new set of basis function vector \mathbf{f} for design point P , the predicted response and the variance of the estimation are given as

$$\hat{y}(\mathbf{x}) = \sum_{j=1}^{N_{\beta_1}} b_j f_j(\mathbf{x}), \quad \text{and} \quad V(\hat{y}(\mathbf{x})) = \sigma^2 (\mathbf{f}^T (X^T X)^{-1} \mathbf{f}). \quad (2.10)$$

Kriging Modeling (KRG)

Kriging is named after the pioneering work of D.G. Krige (a South African mining engineer), and was formally developed by Matheron (1963). More recently, Sacks et al. (1989, 1993), and Jones et al. (1998) made it well-known in the context of the modeling, and optimization of deterministic functions, respectively. The kriging method in its basic formulation estimates the value of a function (response) at some unsampled location as the sum of two components: the linear model (e.g., polynomial trend), and a systematic departure representing low (large scale) and high frequency (small scale) variation components, respectively.

The systematic departure component represents the fluctuations around the trend, with the basic assumption being that these are correlated, and the correlation is a function of distance between the locations under consideration. More precisely, it is represented by a zero mean, second-order, stationary process (mean and variance constant with a correlation depending on a distance) as described by a correlation model.

Hence, these models (ordinary kriging) suggest estimating deterministic functions as:

$$y(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}), \quad E(\varepsilon) = 0, \quad \text{cov}(\varepsilon(\mathbf{x}^{(i)}), \varepsilon(\mathbf{x}^{(j)})) \neq 0 \quad \forall i, j, \quad (2.11)$$

where μ is the mean of the response at sampled design points, and ε is the error with zero expected value, and with a correlation structure that is a function of a generalized distance between the sample data points. A possible correlation structure (Sacks et al., 1989) is given by:

$$\text{cov}(\varepsilon(\mathbf{x}^{(i)}), \varepsilon(\mathbf{x}^{(j)})) = \sigma^2 \exp\left(-\sum_{k=1}^{N_v} \phi_k (x_k^{(i)} - x_k^{(j)})^2\right), \quad (2.12)$$

where N_v denotes the number of dimensions in the set of design variables \mathbf{x} ; σ identifies the standard deviation of the response at sampled design points, and, ϕ_k is a parameter, which is a measure of the degree of correlation among the data along the k^{th} direction. Specifically, the parameters μ , ε , and ϕ are estimated using a set of N_s samples (\mathbf{x}, y) , such that a likelihood function is maximized (Sacks et al., 1989). Given a probability distribution and the corresponding parameters, the likelihood function is a measure of the probability of the sample data being drawn from it. The model estimates at unsampled points is:

$$E(f_p(\mathbf{x})) = \hat{\mu} + \mathbf{r}^T R^{-1}(\mathbf{f} - \mathbf{1}\hat{\mu}), \quad (2.13)$$

where $\hat{\cdot}$ above the letters denotes estimates, \mathbf{r} identifies the correlation vector between the set of prediction points and the points used to construct the model, R is the correlation matrix among the N_s sample points, and $\mathbf{1}$ denotes an N_s -vector of ones. On the other hand, the estimation variance at unsampled design points is given by:

$$V(y(\mathbf{x})) = \sigma^2 \left[1 - \mathbf{r}^T R^{-1} \mathbf{r} + \frac{(\mathbf{1} - \mathbf{1}^T R^{-1} \mathbf{r})}{\mathbf{1}^T R^{-1} \mathbf{1}} \right]. \quad (2.14)$$

Gaussian processes (Williams and Rasmussen, 1996, Gibbs, 1997), another well-known approach to surrogate modeling, can be shown to provide identical expressions for the prediction and prediction variance to those provided by kriging, under the stronger assumption that the

available data (model responses) is a sample of a multivariate normal distribution (Rao, 2002, Section 4a).

Radial Basis Functions (RBF)

Radial basis functions have been developed for the interpolation of scattered multivariate data. The method uses linear combinations of N_{RBF} radially symmetric functions $h_i(\mathbf{x})$, based on Euclidean distance or other such metric, to approximate response functions as,

$$y(\mathbf{x}) = \sum_{i=1}^{N_{RBF}} w_i h_i(\mathbf{x}) + \varepsilon_i, \quad (2.15)$$

where \mathbf{w} represents the coefficients of the linear combinations, $h_i(\mathbf{x})$ the radial basis functions, and ε_i independent errors with variance σ^2 .

The flexibility of the model, that is its ability to fit many different functions, derives from the freedom to choose different values for the weights. Radial basis functions are a special class of functions with their main feature being that their response decreases (or increases) monotonically with distance from a central point. The center, the distance scale, and the precise shape of the radial function are parameters of the model.

A typical radial function is the Gaussian, which in the case of a scalar input, is expressed as,

$$h_i(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x} - \mathbf{c})^2}{\delta^2}\right). \quad (2.16)$$

The parameters are its center \mathbf{c} and its radius δ . Note that the response of the Gaussian RBF decreases monotonically with the distance from the center, giving a significant response only in the center neighborhood.

Given a set of N_s input/output pairs (sample data), a radial basis function model can be expressed in matrix form as,

$$\mathbf{f} = H\mathbf{w}, \quad (2.17)$$

where H is a matrix given by,

$$H = \begin{bmatrix} h_1(\mathbf{x}^{(1)}) & h_2(\mathbf{x}^{(1)}) & \cdots & h_{N_{RBF}}(\mathbf{x}^{(1)}) \\ h_1(\mathbf{x}^{(2)}) & h_2(\mathbf{x}^{(2)}) & \cdots & h_{N_{RBF}}(\mathbf{x}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\mathbf{x}^{(N_s)}) & h_2(\mathbf{x}^{(N_s)}) & \cdots & h_{N_{RBF}}(\mathbf{x}^{(N_s)}) \end{bmatrix}. \quad (2.18)$$

Similar to polynomial response surface approximation method, by solving Equation (2.17), the optimal weights (in the least squares sense) can be found to be,

$$\hat{\mathbf{w}} = A^{-1}H^T \mathbf{y}, \quad (2.19)$$

where A^{-1} is a matrix given by,

$$A^{-1} = (H^T H)^{-1}. \quad (2.20)$$

The error variance estimate can be shown to be given by,

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T P_r^2 \mathbf{y}}{tr(P_r)}, \quad (2.21)$$

where P_r is a projection matrix,

$$P_r = I - HA^{-1}H^T. \quad (2.22)$$

The RBF model estimate for a new set input values is given by,

$$\hat{y}(\mathbf{x}) = \mathbf{h}^T \hat{\mathbf{w}}, \quad (2.23)$$

where \mathbf{h} is a column vector with the radial basis functions evaluations,

$$\mathbf{h} = \begin{Bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ \vdots \\ h_{N_{RBF}}(\mathbf{x}) \end{Bmatrix}. \quad (2.24)$$

On the other hand, the prediction variance is the variance of the estimated model $\hat{y}(\mathbf{x})$ plus the error variance and is given by:

$$V(y(\mathbf{x})) = V(\mathbf{h}^T \hat{\mathbf{w}}) + V(\varepsilon) = \left(\mathbf{h}^T (H^T H)^{-1} \mathbf{h} + 1 \right) \frac{\mathbf{y}^T P_r \mathbf{y}}{N_s - N_{RBF}}. \quad (2.25)$$

Radial basis function is also known as radial basis neural networks (RBNN) as described by Orr (1996, 1999a-b). The MATLAB implementation of radial basis functions or RBNN (function ‘newrb’), used in this study, is described as follows. Radial basis neural networks are two-layer networks consisting of a radial-basis function and a linear output layer. The output of each neuron is given by

$$f = \text{radbas} \left(\left\| \mathbf{w} - \mathbf{x} \right\| \times 0.8326 / \text{spread} \right), \quad (2.26)$$

$$\text{radbas}(x) = \exp(-x^2), \quad (2.27)$$

where \mathbf{w} is the weight vector associated with each neuron, \mathbf{x} is the input design vector, ‘spread’ is a user defined value that controls the radius of influence for each neuron where the radius is half the value of parameter ‘spread’. Specifically, the radius of influence is the distance at which the influence reaches a certain small value. If ‘spread’ is too small, the prediction will be poor in regions that are not near the position of a neuron; and if ‘spread’ is too large, the sensitivity of the neurons will be small. Neurons are added to the network one by one until a specified mean square error goal ‘goal’ is reached. If the error goal is set to zero, neurons will be added until the network exactly predicts the input data. However, this can lead to over-fitting of the data, which may result in poor prediction between data points. On the other hand, if error goal is large, the

network will be under-trained and predictions even on data points will be poor. For this reason, an error goal is judiciously selected to prevent overfitting while keeping the overall prediction accuracy high.

Kernel-based Regression

The basic idea of RBF can be generalized to consider alternative loss functions and basis functions in a scheme known as kernel-based regression. With reference to Equation (2.1), it can be shown that independent of the form of the loss function L , the solution of the variational problem has the form (the Representer Theorem: see Girosi, 1998; Poggio and Smale, 2003):

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i G(\mathbf{x}, \mathbf{x}^{(i)}) + b, \quad (2.28)$$

where $G(\mathbf{x}, \mathbf{x}^{(i)})$ is a (symmetric) kernel function that determines the smoothness properties of the estimation scheme. Table 2-2 shows the kernel functions of selected estimation schemes with the kernel parameters being estimated by model selection approaches (see next section for details).

If the loss function L is quadratic, the unknown coefficients in Equation (2.28) can be obtained by solving the linear system,

$$(N_s \lambda I + G) \alpha = \mathbf{y}, \quad (2.29)$$

where I is the identity matrix, and G a square positive definite matrix with elements

$$G_{i,j} = G(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}). \text{ Note that, the linear system is well posed, since } (N_s \lambda I + G) \text{ is strictly}$$

positive and well conditioned for large values of $N_s \lambda$. If loss function L is non-quadratic, the solution of the variational problem still has the form of Equation (2.28) but the coefficients α_i are found by solving a quadratic programming problem in what is known as ‘support vector regression’ (Vapnik, 1998).

Major characteristics of different surrogate models are summarized in Table 2-3. Comparative studies have shown that depending on the problem under consideration, a particular modeling scheme (e.g., polynomial response surface approximation, kriging, radial basis functions) may outperform the others and in general, it is not known *a priori* which one should be selected. See for example, the works of Friedman and Stuetzle (1981), Yakowitz & Szidarovsky (1985), Laslett (1994), Giunta and Watson (1998), Simpson et al. (2001a-b), Jin et al. (2001). Considering, plausible alternative surrogate models can reasonably fit the available data, and the cost of constructing surrogates is small compared to the cost of the simulations, using multiple surrogates may offer advantages compared to the use of a single surrogate. Recently, multiple surrogate-based analysis and optimization approaches have been suggested by Zerpa et al. (2005) and Goel et al. (2006b) based on the model averaging ideas of Perrone and Cooper (1993), and Bishop (1995). The multiple surrogate-based analysis approach is based on the use of weighted average models, which can be shown to reduce the prediction variance with respect to that of the individual surrogates. The idea of multiple surrogate-based approximations is discussed in Chapter 4.

Model Selection and Validation

Generalization error estimates assess the quality of the surrogates for prediction and can be used for model selection among alternative models; and establish the adequacy of surrogate models for use in analysis and optimization studies (validation). This section discusses the most prominent approaches in the context of surrogate modeling.

Split Sample (SS)

In this scheme, the sample data is divided into training and test sets. The former is used for constructing the surrogate while the latter, if properly selected, allows computing an unbiased estimate of the generalization error. Its main disadvantages are, that the generalization error

estimate can exhibit a high variance (it may depend heavily on which points end up in the training and test sets), and that it limits the amount of data available for constructing the surrogates.

Cross Validation (CV)

It is an improvement on the split sample scheme that allows the use of the most, if not all, of the available data for constructing the surrogates. In general, the data is divided into k subsets (k -fold cross-validation) of approximately equal size. A surrogate model is constructed k times, each time leaving out one of the subsets from training, and using the omitted subset to compute the error measure of interest. The generalization error estimate is computed using the k error measures obtained (e.g., average). If k equals the sample size, this approach is called leave-one-out cross-validation (known also as PRESS in the polynomial response surface approximation terminology). Equation (2.30) represents a leave-one-out calculation when the generalization error is described by the mean square error (GMSE).

$$GMSE = \frac{1}{k} \sum_{i=1}^k (y_i - \hat{y}_i^{(-i)})^2, \quad (2.30)$$

where $\hat{y}_i^{(-i)}$ represents the prediction at $\mathbf{x}^{(i)}$ using the surrogate constructed from all sample points except $(\mathbf{x}^{(i)}, y_i)$. Analytical expressions are available for leave-one-out case for the GMSE without actually performing the repeated construction of the surrogates for both polynomial response surface approximation (Myers and Montgomery, 1995, Section 2.7) and kriging (Martin and Simpson, 2005).

The advantage of cross-validation is that it provides nearly unbiased estimate of the generalization error, and the corresponding variance is reduced (when compared to split-sample) considering that every point gets to be in a test set once, and in a training set $k-1$ times

(regardless of how the data is divided); the variance of the estimation though may still be unacceptably high in particular for small data sets. The disadvantage is that it requires the construction of k surrogate models; this is alleviated by the increasing availability of surrogate modeling tools. A modified version of the CV approach called GCV-generalized cross validation, which is invariant under orthogonal transformations of the data (unlike CV) is also available (Golub et al., 1979).

If the Tikhonov regularization approach for regression is adopted, the regularization parameter λ can be identified using one or more of the following alternative approaches: CV-cross validation (leave-one-out estimates), GCV (smoothed version of CV), or the L -curve (explained below). While CV and GCV can be computed very efficiently (Wahba, 1983; Hutchinson and de Hoog, 1985), they may lead to very small values of λ even for large samples (e.g., very flat GCV function). The L -curve (Hansen, 1992) is claimed to be more robust and have the same good properties of GCV. The L -curve is a plot of the residual norm (first term) versus the norm $\int \|D^m \hat{y}\|_H$ of the solution for different values of the regularization parameter and displays the compromise in the minimization of these two quantities. The best regularization parameter is associated with a characteristic L -shaped “corner” of the graph.

Bootstrapping

This approach has been shown to work better than cross-validation in many cases (Efron, 1983). In its simplest form, instead of splitting the data into subsets, subsamples of the data are considered. Each subsample is a random sample with replacement from the full sample, that is, it treats the data set as a population from which samples can be drawn. There are different variants of this approach (Hall, 1986; Efron and Tibshirani, 1993; Hesterberg et al., 2005) that can be

used for model identification as well as for identifying confidence intervals for surrogate model outputs. However, this may require considering several dozens or even hundreds of subsamples.

For example, in the case of polynomial response surface approximation (given a model), regression parameters can be estimated for each of the subsamples and a probability distribution (and then confidence intervals) for the parameters can be identified. Once the parameter distributions are estimated, confidence intervals on model outputs of interest (e.g., mean) can also be obtained.

Bootstrapping has been shown to be effective in the context of neural network modeling; recently, its performance in the context of model identification in regression analysis is also being explored (Ohtani, 2000, Kleijnen and Deflandre 2004).

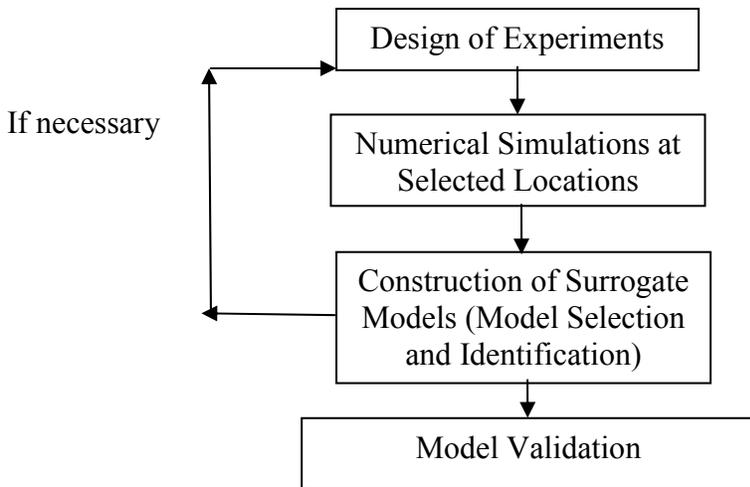


Figure 2-1. Key stages of the surrogate-based modeling approach.

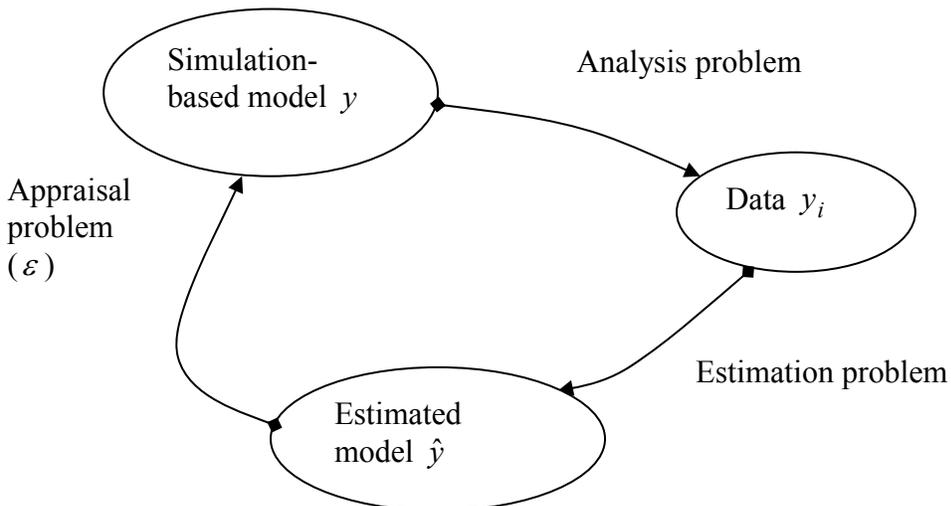


Figure 2-2. Anatomy of surrogate modeling: model estimation + model appraisal. The former provides an estimate of function while the latter forecasts the associated error.

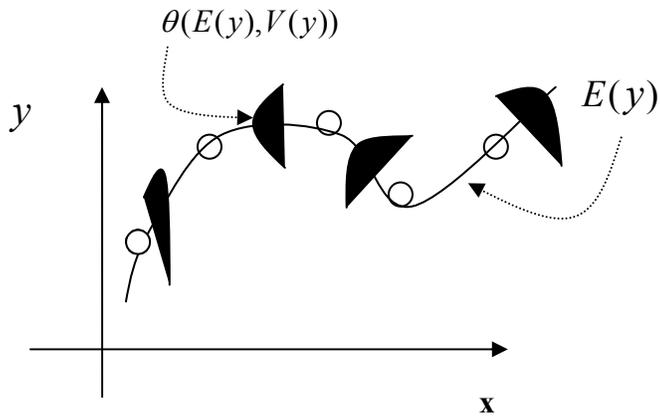


Figure 2-3. A surrogate modeling scheme provides the expected value of the prediction $E(y)$ (solid line) and the uncertainty associated with that prediction, illustrated here using a probability density function θ .

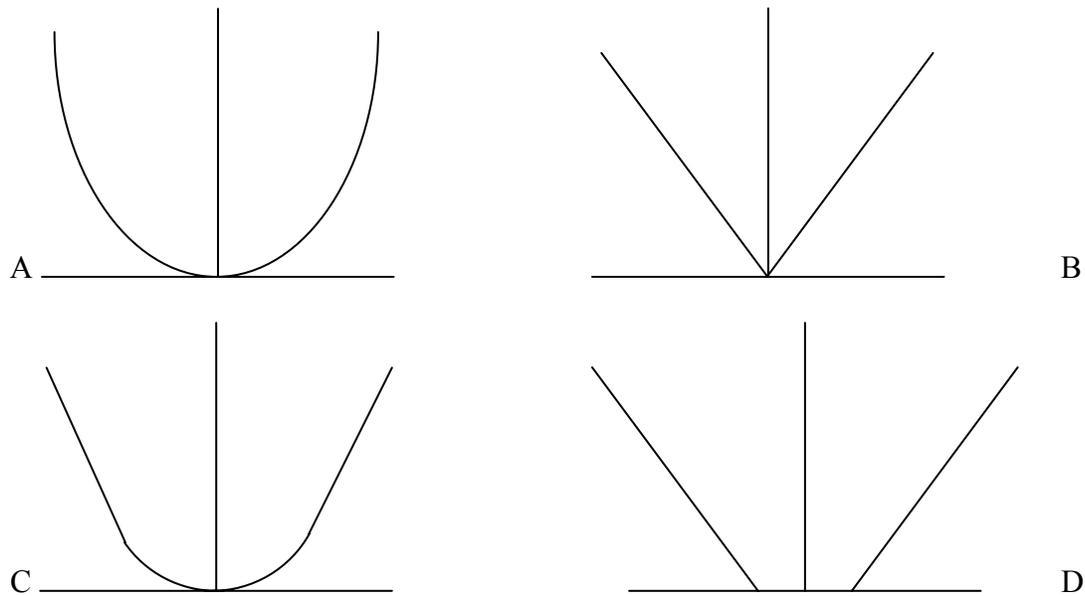


Figure 2-4. Alternative loss functions for the construction of surrogate models. A) Quadratic. B) Laplace. C) Huber. D) ϵ -loss function.

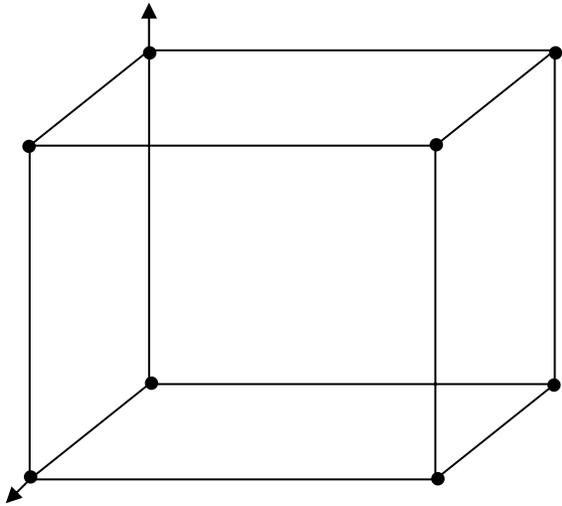


Figure 2-5. A two-level full factorial design of experiment for three variables.

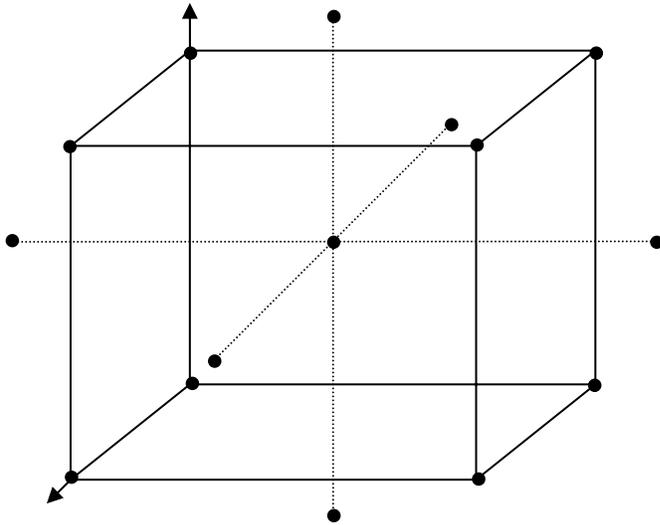


Figure 2-6. A central composite design for three-dimensional design space.

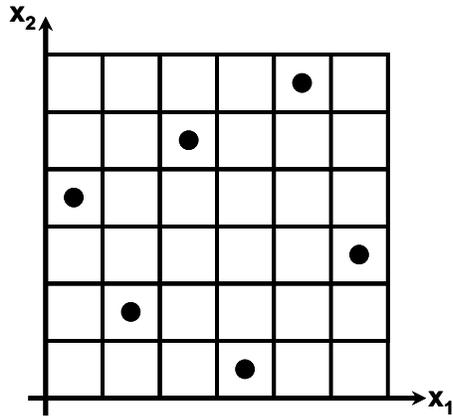


Figure 2-7. A representative Latin hypercube sampling design with $N_s = 6, N_v = 2$ for uniformly distributed variables in the unit square.

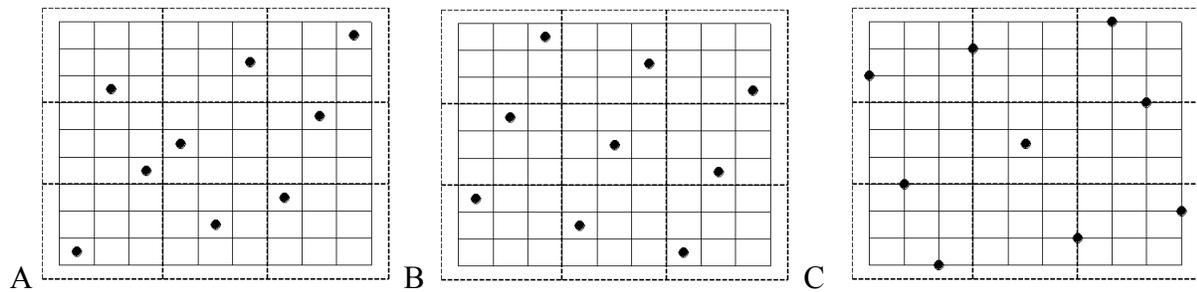


Figure 2-8. LHS designs with significant differences in terms of uniformity (Leary et al., 2003). A) Random LHS. B) Correlation optimized LHS. C) OA-based LHS. (Figure reprinted with kind permission of Taylor and Francis group, Leary et al., 2003, Figure 1).

Table 2-1. Summary of main characteristics of different DOEs.

| Name | Main features | Limitations |
|--------------------------------|--|---|
| Factorial designs | Used to investigate main effect and interaction of variables for box-shaped domains, gives orthogonal designs, caters to noise | Irregular domains , not good for $N_v > 6$ |
| CCD | Cater to noise, applicable for box-shaped domains, repetition of points on center to improve stability | Irregular domains, not good for $N_v > 8$, repetition of points not useful for simulations |
| Variance optimal designs | Cater to noise, Applicable to irregular domains too | High computational cost, not good when noise is low |
| <i>D</i> -optimal | - Maximize confidence in coefficients | |
| <i>A</i> -optimal | - Minimize the sum of variances of coefficients | |
| <i>G</i> -optimal | - Minimize the maximum of prediction variance | |
| <i>I</i> -optimal | - Minimize the integral of prediction variance over design domain | |
| Latin hypercube sampling (LHS) | Caters to bias error, stratified sampling method, good for high number of variables | Not good when noise is significant, occasional poor DOE due to random components |
| Orthogonal arrays (OA) | Box-shaped domains, the moment matrix is diagonal for monomial basis functions so coefficients of approximation are uncorrelated | Limited number of orthogonal arrays, difficult to create OAs |
| OA-based LHS | Combine OA and LHS designs to improve distribution of points | Limited OAs, may leave large holes in design space |

Table 2-2. Examples of kernel functions and related estimation schemes.

| Kernel function | Estimation scheme |
|--|--------------------------------------|
| $G(x, x_i) = (1 + \alpha x_i)^d$ | Polynomial of degree d (PRD) |
| $G(x, x_i) = \ x - x_i\ $ | Linear splines (LSP) |
| $G(x, x_i) = \exp\left(-\frac{\ x - x_i\ ^2}{\gamma_i}\right)$ | Gaussian radial basis function (GRF) |

Table 2-3. Summary of main characteristics of different surrogate models.

| Name | Main features |
|---|--|
| Polynomial response surface approximation | Global parametric approach, good for slow varying functions, easy to construct, good to handle noise, not very good for simulations based data |
| Kriging | Global parametric approach, handles smooth and fast varying functions, computationally expensive for large amount of data |
| Radial basis function | Local, non-parametric approach, computationally expensive, good for fast varying functions |
| Kernel-based functions | Global, non-parametric approach, uses different loss functions, relatively new approach |

CHAPTER 3 PITFALLS OF USING A SINGLE CRITERION FOR SELECTING EXPERIMENTAL DESIGNS

Introduction

Polynomial response surface (PRS) approximations are widely adopted for solving optimization problems with high computational or experimental cost as they offer a computationally less expensive way of evaluating designs. It is important to ensure the accuracy of PRSs before using them for design and optimization. The accuracy of a PRS, constructed using a limited number of simulations, is primarily affected by two factors: (1) noise in the data; and (2) inadequacy of the fitting model (called modeling error or bias error). In experiments, noise may appear due to measurement errors and other experimental errors. Numerical noise in computer simulations is usually small, but it can be high for ill-conditioned problems, or if there are some unconverged solutions such as those encountered in computational fluid dynamics or structural optimization. The true model representing the data is rarely known, and due to limited data available, usually a simple model is fitted to the data. For simulation-based PRS, modeling/bias error due to an inadequate model is mainly responsible for the error in prediction.

In design of experiments techniques, sampling of the points in design space seeks to reduce the effect of noise and reduce bias errors simultaneously. However, these objectives (noise and bias error) often conflict. For example, noise rejection criteria, such as D-optimality, usually produce designs with more points near the boundary, whereas the bias error criteria tend to distribute points more evenly in design space. Thus, the problem of selecting an experimental design (also commonly known as design of experiment or DOE) is a multi-objective problem with conflicting objectives (noise and bias error). The solution to this problem would be a Pareto optimal front of experimental designs that yields different tradeoffs between noise and bias errors. Seeking the optimal experimental designs considering only one criterion, though popular,

may yield minor improvements in the selected criterion with significant deterioration in other criteria.

In the past, the majority of the work related to the construction of experimental designs is done by considering only one design objective. When noise is the dominant source of error, there are a number of experimental designs that minimize the effect of variance (noise) on the resulting approximation, for example, the D-optimal design, that minimizes the variance associated with the estimates of coefficients of the response surface model. Traditional variance-based designs minimize the effect of noise and attempt to obtain uniformity (ratio of maximum to minimum error in design space) over the design space, but they do not address bias errors.

Classical minimum bias designs consider only space-averaged or integrated error measures (Myers and Montgomery, 1995, pp. 208-279) in experimental designs. The bias component of the averaged or integrated mean squared error is minimized to obtain so-called minimum bias designs. The fundamentals of minimizing integrated mean squared error and its components can be found in Myers and Montgomery (1995, Chapter 9), and Khuri and Cornell (1996, Chapter 6). Venter and Haftka (1997) developed an algorithm implementing a minimum-bias criterion for irregularly shaped design spaces where no closed form solution exists for experimental design. They compared minimum-bias and D-optimal experimental designs for two problems with two and three variables. The minimum-bias experimental design was found to be more accurate than D-optimal for average error but not for maximum error. Qu et al. (2004) implemented Gaussian quadrature-based minimum bias design and presented minimum bias central composite designs for up to six variables.

There is some work done on developing experimental designs by minimizing the integrated mean squared error accounting for both variance and bias errors. Box and Draper

(1963) minimized integrated mean squared errors averaged over the design space by combining average weighted variance and average bias errors. Draper and Lawrence (1965) minimized the integrated mean square error to account for model inadequacies. Kupper and Meydrecht (1973) specified bounds on the coefficients associated with the assumed true function to minimize integrated mean squared error. Welch (1983) used a linear combination of variance and bias errors to minimize mean squared error. Montepiedra and Fedorov (1997) investigated experimental designs minimizing the bias component of the integrated mean square error subject to a constraint on the variance component or *vice-versa*. Fedorov et al. (1999) later studied design of experiments *via* weighted regression prioritizing regions where the approximation is needed to predict the response. Their approach considered both variance and bias components of the estimation error.

Bias error averaged over the design space has been studied extensively, but there is a relatively small amount of work to account for pointwise variation of bias errors because of inherent difficulties. An approach for estimating bounds on bias errors in PRS by a pointwise decomposition of the mean squared error into variance and the square of bias was developed by Papila and Haftka (2001). They used the bounds to obtain experimental designs (EDs) that minimize the maximum absolute bias error. Papila et al. (2005) extended the approach to account for the data and proposed data-dependent bounds. They assumed that the true model is a higher degree polynomial than the approximating polynomial, and that it satisfies the given data exactly. Goel et al. (2006a) generalized this bias error bounds estimation method to account for inconsistencies between the assumed true model and actual data. They demonstrated that the bounds can be used to develop adaptive experimental designs to reduce the effect of bias errors in the region of interest. Recently, Goel et al. (2006c) presented a method to estimate pointwise

root mean square (RMS) bias errors in approximation prior to the generation of data. They applied this method to construct experimental designs that minimize maximum RMS bias error (min-max RMS bias designs).

Since minimum-bias designs do not achieve uniformity, designs that distribute points uniformly in design space (space filling designs like Latin hypercube sampling) are popular even though these designs have no claim to optimality. Since Latin hypercube sampling (LHS) designs can create poor designs, as illustrated by Leary et al. (2003), different criteria like, maximization of minimum distance between points, or minimization of correlation between points are used to improve its performance. We will demonstrate in this chapter that even optimized LHS designs can occasionally leave large holes in design space, which may lead to poor predictions. Thus, there is a need to consider multiple criteria. Some previous efforts of considering multiple criteria are as follows. In an effort to account for variance, Tang (1993) and Ye (1998) presented orthogonal array based LHS designs that were shown to be better than the conventional LHS designs. Leary et al. (2003) presented strategies to find optimal orthogonal array based LHS designs in a more efficient manner. Palmer and Tsui (2001) generated minimum-bias Latin hypercube experimental designs for sampling from deterministic simulations by minimizing integrated squared bias error. Combination of face-centered cubic design and LHS designs is quite widely used (Goel et al., 2006d).

The primary objective of this work is to demonstrate the risks associated with using a single criterion to construct experimental designs. Firstly, we compare LHS and D-optimal designs, and demonstrate that both these designs can leave large unsampled regions in design space that may potentially yield high errors. In addition, we illustrate the need to consider multiple criteria to construct experimental designs, as single-criterion based designs may

represent extreme tradeoffs among different criteria. Min-max RMS bias design, which yields a small reduction in the maximum bias error at the cost of a huge increase in the maximum variance, is used as an example. While the above issue of tradeoff among multiple criteria requires significant future research effort, we explore several strategies for the simultaneous use of multiple criteria to guard against selecting experimental designs that are optimal according to one criterion but may yield very poor performance on other criteria. In this context, we firstly discuss which criteria can be simultaneously used meaningfully; secondly, we explore how to combine different criteria. We show that complimentary criteria may cater to competing needs of experimental designs. Next, we demonstrate improvements by combining a geometry-based criterion LHS and a model-based D-optimality criterion, to obtain experimental designs. We also show that poor experimental designs can be filtered out by creating multiple experimental designs and selecting one of them using an appropriate error-based (pointwise) criterion. Finally, we combine the above mentioned strategies to construct experimental designs.

The chapter is organized as follows: Different error measures used in this study are summarized in the next section. Following that, we show major results of this study. We illustrate the issues associated with single criterion-based experimental designs, and show a few strategies to accommodate multiple criteria. We close the chapter by recapitulating major findings.

Error Measures for Experimental Designs

Let the true response $\eta(\mathbf{x})$ at a design point \mathbf{x} be represented by a polynomial $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$, where $\mathbf{f}(\mathbf{x})$ is the vector of basis functions and $\boldsymbol{\beta}$ is the vector of coefficients. The vector $\mathbf{f}(\mathbf{x})$ has two components: $\mathbf{f}^{(1)}(\mathbf{x})$ is the vector of basis functions used in the PRS or fitting model, and $\mathbf{f}^{(2)}(\mathbf{x})$ is the vector of additional basis functions that are missing in the linear regression model

(assuming that the true model is a polynomial). Similarly, the coefficient vector $\boldsymbol{\beta}$ can be written as a combination of vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ that represent the true coefficients associated with the basis function vectors $\mathbf{f}^{(1)}(\mathbf{x})$ and $\mathbf{f}^{(2)}(\mathbf{x})$, respectively. Precisely,

$$\eta(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} = \begin{bmatrix} \mathbf{f}^{(1)T} & \mathbf{f}^{(2)T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}^{(2)}. \quad (3.1)$$

Assuming normally distributed noise ε with zero mean and variance σ^2 ($N(0, \sigma^2)$), the observed response $y(\mathbf{x})$ at a design point \mathbf{x} is given as

$$y(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon. \quad (3.2)$$

The predicted response $\hat{y}(\mathbf{x})$ at a design point \mathbf{x} is given as a linear combination of approximating basis functions vector $\mathbf{f}^{(1)}(\mathbf{x})$ with corresponding estimated coefficient vector \mathbf{b} :

$$\hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}. \quad (3.3)$$

The estimated coefficient vector \mathbf{b} is evaluated using the data (\mathbf{y}) for N_s design points as (Myers and Montgomery, 1995, Chapter 2):

$$\mathbf{b} = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} \mathbf{y}, \quad (3.4)$$

where $X^{(1)}$ is the Gramian matrix constructed using $\mathbf{f}^{(1)}(\mathbf{x})$ (refer to Appendix A).

The error at a general design point \mathbf{x} is the difference between the true response and the predicted response, $e(\mathbf{x}) = \eta(\mathbf{x}) - \hat{y}(\mathbf{x})$. When noise is dominant, estimated standard error $e_{es}(\mathbf{x})$, used to appraise error, is given as (Myers and Montgomery, 1995)

$$e_{es}(\mathbf{x}) = \sqrt{\text{Var}[\hat{y}(\mathbf{x})]} = \sqrt{\sigma_a^2 \mathbf{f}^{(1)T}(\mathbf{x}) (X^{(1)T} X^{(1)})^{-1} \mathbf{f}^{(1)}(\mathbf{x})}, \quad (3.5)$$

where σ_a^2 is the estimated variance of the noise, and σ_a is the standard error in approximation.

When bias error is dominant, the root mean square of bias error $e_b^{rms}(\mathbf{x})$ at design point \mathbf{x} can be obtained as (Goel et al. 2006c, and Appendix A)

$$e_b^{rms}(\mathbf{x}) = \sqrt{E_\beta(e_b^2(\mathbf{x}))} \approx \sqrt{[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] E_\beta(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)T}) [\mathbf{f}^{(2)}(\mathbf{x}) - A^T\mathbf{f}^{(1)}(\mathbf{x})]}, \quad (3.6)$$

where $E_\beta(g(x))$ is the expected value of $g(x)$ with respect to β , and A is the alias matrix

$A = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} X^{(2)}$. However, the true model may not be known, and Equation (3.6) is only satisfied approximately for the assumed true model.

Prior to generation of data, all components of the coefficient vector $\boldsymbol{\beta}^{(2)}$ are assumed to have a uniform distribution between $-\gamma$ and γ (γ is a constant) such that $E_\beta(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)T}) = \gamma^2/3 I$, where I is an $N_{\beta_2} \times N_{\beta_2}$ identity matrix. Substituting this in Equation (3.6), the pointwise RMS bias error (Goel et al. 2006c, and Appendix A) is

$$\begin{aligned} e_b^{rms}(\mathbf{x}) &\approx \sqrt{[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] \left(\gamma^2/3 I \right) [\mathbf{f}^{(2)}(\mathbf{x}) - A^T\mathbf{f}^{(1)}(\mathbf{x})]} \\ &= \gamma/\sqrt{3} \sqrt{[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] [\mathbf{f}^{(2)}(\mathbf{x}) - A^T\mathbf{f}^{(1)}(\mathbf{x})]}. \end{aligned} \quad (3.7)$$

Since γ has a uniform scaling effect, prior to generation of data it can be taken as unity for the sake of simplicity. It is clear from Equation (3.7) that the RMS bias error at a design point \mathbf{x} can be obtained from the location of data points (defines alias matrix A) used to fit the response surface, the form ($\mathbf{f}^{(1)}$ and $\mathbf{f}^{(2)}$) of the assumed true function (which is a higher order polynomial than the approximating polynomial), and the constant γ . Goel et al. (2006c) demonstrated with examples that this error prediction model gives good estimates of actual error fields both when the true function is polynomial and when it is non-polynomial. Two representative examples of a

polynomial true function and a non-polynomial true function (trigonometric function with multiple frequencies) are presented in Appendix B, respectively.

Many criteria used to construct/compare different EDs are presented in the literature. A few commonly used error metrics as well as new bias error metrics are listed below.

- Maximum estimated standard error in design space (Equation (3.5) with $\sigma_a = 1$)

$$(e_{es})_{\max} = \max_V (e_{es}(\mathbf{x})). \quad (3.8)$$

- Space-averaged estimated standard error (Equation (3.5) with $\sigma_a = 1$)

$$(e_{es})_{avg} = \frac{1}{vol(V)} \int_V e_{es}(\mathbf{x}) d\mathbf{x}. \quad (3.9)$$

- Maximum absolute bias error bound (Papila et al., 2005) in design space

$$(e_b^I)_{\max} = \max_V (e_b^I(\mathbf{x})), \quad (3.10)$$

where $\mathbf{c}^{(2)} = 1$ and

$$(e_b^I)_{\max} = \sum_{j=1}^{N_{\beta_2}} \left| f_j^{(2)}(\mathbf{x}) - \sum_{i=1}^{N_{\beta_1}} A_{ij} f_i^{(1)}(\mathbf{x}) \right| \mathbf{c}_j^{(2)}. \quad (3.11)$$

- Maximum RMS bias error in design space (Equation (3.7) with $\gamma = 1$)

$$(e_b^{rms})_{\max} = \max_V (e_b^{rms}(\mathbf{x})). \quad (3.12)$$

- Space-averaged RMS bias error (Equation (3.7) with $\gamma = 1$)

$$(e_b^{rms})_{avg} = \frac{1}{vol(V)} \int_V e_b^{rms}(\mathbf{x}) d\mathbf{x}. \quad (3.13)$$

This criterion is the same as space-averaged bias error.

Among all the above criteria, the standard error based criteria are the most commonly used.

For all test problems, a design space coded as an N_v -dimensional cube $V = [-1, 1]^{N_v}$ is used and

bias errors are computed following the common assumption that the true function and the response surface model are cubic and quadratic polynomials, respectively.

Besides the above error-metric based criteria, the following criteria are also frequently used.

- D-efficiency (Myers and Montgomery, 1995, pp. 393)

$$D_{eff} = \left(\frac{|M|}{\max |M|} \right)^{1/N_{\beta_1}} ; \quad |M| = \frac{|X^{(1)T} X^{(1)}|}{N_s^{N_{\beta_1}}} . \quad (3.14)$$

Here, $\max |M|$ in Equation (3.14) is taken as the maximum of all experimental designs.

This criterion is primarily used to construct D-optimal designs. A high value of D-efficiency is desired to minimize the variance of the estimated coefficients \mathbf{b} .

- Radius of the largest unoccupied sphere (r_{max})

We approximate the radius of the largest sphere that can be placed in the design space such that there are no experimental design points inside this sphere. A large value of r_{max} indicates large holes in the design space and hence a potentially poor experimental design. This criterion is not used to construct experimental designs, but this allows us to measure the space-filling capability of any experimental design.

Test Problems and Results

This section is divided into two parts. In the first subsection, we compare widely used experimental designs, like LHS designs, D-optimal designs, central composite designs and their minimum bias error counterparts. We show that different designs offer tradeoffs among multiple criteria and experimental designs based on a single error criterion may be susceptible to high errors on other criteria. In the second subsection, we discuss a few possible strategies to simultaneously accommodate multiple criteria. Specifically, we present two strategies, (1)

combination of a geometry-based criterion (LHS) with a model-based criterion (D-optimality), and (2) simultaneous use of multiple experimental designs combined with pointwise error estimates as a filtering criterion to seek protection against poor designs.

Comparison of Different Experimental Designs

Space filling characteristics of D-optimal and LHS designs

Popular experimental designs, like LHS designs that cater to bias errors by evenly distributing points in design space or numerically obtained D-optimal designs that reduce the effect of noise by placing the design points as far apart as possible, can occasionally leave large holes in the design space due to the random nature of the design (D-optimal) or due to convergence to local optimized LHS designs. This may lead to poor approximation. Firstly, we demonstrate that for D-optimal and optimized LHS designs, a large portion of design space may be left unsampled even for moderate dimensional spaces. For demonstration, we consider two- to four-dimensional spaces $V = [-1,1]^{N_v}$. The number of points in each experimental design is twice the number of coefficients in the corresponding quadratic polynomial, that is, 12 points for two dimensions, 20 points for three dimensions, and 30 points for four-dimensional design spaces. We also create experimental designs with 40 points in four-dimensional design space. We generate 100 designs in each group to alleviate the effect of randomness.

D-optimal designs were obtained using the MATLAB routine ‘candexch’ such that duplicate points are not allowed (duplicate points are not useful to approximate data from deterministic functions or numerical simulations). We supplied a grid of points (grid includes corner/face points and points sampled at a grid spacing randomly selected between 0.15 and 0.30 units) and allocated a maximum of 50000 iterations to find a D-optimal design. LHS designs

were constructed using the MATLAB routine ‘lhsdesign’ with the ‘maximin’ criterion that maximizes the minimum distance between points. We allocated 40 iterations for optimization.

For each experimental design, we estimated the radius (r_{max}) of the largest unsampled sphere that fits inside the design space and summarized the results with the help of boxplots in Figure 3-1. The box encompasses the 25th to 75th percentiles and the horizontal line within the box shows the median value. The notches near the median represent the 95% confidence interval of the median value. It is obvious from Figure 3-1 that r_{max} increases with the dimensionality of the problem, i.e., the distribution of points in high dimensional spaces tends to be sparse. As expected, an increase in the density of points reduced r_{max} (compare four-dimensional space with 30- and 40- points). The reduction in r_{max} was more pronounced for D-optimal designs than LHS designs. LHS designs had a less sparse distribution compared to D-optimal designs, however, the median r_{max} of approximately 0.75 units in four-dimensional space for LHS designs indicated that a very large region in the design space remained unsampled and data points are quite far from each other.

The sparse distribution of points in the design space is illustrated with the help of a three-dimensional example with 20 points in Figure 3-2, where the largest unsampled sphere is shown. For both D-optimal and LHS designs, the large size of the sphere clearly demonstrates the presence of large gaps in the design space that makes the surrogate predictions susceptible to errors. This problem is expected to become more severe in high dimensional spaces. The results indicate that a single criterion (D-optimality for D-optimal designs, and max-min distance for LHS designs) based experimental design may lead to poor performance on other criteria.

Tradeoffs among various experimental designs

Next, we illustrate tradeoffs among different experimental designs by comparing min-max RMS bias design (refer to Appendix B), face-centered cubic design (FCCD), D-optimal design

(obtained using JMP, Table 3-1), and LHS design (generated using MATLAB routine ‘lhsdesign’ with ‘maximin’ criterion, and allocating 1000 iterations to get a design, Table 3-2) in four-dimensional space. Note that all experimental designs, except FCCD, optimize a single criterion, i.e., D-optimal designs optimize D-efficiency, LHS designs maximize the minimum distance between points, and min-max RMS bias designs minimize the influence of bias errors. On the other hand, FCCD is an intuitive design obtained by placing the points on the faces and vertices.

The designs were tested using a uniform 11^4 grid in the design space $V = [-1, 1]^4$ and different metrics are documented in Table 3-3. We observed that no single design (used in the generic sense, meaning a class of designs) outperformed other designs on ‘all’ criteria. The D-optimal and the face-centered cubic design had high D-efficiency; the min-max RMS bias design and the LHS design had low D-efficiency. The min-max RMS bias design performed well on bias error based criteria but caused a significant deterioration in standard error based criteria, due to the peculiar placement of axial points near the center. While the D-optimal design was a good experimental design according to standard error based criteria, large holes in the design space ($r_{max} = 1$) led to poor performance on bias error based criteria. Since LHS designs neglect boundaries, they resulted in very high maximum standard error and bias errors. However, LHS designs yielded the least space-averaged bias error estimate. The FCCD design, which does not optimize any criterion, performed reasonably on all the metrics. However, we note that FCCD designs in high dimensional spaces are not practical due to the high ratio of the number of simulations to the number of polynomial coefficients.

We used polynomial examples to illustrate the risks in using experimental designs constructed with a single criterion. Firstly, we considered a simple quadratic function $F_1(\mathbf{x})$

(Equation (3.15) in four-dimensional space) with normally distributed noise ε (zero mean and unit variance),

$$\begin{aligned} F_1(\mathbf{x}) &= \eta(\mathbf{x}) + \varepsilon, \\ \eta(\mathbf{x}) &= 10(1 + x_1 + x_3 + x_2^2 + x_4^2). \end{aligned} \tag{3.15}$$

We construct a quadratic polynomial response surface approximation using the data at 25 points sampled with different experimental designs (min-max RMS bias design; FCCD; D-optimal, Table 3-1; and LHS, Table 3-2) and compute actual absolute error in approximation at a uniform grid of 11^4 points in the design space $V = [-1, 1]^4$. The accuracy measures based on the data, that are most commonly used, are the adjusted coefficient of determination R_{adj}^2 and the standard error normalized by the range of the function (RMSE) in Table 3-4. Very high values of normalized maximum, root mean square, and average absolute errors (normalized by the range of the data for the respective experimental design) in Table 3-4 indicate that the min-max RMS bias design (also referred to as RMS bias CCD) is indeed a poor choice of experimental design when the error is due to noise, though all approximation accuracy measures (R_{adj}^2 , standard error) suggested otherwise. That is, the high errors come with no warning from the fitting process! High values of the ratio of space-averaged, root mean square, or maximum actual errors to standard error indicate the risks associated with relying on measures such as R_{adj}^2 to determine the accuracy of approximations (We pursue this issue in detail in a Chapter 5). Among other experimental designs, LHS design had a high normalized maximum error near the corners, where no data is sampled. FCCD and D-optimal designs performed reasonably, with FCCD being the best design.

Secondly, we illustrate errors due to large holes in design space observed in the previous section. A simple function that is likely to produce large errors would be a cosine with maximum

at the center of the sphere. However, to ensure a reasonable approximation of the true function with polynomials, we used a truncated Maclaurin series expansion of a translated radial cosine function $\cos(k\|\mathbf{x}-\mathbf{x}_{lhs}^c\|)$, namely

$$F(\mathbf{x}) = 20\left(1 - \frac{r^2}{2} + \frac{r^4}{4!}\right); \quad r = k\|\mathbf{x}-\mathbf{x}_{ed}^c\|, \quad (3.16)$$

where \mathbf{x}_{ed}^c is a fixed point in design space, and k is a constant. We considered two instances of Equation (3.16) by specifying the center of the largest unoccupied sphere associated with LHS design (Table 3-2) and D-optimal design (Table 3-1) as the fixed point.

$$F_2(\mathbf{x}) = 20\left(1 - \frac{r^2}{2} + \frac{r^4}{4!}\right); r = k_2\|\mathbf{x}-\mathbf{x}_{lhs}^c\|, \mathbf{x}_{lhs}^c = [-0.168, -0.168, -0.141, 0.167], \quad (3.17)$$

$$F_3(\mathbf{x}) = 20\left(1 - \frac{r^2}{2} + \frac{r^4}{4!}\right); r = k_3\|\mathbf{x}-\mathbf{x}_{dopt}^c\|, \mathbf{x}_{dopt}^c = [0.0, 0.0, 0.0, 0.0], \quad (3.18)$$

The constants k_2 and k_3 were obtained by maximizing the normalized maximum actual absolute error in approximation of $F_2(\mathbf{x})$ using LHS experimental design and approximation of $F_3(\mathbf{x})$ using D-optimal experimental design, respectively, subject to a reasonable approximation (determined by the condition, $R_{adj}^2 \geq 0.90$) of the two functions by all considered experimental designs (FCCD, D-optimal, LHS, and RMS bias CCD). As earlier, the errors were normalized by dividing the actual absolute errors by the range of data values used to construct the experimental design. Subsequently, the optimal value of the constants were $k_2 = 1.13$ and $k_3 = 1.18$.

We used quadratic polynomials to approximate $F(\mathbf{x})$ and errors are evaluated at a uniform grid of 11^4 points in the design space $V = [-1, 1]^4$. The quality of fit, maximum, root mean square, and average actual absolute errors in approximation for each experimental design are summarized in Table 3-4. We observed that despite a good quality of fit (high R_{adj}^2 and low

normalized standard error), the normalized maximum actual absolute errors were high for all experimental designs. In particular, the approximations constructed using data sampled at the D-optimal and the LHS designs performed very poorly. This means that the accuracy metrics, though widely trusted, can mislead the actual performance of the experimental design. The high maximum error in approximations using the LHS designs occurred at one of the corners that was not sampled (thus extrapolation error), however, we note that LHS designs yielded the smallest normalized space-averaged and root mean square error in approximation. On the other hand, the maximum error in approximations using D-optimal designs appeared at a test point closest to the center \mathbf{x}_{lhs}^c in the case of $F_2(\mathbf{x})$, and near \mathbf{x}_{dopt}^c in the case of $F_3(\mathbf{x})$. Besides, high normalized space-averaged errors indicated poor approximation of the true function $F(\mathbf{x})$. The other two experimental designs, FCCD and RMS bias CCD, performed reasonably on maximal errors. The relatively poor performance of RMS bias CCD for average and RMS errors is explained by recalling that the experimental design was constructed by assuming the true function to be a cubic polynomial, whereas $F(\mathbf{x})$ was a quartic polynomial.

An important characteristic for all experimental designs is the ratio of space-averaged, root mean square, or maximum actual absolute error to estimated standard error. When this ratio is large, the errors are unexpected and therefore, potentially damaging. The FCCD design provided a reasonable value of the ratio of actual to estimated standard errors, however, RMS bias design performed very poorly as the actual errors were much higher than the standard estimated error. This means that the estimated standard error is misleading about the actual magnitude of error that cannot be detected in an engineering example where we do not have the luxury of using a large number of data points to test the accuracy of approximation. Similarly, for all functions, the ratio of maximum actual absolute error to standard error for LHS designs (29–52) was much

higher than for D-optimal designs (about 9). The surprise element is also evident by the excellent values of R_{adj}^2 of 0.99 and 1.00 compared to 0.90 for the D-optimal design.

The results presented here clearly suggest that different experimental designs were non-dominated with respect to each other and offered multiple (sometimes extreme) tradeoffs, and that it might be dangerous to use a single criterion based experimental design without thorough knowledge of the problem (which is rarely the case in practice).

Extreme example of risks in single criterion based design: Min-max RMS bias CCD

A more pathological case, demonstrating the risks in developing experimental designs using a single criterion, was encountered for moderate dimensional cases while developing the central composite design counterpart of the minimum bias design, i.e., minimizing the maximum RMS bias error. The performance of the min-max RMS bias designs constructed using two parameters (refer to Appendix B) for two- to five-dimensional spaces on different metrics is given in Table 3-5. For two- and three-dimensional spaces, the axial points (given by α_2) were located at the face and the vertex points (given by α_1) were placed slightly inwards to minimize the maximum RMS bias errors. The RMS bias designs performed very reasonably on all the error metrics. A surprising result was obtained for optimal designs for four- and five-dimensional spaces: while the parameter corresponding to vertex points (α_1) was at its upper limit (1.0), the parameter corresponding to the location of axial points (α_2) hit the corresponding lower bound (0.1). This meant that to minimize maximum RMS bias error, the points should be placed near the center. The estimated standard error was expectedly very high for this design. Contrasting face-centered cubic design for four-dimensional cases with three- and four-dimensional min-max RMS bias designs (Table 3-5) isolated the effect of dimensionality and the change in experimental design (location of axial points) on different error metrics. The increase in bias

errors (bounds and RMS error) was attributed to increase in dimensionality (small variation in bias errors with different experimental designs in four-dimensional design space), and the increase in standard error for min-max RMS bias design was the outcome of the change in experimental design (the location of axial points given by α_2). This unexpected result for four- and higher dimensional cases is supported by theoretical reasoning (Appendix B), and very strong agreement between the predicted and the actual RMS bias errors for the min-max RMS bias design and the face-centered central composite design (Appendix B).

To further illustrate the severity of the risks in using a single criterion, we show the tradeoffs among the maximum errors (RMS bias error, estimated standard error, and bias error bound) for a four-dimensional design space $[-1,1]^4$, obtained by varying the location of the axial points (α_2) from near the center ($\alpha_2=0.1$, min-max RMS bias design) to the face of the design space ($\alpha_2=1.0$, central composite design), while keeping the vertex locations ($\alpha_1=1.0$) fixed. The tradeoff between maximum RMS bias error and maximum estimated standard error is shown in Figure 3-3(A), and the tradeoff between maximum RMS bias error and maximum bias error bound is shown in Figure 3-3(B). Moving the axial point away from the center reduced the maximum bias error bound and the maximum estimated standard error but increased the maximum RMS bias error. The relatively small variation in maximum RMS bias error compared to the variation in maximum estimated standard error and maximum bias error bound demonstrated the low sensitivity of maximum RMS bias error with respect to the location of axial points (α_2), and explains the success of the popular central composite designs ($\alpha_2=1.0$) in handling problems with bias errors. While we noted that each design on the curves in Figure 3-3 corresponds to a non-dominated (tradeoff) point, a small increase in maximum RMS bias error permits a large reduction in maximum estimated standard error, or in other words, the

minimization with respect to a single criterion (here maximum RMS bias error) may lead to small gains at a cost of significant loss with respect to other criteria. Tradeoff between maximum bias error bound and maximum RMS bias error also reflected similar results, though the gradients were relatively small.

The most important implication of the results presented in this section is that it may not be wise to select experimental designs based on a single criterion. Instead, tradeoff between different metrics should be explored to find a reasonable experimental design. While detailed exploration of this issue requires significantly more research, our initial attempts to simultaneously accommodate multiple criteria are illustrated next.

Strategies to Address Multiple Criteria for Experimental Designs

As discussed in the previous section, the experimental designs optimized using a single criterion may perform poorly on other criteria. While a bad experimental design can be identified by visual inspection in low dimensional spaces, we need additional measures to filter out bad designs in high dimensional spaces (Goel et al., 2006e). We explored several strategies to simultaneously accommodate multiple criteria in an attempt to avoid poor experimental designs. In this context, we discuss two issues:

- Which criteria are meaningful for different experimental designs? and
- How can we combine different criteria?

Since the experimental designs are constructed to minimize the influence of bias error and noise, a sensible choice of suitable criteria for any experimental design should seek balance among the two sources of errors, i.e., bias and noise. Consequently, if we select an experimental design that primarily caters to one source of error, for example, noise, the secondary criterion should be introduced to address the other source of error, bias error in this case, and *vice-versa*. We elaborate on this idea in a following subsection.

Once we have identified criteria to construct experimental designs, we seek ways to combine different criteria. Taking inspiration from multi-objective optimization problems, we can accommodate multiple criteria according to several methods, for example,

- Optimize the experimental design to minimize a composite function that represents the weighted sum of criteria,
- Optimize the experimental design to minimize the primary criterion while satisfying constraints on the secondary criteria, and
- Solve a multi-objective optimization problem to identify different tradeoffs and then select a design that suits the requirements the most.

Here, we show two ways to avoid poor experimental designs using a four-dimensional example. Firstly, we present a method to combine the model-based D-optimality criterion that caters to noise with the geometry-based LHS criterion that distributes points evenly in design space and reduces space-averaged bias errors. Secondly, we demonstrate that selecting one out of several experimental designs according to an appropriate pointwise error-based criterion reduces the risk of obtaining poor experimental designs. Further, we show that the coupling of multiple criteria and multiple experimental designs may be effective to avoid poor designs.

Combination of model-based D-optimality criterion with geometry based LHS criterion

We used an example of constructing an experimental design for a four-dimensional problem with 30 points (response surface model and assumed true model were quadratic and cubic polynomials, respectively). Three sets of experimental designs were generated as follows. The first set comprised 100 LHS experimental designs generated using the MATLAB routine ‘lhsdesign’ with the ‘maximin’ criterion (a maximum of 40 iterations were assigned to find an experimental design). The second set comprised 100 D-optimal experimental designs generated using the MATLAB routine ‘candexch’ with a maximum of 40 iterations for optimization. A grid of points, (with grid spacing randomly selected between 0.15 and 0.30) including face and

corner points, was used to search for the D-optimal experimental designs. The third set of (combination) experimental designs was obtained by combining D-optimal (model-based criterion) and LHS designs (geometry-based criterion). We selected 30 design points from a 650 point² LHS design ('lhsdesign' with 'maximin' criterion and a maximum of 100 'iterations' for optimization) using the D-optimality criterion ('candexch' with a maximum of 50000 iterations for optimization). For each design, we computed the radius r_{max} of the largest unsampled sphere, D-efficiency, maximum and space-averaged RMS bias and estimated standard error using a uniform 11^4 grid in the design space $[-1,1]^4$.

We show the tradeoff among different criteria for D-optimal, LHS, and combination designs in Figure 3-4. As can be seen from Figure 3-4(A), the D-optimal designs were the best and LHS designs were the worst with respect to the maximum estimated standard and RMS bias error. Compared to the LHS designs, the combination designs significantly reduced the maximum estimated standard error with marginal improvement on the maximum RMS bias error criterion (Figure 3-4(A)), and improved D-efficiency without sacrificing r_{max} (Figure 3-4(D)). The advantages of using combination designs were more obvious in Figure 3-4(B), where we compared space-averaged bias and estimated standard errors. We see that D-optimal designs performed well on space-averaged estimated standard errors but yielded high space-averaged RMS bias errors. On the other hand, the LHS designs had low space-averaged RMS bias errors but high space-averaged estimated standard errors. The combination designs simultaneously yielded low space-averaged RMS bias and estimated standard errors. This result was expected because the Latin hypercube sampling criterion allows relatively uniform distribution of the

² The average number of points in the uniform grid used to generate D-optimal designs was 1300. So to provide a fair comparison while keeping the computational cost low, we obtain 650 points using LHS and use this set of points to develop combination experimental designs.

points by constraining the location of points that are used to generate combination designs using the D-optimality criterion. Similarly, we observed that unlike D-optimal designs, combination experimental designs performed very well on the space-averaged RMS bias error and the r_{max} criterion (refer to Figure 3-4(C)), and the performance was comparable to that of the LHS designs.

Mean and coefficient of variation (COV) of different metrics for the three sets of experimental designs are tabulated in Table 3-6. D-optimal designs outperformed LHS designs in terms of the ratio of maximum to average error (stability), D-efficiency, maximum RMS bias error, and maximum estimated standard error. Also, for most metrics, the variation in results due to sampling (COV) was the least among the three. As seen before, LHS designs performed the best on space-averaged RMS bias errors. The designs obtained by combining two criteria (D-optimality and LHS), were substantially closer to the best of the two except for $(e_b^{rms})_{max}$. Thus, they reduced the risk of large errors. Furthermore, the variation with samples (COV) is also reduced. The results suggested that though different experimental designs were non-dominated (tradeoffs) with respect to each other, simultaneously considering multiple criteria by combining the model-based D-optimality criterion and the geometry-based LHS criterion may be effective in producing more robust experimental designs with a reasonable tradeoff between bias errors and noise.

Multiple experimental designs combined with pointwise error-based filtering

Next, we demonstrate the potential of using multiple experimental designs to reduce the risk of finding poor experimental designs. The main motivation is that the cost of generating experimental designs is not high so we can construct two or three experimental designs using LHS, or D-optimality, or a combination of the two criteria, and pick the best according to an

appropriate criterion. To illustrate the improvements by using three EDs over a single ED, each of the two criteria—maximum RMS bias error and maximum estimated standard error—were used to select the best (least error) of the three EDs. For illustration, 100 such experiments were conducted with LHS designs, D-optimal designs, and the combination of LHS and D-optimal designs (as described above).

Actual magnitudes of maximum RMS bias error and maximum estimated standard error for all 300 designs and the 100 designs obtained after filtering using min-max RMS bias or maximum estimated standard error criteria are plotted in Figure 3-5 for three sets of (100) experimental designs. As is evident by the shortening of the upper tail and the size of the boxplots in Figure 3-5, both the min-max RMS bias and maximum estimated standard error criteria helped eliminate poor experimental designs for all three sets. Filtering had more impact on the maximum error estimates than the space-averaged error estimates. The numerical quantification of improvements in actual magnitude of maximum and space-averaged error based on 100 experiments is summarized in Table 3-7. We observed that the pointwise error-based (min-max RMS bias or estimated standard error) filtering significantly reduced the mean and COV of maximum errors. We also noted improvements in the individual experimental designs using multiple criteria. LHS designs were most significantly improved by picking the best of three based on estimated maximum standard error. D-optimal designs combined with the min-max RMS bias error based filtering criterion helped eliminate poor designs according to the RMS bias error criterion. It can be concluded from this exercise that potentially poor designs can be filtered out by considering a small number of experimental designs with an appropriate (min-max RMS bias or maximum estimated standard) error criterion. The filtering criterion should be complimentary to the criterion used for construction of the experimental design, i.e., if a group of

EDs are constructed using a variance based criterion, then the selection of an ED from the group should be based on bias error criterion, and vice-versa.

Results presented in this section indicate that use of multiple criteria (LHS and D-optimality) and multiple EDs help reduce maximum and space-averaged bias and estimated standard errors. Implementing the above findings, we can obtain experimental designs with reasonable tradeoff between bias error and noise in three steps as follows:

- Generate a large number of LHS experimental design points,
- Select a D-optimal subset within the LHS design (combine model-based and geometry-based criteria),
- Repeat first two steps three times and select the design that is the best according to one of the min-max RMS bias or maximum estimated standard error criteria (filtering using pointwise error-based criterion).

Concluding Remarks

In this chapter, we demonstrated the risks of using a single criterion to construct experimental designs. We showed that constructing experimental designs by combining multiple (model, geometry, and error based) criteria and/or using multiple experimental designs reduces the risk of using a poor experimental design.

For four-dimensional space, comparison of computed LHS and D-optimal designs, that involve random components and may yield poor approximation due to random components or convergence to local optima, revealed that the D-optimal designs were better for maximum errors, and LHS designs were better for space-averaged RMS bias errors. Both designs were susceptible to leaving large spheres in design space unsampled. A comparison of popular experimental designs (face-centered cubic design, min-max RMS bias design, D-optimal design, and LHS design) revealed the non-dominated (tradeoff among different criteria) nature of different designs. The min-max RMS bias design, obtained by placing the axial points close to

the center, performed the best in reducing maximum RMS bias error, but was the worst design for estimated standard error metrics and D-efficiency. LHS design gave the best performance in terms of space-averaged bias errors. However, face-centered cubic design that is an intuitive design yielded a reasonable tradeoff between bias error and noise reduction on all metrics. The same conclusions were supported by approximation of three example polynomials that highlighted the susceptibility of different experimental designs to the nature of the problem, despite the fact that the accuracy metrics suggested a very good fit for each example. We concluded that different experimental designs, constructed using one error criterion, do not perform the best on all criteria. Instead, they offer tradeoffs.

In moderate dimensional spaces these single criterion-based designs can often lead to extreme tradeoffs, particularly by using the maximum RMS bias error measure as a design criterion, such that small gains in the desired criterion are achieved at the cost of significant deterioration of performance in other criteria. A tradeoff study, conducted to study the variation of different error metrics with the location of axial points in central-composite designs, illustrated the perils of using a single criterion to construct experimental designs and emphasized the need to consider multiple criteria to tradeoff bias error and noise reduction.

To address the risk of using a poor experimental design by considering a single criterion, we explored a few strategies to accommodate multiple criteria. We demonstrated that the experimental design obtained by combining two criteria, the D-optimality criterion with LHS design, offered a reasonable tradeoff between space-averaged RMS bias and estimated standard error, and space-filling criteria. Specifically, combination designs significantly improved the poor experimental designs. We showed that the risk of getting a poor experimental design could be further reduced by choosing one out of three experimental designs using a pointwise error-

based criterion, e.g., min-max RMS bias or maximum estimated standard error criterion. The combination of D-optimal designs and min-max RMS bias error was particularly helpful in reducing bias errors. Finally, we adopted selection of experimental designs by combining the D-optimality criterion with LHS design and selecting one out of three such combination designs to cater to both bias error and noise reduction. However, since these results are based on a limited number of examples, we note the need of future research to address the issues related to accommodating multiple criteria while constructing experimental designs.

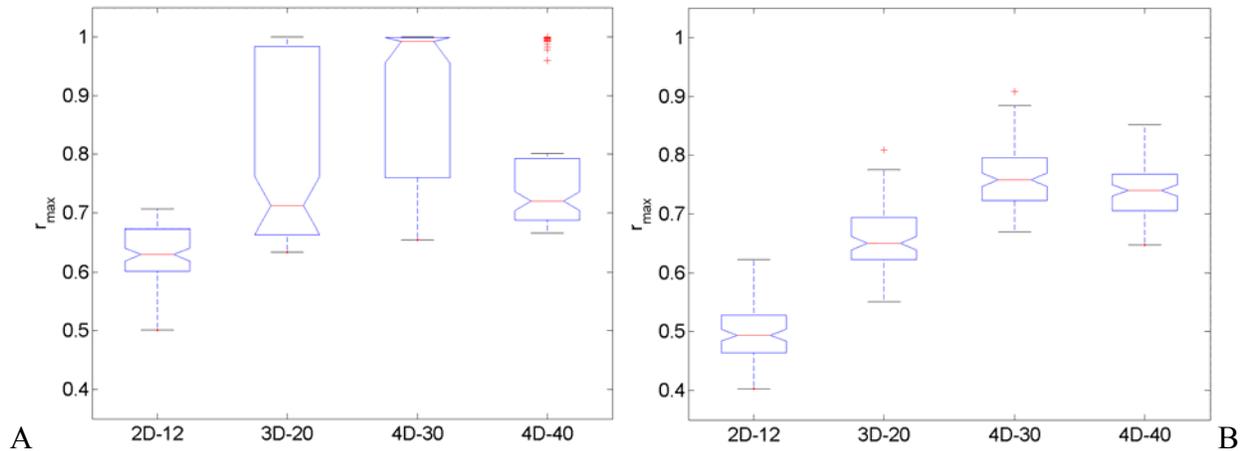


Figure 3-1. Boxplots (based on 100 designs) of radius (r_{max}) of the largest unoccupied sphere inside the design space $[-1, 1]^{N_v}$ (where N_v is the number of variables). x -axis shows the dimensionality of the design space and corresponding number of points in the experimental design. Smaller r_{max} is desired to avoid large unsampled regions. D-optimal designs are selected using MATLAB routine ‘candexch’ (we specify a grid of points with grid spacing between 0.15 and 0.30, and a maximum of 50000 iterations for optimization). LHS designs are generated using MATLAB routine ‘lhsdesign’ with a maximum of 40 iterations for optimization. A) D-optimal designs. B) LHS designs.

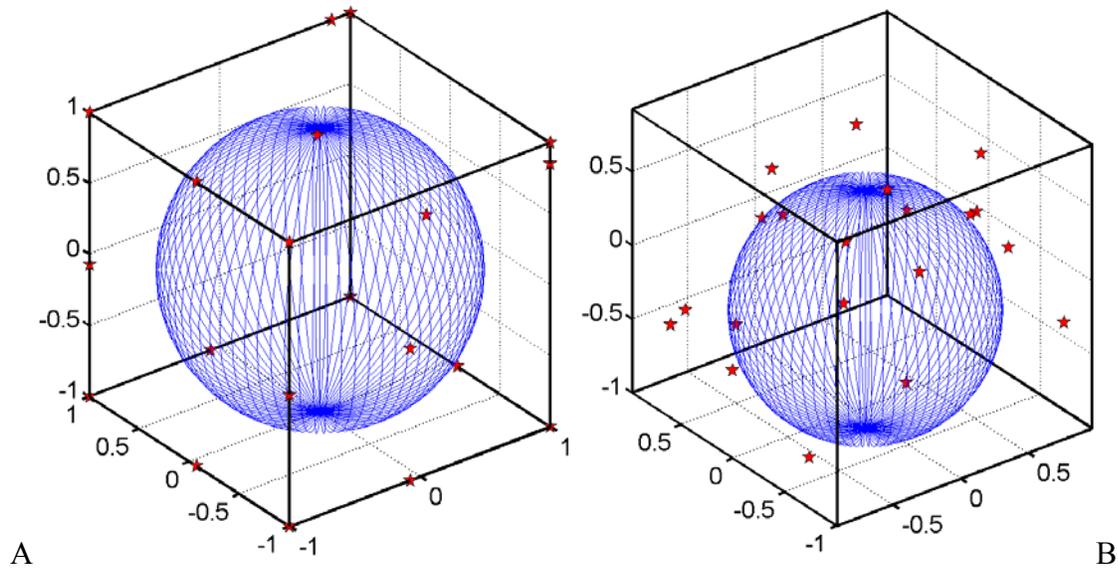


Figure 3-2. Illustration of the largest spherical empty space inside the 3D design space $[-1, 1]^3$ (20 points). A) D-optimal designs. B) LHS designs.

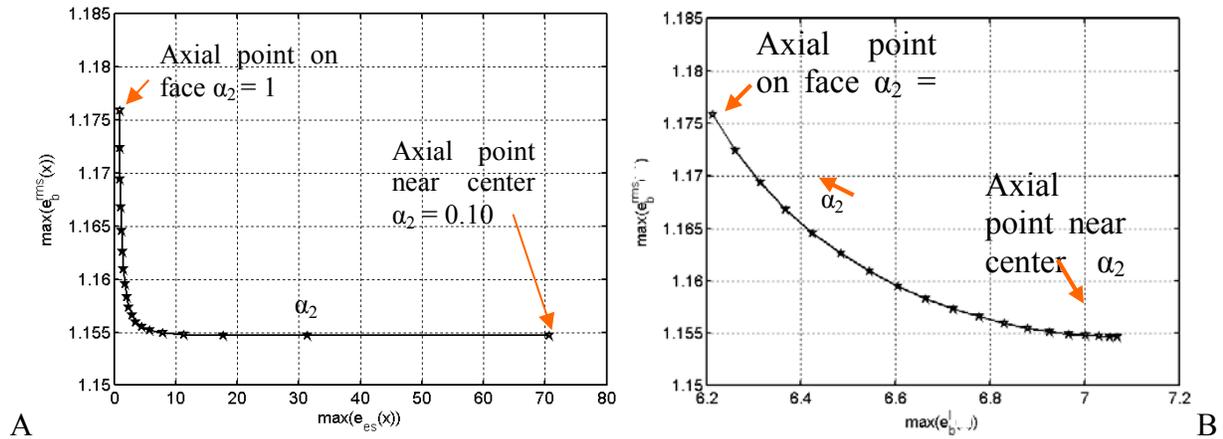


Figure 3-3. Tradeoffs between different error metrics. A) Maximum estimated standard error $(e_{es})_{\max}$ and maximum RMS bias error $(e_b^{rms})_{\max}$. B) Maximum bias error bound $(e_b^l)_{\max}$ and maximum RMS bias error $(e_b^{rms})_{\max}$ for four-dimensional space. 25 points were used to construct central composite experimental designs with vertex location fixed at $\alpha_1 = 1.0$.

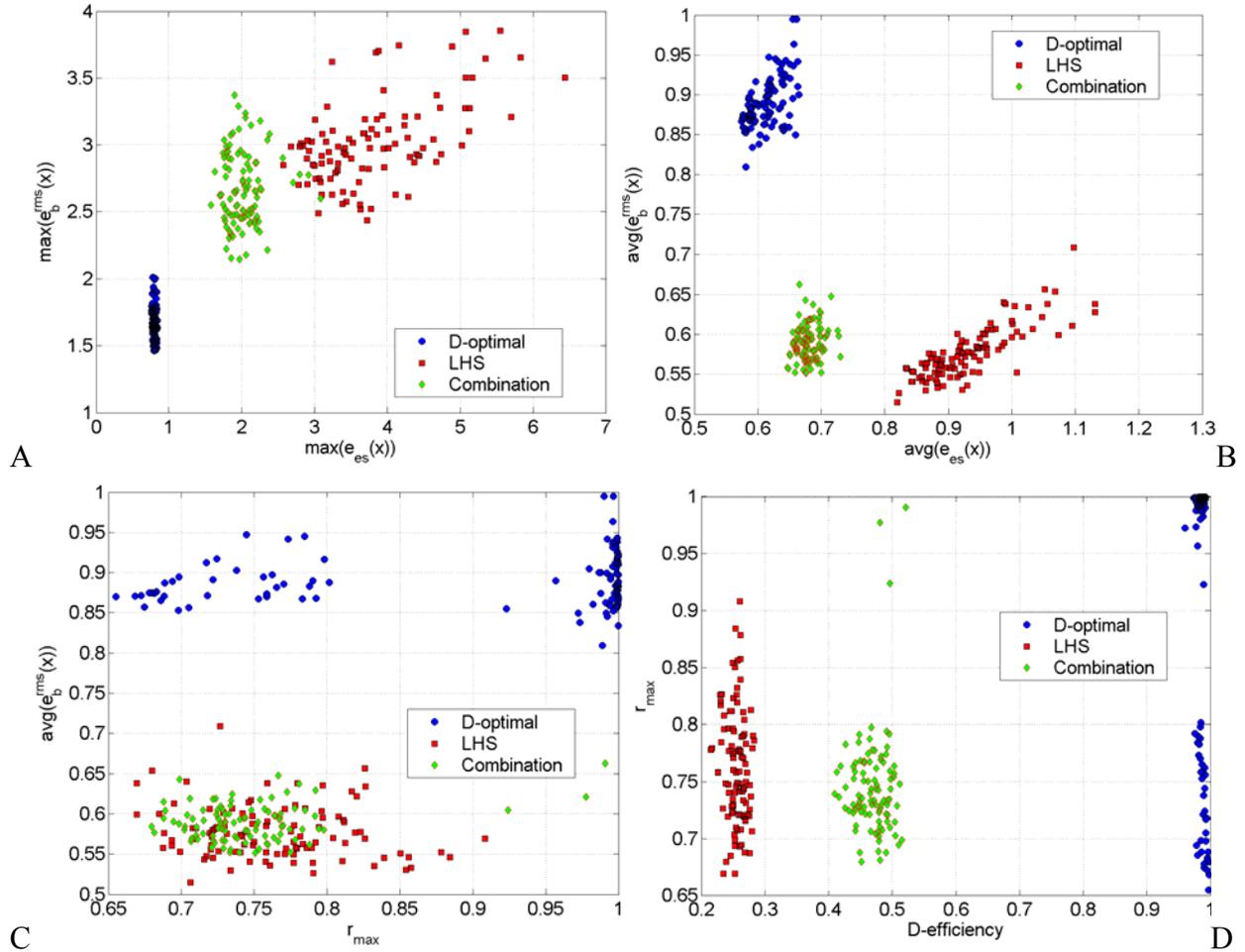


Figure 3-4. Comparison of 100 D-optimal, LHS, and combination (D-optimality + LHS) experimental designs in four-dimensional space (30 points) using different metrics. r_{\max} : radius of the largest unsampled sphere, $e_b^{rms}(\mathbf{x})$: RMS bias error, $e_{es}(\mathbf{x})$: estimated standard error, $(\cdot)_{\max}$: maximum of the quantity inside parenthesis, $(\cdot)_{\text{avg}}$: spatial average of the quantity inside parenthesis. All metrics, except D-efficiency (D_{eff}), are desired to be low. A) $(e_{es})_{\max}$ vs. $(e_b^{rms})_{\max}$. B) $(e_{es})_{\text{avg}}$ vs. $(e_b^{rms})_{\text{avg}}$. C) r_{\max} vs. $(e_b^{rms})_{\text{avg}}$. D) D_{eff} vs. r_{\max} .

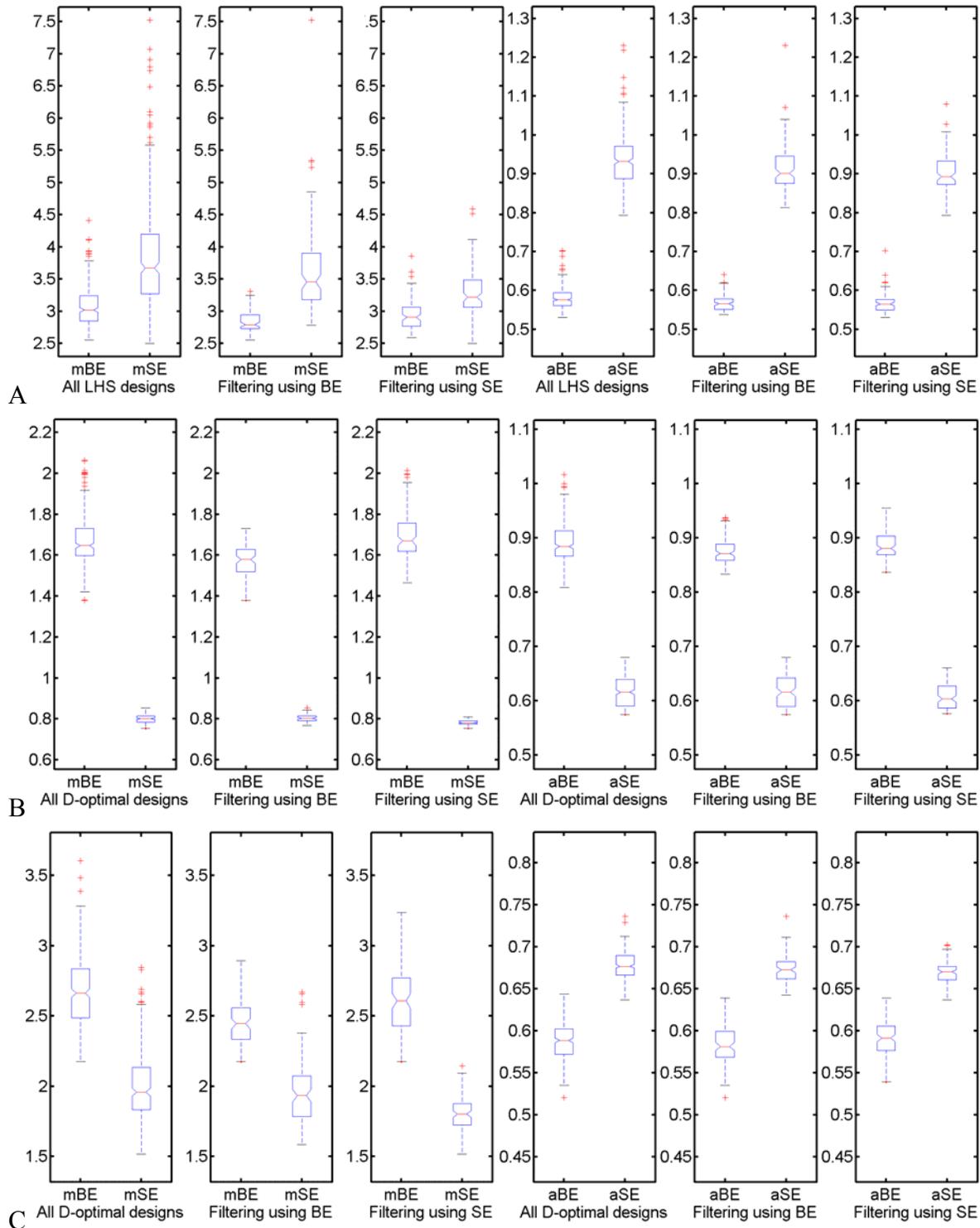


Figure 3-5. Simultaneous use of multiple experimental designs concept, where one out of three experimental designs is selected using appropriate criterion (filtering). Boxplots show maximum and space-averaged RMS bias, and standard estimated errors in four-dimensional design space; considering all 300 designs, 100 designs filtered using min-max RMS bias error as a criterion and 100 designs filtered using min max estimated standard error as a criterion (mBE denotes maximum RMS bias error, mSE

denotes maximum estimated standard error, aBE denotes space-averaged RMS bias error, aSE denotes space-averaged estimated standard error). Each experimental design has 30 points and errors are computed in design space $[-1, 1]^4$ on 11^4 uniform grid. A) LHS designs. B) D-optimal design. C) Combination designs.

Table 3-1. D-optimal design (25 points, 4-dimensional space) obtained using JMP.

| # no. | x_1 | x_2 | x_3 | x_4 | # no. | x_1 | x_2 | x_3 | x_4 |
|-----------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|
| 1 | -1 | -1 | -1 | -1 | 13 | 0 | 0 | -1 | -1 |
| 2 | -1 | -1 | -1 | 1 | 15 | 0 | 1 | 0 | 0 |
| 3 | -1 | -1 | 0 | -1 | 16 | 1 | -1 | -1 | -1 |
| 4 | -1 | -1 | 1 | 0 | 17 | 1 | -1 | -1 | 0 |
| 5 | -1 | -1 | 1 | 1 | 18 | 1 | -1 | 0 | 1 |
| 6 | -1 | 0 | 0 | 1 | 19 | 1 | -1 | 1 | -1 |
| 7 | -1 | 0 | 1 | -1 | 20 | 1 | -1 | 1 | 1 |
| 18 | -1 | 1 | -1 | -1 | 21 | 1 | 0 | 1 | 0 |
| 9 | -1 | 1 | -1 | 1 | 22 | 1 | 1 | -1 | -1 |
| 10 | -1 | 1 | 1 | -1 | 23 | 1 | 1 | -1 | 1 |
| 11 | -1 | 1 | 1 | 1 | 24 | 1 | 1 | 1 | -1 |
| 12 | 0 | -1 | -1 | 1 | 25 | 1 | 1 | 1 | 1 |
| 13 | 0 | -1 | 1 | -1 | | | | | |

Table 3-2. LHS designs (25 points, 4-dimensional space) obtained using the MATLAB routine 'lhsdesign' with 'maximin' criterion and a maximum of 1000 iterations for optimization.

| # no. | x_1 | x_2 | x_3 | x_4 | # no. | x_1 | x_2 | x_3 | x_4 |
|-----------|--------|--------|--------|--------|-----------|--------|--------|--------|--------|
| 1 | -0.970 | -0.344 | 0.055 | -0.809 | 13 | -0.717 | -0.812 | -0.125 | 0.396 |
| 2 | 0.679 | 0.123 | 0.709 | 0.520 | 15 | -0.376 | 0.986 | -0.686 | -0.463 |
| 3 | -0.093 | -0.233 | -0.929 | -0.659 | 16 | 0.002 | -0.361 | -0.101 | -0.936 |
| 4 | -0.578 | 0.659 | 0.943 | -0.531 | 17 | -0.464 | -0.862 | -0.589 | -0.256 |
| 5 | 0.798 | -0.657 | -0.829 | 0.355 | 18 | 0.972 | -0.547 | 0.424 | -0.411 |
| 6 | 0.328 | 0.282 | -0.619 | 0.051 | 19 | 0.689 | 0.557 | -0.324 | 0.979 |
| 7 | 0.908 | -0.059 | -0.207 | -0.714 | 20 | 0.591 | -0.175 | 0.010 | 0.471 |
| 8 | -0.142 | 0.027 | 0.879 | 0.271 | 21 | 0.065 | 0.457 | 0.255 | -0.281 |
| 9 | -0.679 | 0.427 | -0.507 | 0.700 | 22 | 0.160 | -0.935 | -0.901 | -0.075 |
| 10 | 0.427 | 0.265 | 0.552 | -0.900 | 23 | -0.839 | 0.702 | 0.329 | 0.618 |
| 11 | 0.470 | 0.090 | 0.604 | -0.160 | 24 | -0.916 | -0.691 | 0.777 | 0.808 |
| 12 | 0.226 | -0.474 | 0.504 | 0.006 | 25 | -0.327 | 0.806 | -0.408 | 0.142 |
| 13 | -0.224 | 0.873 | 0.150 | 0.872 | | | | | |

Table 3-3. Comparison of RMS bias CCD, FCCD, D-optimal, and LHS designs for 4-dimensional space (all designs have 25 points), errors are computed on a uniform 11^4 grid in space $V = [-1, 1]^4$. (**Table 3-1, *Table 3-2)

| Experimental design | D-efficiency | $(e_{es})_{\max}$ | $(e_{es})_{\text{avg}}$ | $ e_b^j _{\max}$ | $(e_b^{\text{rms}})_{\max}$ | $(e_b^{\text{rms}})_{\text{avg}}$ | r_{\max} |
|---------------------|--------------|-------------------|-------------------------|------------------|-----------------------------|-----------------------------------|------------|
| RMS bias CCD | 0.148 | 70.71 | 35.22 | 6.996 | 1.155 | 0.927 | 0.65 |
| FCCD | 0.932 | 0.877 | 0.585 | 6.208 | 1.176 | 0.827 | 0.67 |
| LHS* | 0.256 | 3.655 | 1.032 | 21.48 | 3.108 | 0.588 | 0.83 |
| D-optimal** | 1.000 | 0.933 | 0.710 | 12.00 | 1.996 | 1.004 | 1.00 |

Table 3-4. Prediction performance of different 25-point experimental designs in approximation of example functions F_1 (Equation (3.15)), F_2 (Equation (3.17)), and F_3 (Equation (3.18)) in four-dimensional spaces. We use quadratic polynomial response surface approximations and errors are computed on a uniform 11^4 grid in space $V = [-1, 1]^4$. R_{adj}^2 indicates the quality of approximation. The true standard error, space-averaged actual absolute error, root mean square error in space, and maximum actual absolute error are normalized by the range of the data at respective experimental designs. D-optimal design is shown in Table 3-1 and LHS design is given in Table 3-2.

| | ED | R_{adj}^2 | Standard error | Average error | RMS error | Maximum error |
|-----------------------------------|--------------|-------------|----------------|---------------|-----------|---------------|
| F ₁ (noise) | RMS bias CCD | 1.00 | 0.021 | 0.84 | 1.05 | 3.22 |
| | FCCD | 1.00 | 0.018 | 0.0081 | 0.010 | 0.044 |
| | LHS | 0.99 | 0.032 | 0.032 | 0.043 | 0.26 |
| | D-optimal | 0.99 | 0.022 | 0.023 | 0.025 | 0.042 |
| F ₂ (LHS sphere) | RMS bias CCD | 1.00 | 0.035 | 0.15 | 0.18 | 0.36 |
| | FCCD | 0.91 | 0.094 | 0.090 | 0.11 | 0.25 |
| | LHS | 0.99 | 0.024 | 0.042 | 0.072 | 1.29 |
| | D-optimal | 0.90 | 0.090 | 0.25 | 0.31 | 0.82 |
| F ₃ (D-optimal sphere) | RMS bias CCD | 1.00 | 0.00065 | 0.17 | 0.18 | 0.22 |
| | FCCD | 0.96 | 0.062 | 0.077 | 0.088 | 0.18 |
| | LHS | 0.99 | 0.024 | 0.035 | 0.055 | 0.69 |
| | D-optimal | 0.90 | 0.075 | 0.17 | 0.22 | 0.65 |

Table 3-5. Min-max RMS bias central composite designs for 2-5 dimensional spaces and corresponding design metrics (N_v is the number of variables), errors are computed on a uniform 41^2 grid ($N_v = 2$), 21^3 grid ($N_v = 3$) and, 11^{N_v} grid ($N_v > 3$), in space $V = [-1, 1]^{N_v}$. Refer to Appendix B for details of creating CCD.

| N_v | A_1 | α_2 | $(e_{es})_{\max}$ | $(e_{es})_{\text{avg}}$ | $ e_b^I _{\max}$ | $(e_b^{\text{rms}})_{\max}$ | $(e_b^{\text{rms}})_{\text{avg}}$ |
|-------|-------|------------|-------------------|-------------------------|------------------|-----------------------------|-----------------------------------|
| 2 | 0.954 | 1.000 | 0.973 | 0.688 | 1.029 | 0.341 | 0.269 |
| 3 | 0.987 | 1.000 | 0.913 | 0.607 | 2.832 | 0.659 | 0.518 |
| 4* | 1.000 | 1.000 | 0.877 | 0.585 | 6.208 | 1.176 | 0.827 |
| 4 | 1.000 | 0.100 | 70.71 | 35.22 | 6.996 | 1.155 | 0.927 |
| 5 | 1.000 | 0.100 | 77.46 | 41.60 | 12.31 | 1.826 | 1.200 |

*FCCD for four variables (not min-max RMS bias design)

Table 3-6. Mean and coefficient of variation (based on 100 instances) of different error metrics for various experimental designs in four-dimensional space (30 points). Errors are computed on a uniform 11^4 grid in space $V = [-1, 1]^4$. D_{eff} : D-efficiency, r_{max} : radius of the largest unsampled sphere, e_b^{rms} : RMS bias error, e_{es} : estimated standard error, $(\cdot)_{max}$: maximum of the quantity inside parenthesis, $(\cdot)_{avg}$: average of the quantity inside parenthesis.

| | ED | $(e_{es})_{max}$ | $(e_{es})_{avg}$ | $(e_b^{rms})_{max}$ | $(e_b^{rms})_{avg}$ | r_{max} | D_{eff} |
|------|-------------|------------------|------------------|---------------------|---------------------|-----------|-----------|
| Mean | LHS | 3.82 | 0.94 | 3.02 | 0.57 | 0.76 | 0.26 |
| | D-optimal | 0.80 | 0.61 | 1.67 | 0.89 | 0.90 | 0.99 |
| | Combination | 2.02 | 0.68 | 2.67 | 0.59 | 0.75 | 0.47 |
| COV | LHS | 0.209 | 0.074 | 0.105 | 0.058 | 0.066 | 0.056 |
| | D-optimal | 0.025 | 0.042 | 0.075 | 0.037 | 0.15 | 0.006 |
| | Combination | 0.126 | 0.026 | 0.102 | 0.038 | 0.064 | 0.053 |

Table 3-7. Reduction in errors by considering multiple experimental designs and picking one experimental design using appropriate criterion (filtering). We summarize actual magnitudes of maximum and space-averaged errors for LHS/D-optimal/combination designs in four-dimensional space with 30 points. All-ED refers to error data corresponding to all 300 LHS/D-optimal/combination designs. BE and SE refer to error data from 100 EDs selected using min-max RMS bias error and min-max estimated standard error criterion, respectively. Errors are computed on a uniform 11^{N_v} grid in space $V = [-1, 1]^{N_v}$. Mean and coefficient of variation (COV) are based on 300 (ALL-ED) or 100 (BE/SE) EDs.

| | LHS designs | | | D-optimal designs | | | Combination designs | | |
|---------------------------|-------------|-------|-------|-------------------|-------|-------|---------------------|-------|-------|
| | All-ED | BE | SE | All-ED | BE | SE | All-ED | BE | SE |
| Mean- $(e_{es})_{max}$ | 3.85 | 3.62 | 3.29 | 0.80 | 0.80 | 0.78 | 2.00 | 1.96 | 1.80 |
| COV- $(e_{es})_{max}$ | 0.21 | 0.19 | 0.11 | 0.025 | 0.020 | 0.016 | 0.12 | 0.12 | 0.066 |
| Mean- $(e_{es})_{avg}$ | 0.94 | 0.92 | 0.90 | 0.62 | 0.62 | 0.61 | 0.68 | 0.67 | 0.67 |
| COV- $(e_{es})_{avg}$ | 0.071 | 0.063 | 0.053 | 0.043 | 0.045 | 0.039 | 0.024 | 0.024 | 0.019 |
| Mean- $(e_b^{rms})_{max}$ | 3.08 | 2.83 | 2.95 | 1.67 | 1.57 | 1.69 | 2.68 | 2.46 | 2.60 |
| COV- $(e_b^{rms})_{max}$ | 0.099 | 0.053 | 0.083 | 0.074 | 0.049 | 0.069 | 0.094 | 0.066 | 0.091 |
| Mean- $(e_b^{rms})_{avg}$ | 0.58 | 0.57 | 0.57 | 0.89 | 0.88 | 0.89 | 0.59 | 0.58 | 0.59 |
| COV- $(e_b^{rms})_{avg}$ | 0.049 | 0.036 | 0.043 | 0.037 | 0.027 | 0.032 | 0.037 | 0.037 | 0.036 |

CHAPTER 4 ENSEMBLE OF SURROGATES

Surrogate models have been extensively used in the design and optimization of computationally expensive problems. Different surrogate models have been shown to perform well in different conditions. Barthelemy and Haftka (1993) reviewed the application of meta-modeling techniques in structural optimization. Sobieszczanski-Sobieski and Haftka (1997) reviewed different surrogate modeling applications in multi-disciplinary optimization. Giunta and Watson (1998) compared polynomial response surface approximations and kriging on analytical example problems of varying dimensions. Simpson et al. (2001a) reviewed different surrogates and gave recommendations on the usage of different surrogates for different problems. Jin et al. (2001) compared different surrogate models based on multiple performance criteria such as accuracy, robustness, efficiency, transparency, and conceptual simplicity. They recommended using radial basis function for high-order nonlinear problems, kriging for low-order nonlinear problems in high dimension spaces, and polynomial response surfaces for low-order nonlinear problems. They also noted difficulties in constructing different surrogate models. Li and Padula (2005) and Queipo et al. (2005) recently reviewed different surrogate models used in the aerospace industry.

There are also a number of studies comparing different surrogates for specific applications. Papila et al. (2001), Shyy et al. (2001a-b), Vaidyanathan et al. (2004), Mack et al. (2005a-b, 2006) presented studies comparing radial basis neural networks and response surfaces while designing the propulsion systems like, liquid rocket injector, supersonic turbines, and the shape of bluff body for mixing enhancement. For crashworthiness optimization, Stander et al. (2004) compared polynomial response surface approximation, kriging, and neural networks, while Fang

et al. (2005) compared polynomial response surface approximation and radial basis functions. Most researchers found that no single surrogate model was superior in general.

While most researchers have primarily been concerned with the choice among different surrogates, there has been relatively very little work about the use of an ensemble of surrogates. Zerpa et al. (2005) presented one application of simultaneously using multiple surrogates to construct a weighted average surrogate model for the optimization of an alkali-surfactant-polymer flooding process. They suggested that the weighted average surrogate model has better modeling capabilities than individual surrogates.

Typically, the cost of obtaining data required for developing surrogate models is high, and it is desired to extract as much information as possible from the data. Using an ensemble of surrogates, which can be constructed without a significant expense compared to the cost of acquiring data, can prove effective in distilling correct trends from the data and may protect against bad surrogate models. Averaging surrogates is one approach motivated by our inability to find a unique solution to the non-linear inverse problem of identifying the model from a limited set of data (Queipo et al., 2005). In this context, model averaging essentially serves as an approach to account for model uncertainty. In this work, we explore methods to exploit the potential of using an ensemble of surrogates. Specifically, we present the following two aspects:

- Ensemble of surrogates can be used to identify regions, where we expect large uncertainties (contrast).
- Use of an ensemble of surrogates *via* weighted averaging (combination) or selection of best surrogate model based on error statistics to induce robustness in approximation compared to individual surrogates.

We demonstrate the advantages of an ensemble of surrogates using analytical problems and an engineering problem of radial turbine design for space launch vehicle. This chapter is organized as follows. In the next section, we present a method to use an ensemble of surrogates

to identify the regions with large uncertainty, and the conceptual framework of constructing weighted average surrogate models. Thereafter, we discuss the test problems, numerical procedure, and results supporting our claims. We close the chapter by recapitulating salient points.

Conceptual Framework

Identification of Region of Large Uncertainty

Surrogate models are used to predict the response in unsampled regions. There is an uncertainty associated with the predictions. An ensemble of surrogates can be used to identify the regions of large uncertainty. The concept is described as follows: Let there be N_{SM} surrogate models. We compute the standard deviation of the predictions at a design point \mathbf{x} as,

$$s_{resp}(\hat{y}(\mathbf{x})) = \sqrt{\frac{\sum_{i=1}^{N_{SM}} (\hat{y}_i(\mathbf{x}) - \bar{y}(\mathbf{x}))^2}{N_{SM} - 1}}, \quad (4.1)$$

where $\bar{y}(\mathbf{x}) = \frac{\sum_{i=1}^{N_{SM}} \hat{y}_i(\mathbf{x})}{N_{SM}}$.

The standard deviation of the predictions will be high in regions where the surrogates differ greatly. A high standard deviation may indicate a region of high uncertainty in the predictions of any of the surrogates, and additional sampling points in this region can reduce that uncertainty. Note that, while high standard deviation indicates high uncertainty, low standard deviation does not guarantee high accuracy. It is possible for all surrogate models to predict similar response (yielding low standard deviation) yet perform poorly in a region.

Weighted Average Surrogate Model Concept

We develop a weighted average surrogate model as,

$$\hat{y}_{wta}(\mathbf{x}) = \sum_i^{N_{SM}} w_i(\mathbf{x}) \hat{y}_i(\mathbf{x}) \quad (4.2)$$

where $\hat{y}_{wta}(\mathbf{x})$ is the predicted response by the weighted average of surrogate models, $\hat{y}_i(\mathbf{x})$ is the predicted response by the i^{th} surrogate model, and $w_i(\mathbf{x})$ is the weight associated with the i^{th} surrogate model at design point \mathbf{x} . Furthermore, the sum of the weights must be one $\left(\sum_{i=1}^{N_{SM}} w_i = 1 \right)$ so that if all the surrogates agree, $\hat{y}_{wta}(\mathbf{x})$ will also be the same.

A surrogate model, deemed more accurate, should be assigned a large weight, and conversely, a less accurate model should have lower influence on the predictions. The confidence in surrogate models is given by different measures of “goodness” (quality of fit), which can be broadly characterized as: (i) global vs. local measures, and (ii) measures based on surrogate models vs. measures based on data. Weights associated with each surrogate, based on the local measures of goodness, are functions of space $w_i = w_i(\mathbf{x})$. For example, weights, which are based on the pointwise error measures, like prediction variance, mean squared error (surrogate based), or weights based on the interpolated cross-validation errors (data based). When weights are selected based on the global measures of goodness, they are fixed in design space $w_i(\mathbf{x}) = C_i, \forall \mathbf{x}$. For example, weights based on RMS error σ_a for polynomial response surface approximation, process variance for kriging (surrogate based), or weights based on cross-validation error (data based). While variable weights may capture local behavior better than constant weights, reasonable selection of weight functions is a formidable task.

Zerpa et al. (2005) constructed a local weighted average model from three surrogates (polynomial response surface approximation, kriging, and radial basis functions) for the optimization of an alkali-surfactant-polymer flooding process. Their approach was based on the pointwise estimate of the variance predicted by the three surrogate models.

There are different strategies of selecting weights. A few can be enumerated as follows:

Non-parametric surrogate filter (NPSF)

Weights are a function of relative magnitude of (global data-based) errors. The weight associated with i^{th} surrogate is given as:

$$w_i = \frac{\sum_{j=1, j \neq i}^{N_{SM}} E_j}{(N_{SM} - 1) \sum_{j=1}^{N_{SM}} E_j} \quad (4.3)$$

where E_j is a global or a local data-based error measure for the j^{th} surrogate model. This choice of weights gives only a small premium to the better surrogates when N_{SM} is large. For example, the best surrogate has a weight equal to or less than $(N_{SM} - 1)^{-1}$, which becomes unreasonably low when N_{SM} is large. On the positive side, the weights selected this way protect against errors induced by the surrogate models, which perform extremely well at the sampled data points but give poor predictions at unsampled locations.

Best PRESS for exclusive assignments

The traditional method of using an ensemble of surrogates is to select the best model among all considered surrogate models. However, once the choice is made, it is usually kept even as the design of experiment is refined. If the choice is revisited for each new design of experiment, we consider it as a weighting scheme where the model with least (global data-based) error is assigned a weight of one and all other models are assigned zero weight. In this study, we call this strategy the ‘best PRESS model’.

Parametric surrogate filter (PSF)

As discussed above, there are two issues associated with the selection of weights: (i) weights should reflect our confidence in the surrogate model, and (ii) weights should filter out

adverse effects of the model, which represents the data well but performs poorly in unexplored regions. A strategy to select weights, which addresses both issues, can be formulated as follows:

$$w_i^* = \left(E_i + \alpha E_{avg} \right)^\beta, \quad w_i = \frac{w_i^*}{\sum_i w_i^*} \quad (4.4)$$

$$E_{avg} = \sum_{i=1}^{N_{SM}} E_i / N_{SM}; \quad \beta < 0, \alpha < 1.$$

This weighting scheme requires the user to specify two parameters α and β , which control the importance of averaging and importance of individual surrogate, respectively. Small values of α and large negative values of β impart high weights to the best surrogate model. Large α values and small negative β values represent high confidence in the averaging scheme. In this study, we have used $\alpha = 0.05$ and $\beta = -1$. The sensitivity to these parameters is studied in a section on parameter sensitivity.

The above-mentioned formulation of weighting schemes is used with generalized mean square cross-validation error (GMSE; leave-one-out cross validation or PRESS in polynomial response surface approximation terminology), defined in Chapter 2, as global data-based error measure, by replacing E_j with $\sqrt{GMSE_j}$. We have used three surrogate models, polynomial response surface approximation (PRS), kriging (KRG), and radial basis neural networks (RBNN) (Orr, 1996), to construct the weighted average surrogate model. The PRESS-based weighted surrogate model (PWS) can then be given as follows:

$$\hat{y}_{pws} = w_{prs} \hat{y}_{prs} + w_{krg} \hat{y}_{krg} + w_{rbnn} \hat{y}_{rbnn} \quad (4.5)$$

where weights are selected according to the parametric surrogate filter PSF (Equation (4.4)). The rationale behind selecting these surrogate models to demonstrate the proposed approach was, (i)

these surrogate models are commonly used by practitioners, and (ii) they represent different parametric and non-parametric approaches (Queipo et al., 2005).

The cost of constructing surrogate models is usually low compared to that of analysis. If this cost is not small (for example, when using a kriging model and GMSE for large data sets), the user may want to explore surrogate models that provide a compromise solution between accuracy and construction cost. In general, the choice of surrogate models, which are most amenable to averaging and uncertainty identification, remains a question of future research (Sanchez et al., 2006).

Since global measures of error depend on the data and design of experiments, weights implicitly depend on the choice of the design of experiments. This dependence can be seen from Figure 4-1, where we show boxplots of weights obtained for 1000 instances of Latin hypercube sampling (LHS) design of experiments (DOEs) for Camelback function (described in next section). The center line of each boxplot shows the 50th-percentile (median) value and the box encompasses the 25th- and 75th- percentile of the data. The leader lines (horizontal lines) are plotted at a distance of 1.5 times the inter-quartile range in each direction or the limit of the data (if the limit of the data falls within 1.5 times the inter-quartile range). The data points outside the horizontal lines are shown by placing a '+' sign for each point.

We can see that the weights for different surrogates vary over a wide range with DOEs. The weights also give an assessment of relative contribution of different surrogate models to the weighted average surrogate model. In this example, polynomial response surface approximation had the highest weight most of the time (880 times), but not all the times (59 times kriging had the highest weight and 61 times RBNN had the highest weight).

Test Problems, Numerical Procedure, and Prediction Metrics

Test Problems

To test the predictive capabilities of the proposed approach of using an ensemble of surrogates, we employ two types of problems: (i) analytical (Dixon-Szegö, 1978), which are often used to test global optimization methods, and (ii) engineering, a radial turbine design problem (Mack et al., 2006), motivated by space launch. The details of each test problem are given as follows:

Branin-Hoo function

$$f(x, y) = \left(y - 5.1x^2 / 4\pi^2 + 5x / \pi - 6 \right)^2 + 10 \left(1 - 1/8\pi \right) \cos(x) + 10, \quad (4.6)$$
$$x \in [-5, 10], \quad y \in [0, 15].$$

Camelback function

$$f(x, y) = \left(4 - 2.1x^2 + x^4/3 \right) x^2 + xy + (-4 + 4y^2) y^2, \quad (4.7)$$
$$x \in [-3, 3], \quad y \in [-2, 2].$$

Goldstein-Price function

$$f(x, y) = \left[1 + (x + y + 1)^2 (19 - 4x + 3x^2 - 14y + 6xy + 3y^2) \right] \times$$
$$\left[30 + (2x - 3y)^2 (18 - 32x + 12x^2 + 48y - 36xy + 27y^2) \right], \quad (4.8)$$
$$x, y \in [-2, 2].$$

Figure 4-2 depicts these two-variable test problems and shows zones of high gradients.

Hartman functions

$$f(\mathbf{x}) = -\sum_{i=1}^m c_i \exp \left\{ -\sum_{j=1}^n a_{ij} (x_j - p_{ij})^2 \right\}, \quad (4.9)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $x_i \in [0, 1]$.

Two instances of this problem are considered based on the number of design variables. For the chosen examples, $m = 4$.

- **Hartman-3**

This problem has three variables. The choice of parameters is given in Table 4-1 (Dixon-Szegö, 1978).

- **Hartman-6**

This instance of the problem has six design variables and the parameters used in the function are tabulated in Table 4-2 (Dixon-Szegö, 1978).

Figure 4-3 illustrates the complexity of the analytical problems. It shows the boxplots of function values at a uniform grid of points with 21 points in each direction (for Hartman problem with six variables, we used five points in each direction); the mean, coefficient of variation, and median are given in Table 4-3. We can see that for all the problems, the coefficient of variation was close to one or more, which indicates large variation in the function values. It is clear from Figure 4-3 that the function values follow nonuniform distribution, which is also reflected by large differences in the mean and median. These conditions translate into high gradients in the functions and may pose difficulties in accurate modeling of the responses. Goldstein-Price and Hartman problem with six variables had a significant number of points, which had higher function values than the inter-quartile range of the data. This is reflected in the high coefficient of variation of these two functions.

Radial turbine design for space launch

As described by Mack et al. (2006), this six-variable problem is motivated by the design of compact radial turbine used to drive pumps that deliver liquid hydrogen and liquid oxygen to the combustion chamber of a spacecraft. The objective of the design is to increase the work output of a turbine in the liquid rocket expander cycle engine while keeping the overall weight of the turbine low. If the turbine inlet temperature is held constant, the increase in turbine work is directly proportional to the increase in efficiency. Thus, the design goal is to maximize the turbine efficiency while minimizing the turbine weight. Our interest in this problem is to develop

accurate surrogate model(s) of the efficiency as a function of six design variables. The description of design variables and their corresponding ranges are given in Table 4-4 (Mack et al., 2006). The objectives of the design were calculated using a one-dimensional flow analysis ‘meanline’ code (Huber, 2001). Mack et al. (2006) identified the appropriate region of interest by iteratively refining the design space. They also identified the most important variables using global sensitivity analysis.

Numerical Procedure

For all analytical problems, Latin hypercube sampling (LHS) was used to pick design points such that the minimum distance between the design points is maximized. We used MATLAB routine ‘lhsdesign’ with ‘maximin’ criterion (maximize the minimum distance between points) and a maximum of 40 iterations to obtain optimal configuration of points. For the radial turbine design problem, Mack et al. (2006) sampled 323 designs in the six-dimensional region of interest, using LHS and a five level factorial design on the three most important design variables (identified by global sensitivity analysis). Out of these 323 designs, 13 designs were found infeasible. The remaining 310 design points were used to construct and test the surrogate models. For this study, we randomly selected 56 points to construct the surrogate model and used the remaining 254 points to test the surrogate model. To reduce the effect of random sampling for both analytical and radial turbine design problems, we present results based on 1000 instances of design of experiments for all the problems in low dimension spaces. However, to keep computational cost low for six-variable problems, we used 100 design of experiments and then, used 1000 bootstrap (Hesterberg et al., 2005) samples to estimate results.

The numerical settings used to fit different surrogate models for each problem are given in Table 4-5. The total number of test points (on a uniform grid) is p^{N_v} where N_v is the number of

variables and p is the number of points along each direction (Table 4-5), except for the radial turbine problem, where the number represents the total number of test points. We used reduced-quadratic or reduced-cubic polynomials for PRS. A Gaussian correlation function and a linear trend model were used in kriging approximation of all test problems. Parameters ‘spread’ and ‘goal’ for radial basis neural network were selected according to problem characteristics (‘spread’ controls the decay rate of radial basis function; and ‘goal’ is the desired level of accuracy of the RBNN model on training points). It should be pointed out that no attempt was made to improve the predictions of any surrogate model.

Prediction Metrics

The following metrics were used to compare the prediction capabilities of different surrogate models:

Correlation coefficient

The correlation coefficient between actual and predicted response at the test points $r(y, \hat{y})$ is given as

$$r(y, \hat{y}) = \frac{\frac{1}{V} \int (y - \bar{y})(\hat{y} - \bar{y}) dv}{\sigma(y)\sigma(\hat{y})} \quad (4.10)$$

It is numerically evaluated from the data for test points by implementing quadrature³ for integration (Ueberhuber, 1997) as given in Equation (4.11).

$$\begin{aligned} \frac{1}{V} \int y \hat{y} dv &= \sum_{i=1}^{N_{test}} \xi_i y_i \hat{y}_i / N_{test} \quad ; \bar{y} = \sum_{i=1}^{N_{test}} \xi_i y_i / N_{test} \\ \sigma(y) &= \sqrt{\frac{1}{V} \int (y - \bar{y})^2 dv} = \sqrt{\sum_{i=1}^{N_{test}} \xi_i (y_i - \bar{y}_i)^2 / N_{test}} \end{aligned} \quad (4.11)$$

³ Here we used trapezoidal rule for integration.

where \bar{y} is the mean of actual response, $\bar{\hat{y}}$ is the mean of predicted response, N_{test} is the number of test points, and ξ_i is the weight used for integration using the trapezoidal rule. For radial turbine problem, we used a nonuniform set of data points so the correlation coefficient is obtained using Equation (4.11) with weight $\xi_i = 1$. For a high-quality surrogate model, the correlation coefficient should be as high as possible. The maximum value of $r(y, \hat{y})$ is one that defines exact linear relationship between the predicted and the actual response.

RMS error

For all the test problems, the actual response at test points was known, which allowed us to compute error at all test points. The root mean square error (RMSE) in the design domain, as defined in Equation (4.12), was used to assess the goodness of the predictions.

$$RMSE = \sqrt{\frac{1}{V} \int_V (y - \hat{y})^2 dv} \quad (4.12)$$

Equation (4.12) can be evaluated using trapezoidal rule as denoted in Equation (4.13).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{test}} \xi_i (y_i - \hat{y}_i)^2}{N_{test}}} \quad (4.13)$$

For radial turbine problem, we used Equation (4.13) with weight $\xi_i = 1$ to estimate the RMS error. Of course, a good surrogate model gives low RMS error.

Maximum error

Another measure of the quality of prediction of a surrogate is the maximum absolute error at the test points. This is required to be low.

A combination of high correlation coefficient, low RMS error, and low maximum error indicates a good prediction.

Results and Discussion

In this section, we present some numerical results to demonstrate the capabilities of multiple surrogate models using the test problems discussed in previous section.

Identification of Zones of High Uncertainty

We demonstrate the application of an ensemble of surrogates to identify the region of high uncertainty with the help of different test problems. Results for a single instance of a DOE for Branin-Hoo example are presented in detail. Figure 4-4 shows the contour plots of absolute errors in the prediction ($|y(\mathbf{x}) - \hat{y}(\mathbf{x})|$) due to different surrogate models and the standard deviation of the responses.

Figure 4-4(A)-(C) shows contour plots of actual absolute errors in different surrogate models. It can be seen that the middle section of the design space was approximated very well (errors are low) but the left boundary was poorly represented by different surrogate models. The errors (and hence, responses) from PRS, KRG, and RBNN differed in the region close to the top left corner. The contour plot of the standard deviation (Figure 4-4(D)) of predicted responses correctly indicated the region of high uncertainty near the top left corner due to high standard deviation. It also appropriately identified good predictions in the central region of design space. The predictions in the region of high uncertainty can be improved by sampling additional points.

It is also noted that although all surrogate models had high errors near the bottom left corner of the design space, (Figure 4-4(A)-(C)), the standard deviation of the predicted responses was not high. This means that we can use the standard deviation of surrogate models to identify regions of high uncertainty, but we cannot use it to identify regions of high fidelity. This particular situation demands further investigation, if the objective of using an ensemble of surrogates was to identify a region of high error in the predictions.

To further show the independence of the result with respect to the design of experiments, we simulated the Branin-Hoo function with 1000 DOEs. For each DOE, we computed the standard deviation of responses in design space. At the location of maximum standard deviation for each DOE, we computed actual errors in the predictions of different surrogates. Similarly, we calculated actual errors in the predictions of different surrogates at the location of minimum standard deviation. Figure 4-5(A) shows the magnitude of maximum standard deviation and actual errors in predictions using different surrogates for 1000 DOEs, and Figure 4-5(B) shows the magnitude of minimum standard deviation and actual errors in predictions using different surrogates from 1000 DOEs. By comparing Figure 4-5(A) and (B), it is clear that high standard deviation of responses corresponded to the regions with large uncertainties in the predictions and low standard deviation corresponded to regions with low uncertainty and there was an order of magnitude difference.

To generalize the findings, we simulated all test problems and identified the actual errors at the locations of maximum and minimum standard deviation of responses. The results are summarized in Table 4-6 and Table 4-7. A one-to-one comparison of the results for different test problems shows that when the standard deviation of responses was highest, the actual errors in predictions were high, and when the standard deviation of responses was lowest, the actual errors in predictions were low. We note that the results are more useful for a qualitative comparison than quantitative, i.e., identifying the regions where we expect large uncertainties in prediction rather than quantifying the magnitude of actual errors.

We also estimated the maximum (over the entire design space) error due to each surrogate model for different test problems and compared with the maximum standard deviation of responses. The results are presented in Table 4-8. While the maximum standard deviation of

responses was the same order of magnitude as the maximum actual error for all surrogate models, it underestimated the maximum error by a factor of 2.5–4.0. When the number of data points to construct the surrogate model was increased (Branin-Hoo function was modeled with 31 points and Camelback function was modeled with 40 points, refer to a following section for details about modeling with increased sample density), the underestimation of the maximum actual error was reduced.

The main conclusions of the results presented in this section are: (1) dissimilar predictions of surrogate models (high standard deviation of responses) indicate regions of high errors, (2) similar predictions of surrogate models (low standard deviation of responses) do not necessarily imply small errors, and (3) the maximum standard deviation of responses underestimates the actual maximum error.

Robust Approximation *via* Ensemble of Surrogates

Next, we demonstrate the need of robust approximation with the help of Table 4-9 that enlists the number of times each surrogate yields the least PRESS error for all test problems. As can be seen, no surrogate model is universally the best for all problems. Besides, for any given problem, the choice of best surrogate model is affected by the design of experiment (except the radial turbine design problem). The results presented in Table 4-9 clearly establish the need to search robust approximation models (i.e., the same surrogate model can be applied to different problems, and the results produced are not significantly influenced by the choice of DOE).

We present results to reflect the advantages of using an ensemble of surrogates. Firstly, we quantify the number of points (reflects portion of design space), where different surrogates yield errors of opposite signs in Table 4-10. The predictions at these locations can be potentially improved due to error cancellation *via* proposed PRESS-based weighted surrogate (PWS) model. Next, we compare the performance of the PWS model and the surrogate model corresponding to

the best generalization error among the three surrogates (best PRESS model) with individual surrogate models (PRS, kriging, and RBNN). For each problem, the summary of the results based on 1000 DOEs is shown with the help of boxplots. A small size of the box suggests small variation in results with respect to the choice of design of experiment.

Correlations

Figure 4-6 shows the correlation coefficient (between actual and predicted responses) for different test problems. The results were statistically significant (p -value is smaller than $1e-4$) for all problems and DOEs. It is evident that no single surrogate worked the best for all problems, and the correlation coefficient for individual surrogates varied with DOE. Both, the best PRESS and the PWS, models were better than the worst surrogate model and at par with the corresponding best surrogate for most problems. The PWS model generally performed better than the best PRESS model. The variation in results with respect to the design of experiments for both, the PWS model and the best PRESS model, was also comparable to the best surrogate for all problems, except Hartman problem with six variables.

For all problems, we observed that some design of experiments (DOEs) yielded very poor correlations. Analysis of the corresponding experiments revealed two scenarios: (1) Sometimes, the DOE was not satisfactory and a large portion of the design space was unsampled. This led to poor performance of all the surrogate models. (2) For a few poor correlation cases, despite a good DOE, one or more surrogates failed to capture the correct trends. The PWS model and the best PRESS model were able to correct the anomalies in these scenarios to some extent. The tail of the boxplot corresponding to the PWS model and the best PRESS model was shorter compared to the worst surrogate (Figure 4-6).

Table 4-11 shows the mean and the coefficient of variation for different test problems to assess the performance of different surrogate models. It is clear that the average correlation

coefficient for the PWS model was either the best or the second best for all the test problems. Also, the low coefficient of variation underscored the relatively low sensitivity of the PWS model with respect to the choice of design of experiments. Performance of the best PRESS model was also comparable to the best surrogate model for each problem. The overall performance of all three surrogates was comparable. It can also be seen from Table 4-11 that the PWS model outperformed the best PRESS model for all cases, but radial turbine design problem.

The mean of the correlation coefficient for different problems is reported based on one set of 1000 DOEs. Since the distribution of mean is approximately Gaussian, the coefficient of variation of the mean (of correlation coefficient) can be given as $COV \times (N_{DOE})^{-0.5}$, where COV is the coefficient of variation (of correlation coefficient) based on 1000 DOEs ($N_{DOE} = 1000$), leading to a coefficient of variation of the mean that is about 30 times lower than the native coefficient of variation. The number of digits in the table is based on this estimate of the coefficient of variation.

We verified the results by performing the bootstrap analysis (Hesterberg et al., 2005) by considering 1000 samples of 1000 DOEs each. The distribution of the mean for one representative case (mean correlation coefficient predicted using kriging approximation for Branin-Hoo function) is plotted in Figure 4-7. The mean correlation coefficient evidently follows the Gaussian distribution as the data falls on the straight line depicting the normal distribution. Similar results were observed for all other cases. Bootstrapping also confirmed that the coefficient of variation of the mean value followed the simple expression given above.

RMS errors

Next, we compared different surrogate models based on the RMS errors in predictions at test points. Figure 4-8 shows the results on different test problems. While no single surrogate

performed the best on all problems, individual surrogate models approximated different problems better than others. The PRESS-based weighted surrogate (PWS) model and the best PRESS model performed reasonably for all test problems. The results indicate that if we know that a particular surrogate performs the best for a given problem, it is best to use that surrogate model for approximation. However, for most problems, the best surrogate model is not known *a priori* or the choice of best surrogate may get affected by choice of DOE (Table 4-9). Then an ensemble of surrogates (*via* the PWS or the best PRESS model) may prove beneficial to protect against the worst surrogate model.

The mean and coefficient of variation of RMS errors using different surrogates on different problems are tabulated in Table 4-12. Note that, kriging most often had the lowest RMS errors compared to other surrogates. When the RMS errors due to all surrogates were comparable, as was the case for Branin-Hoo and Camelback functions, the predictions using the PWS model were more accurate (lower RMS error) than any individual surrogate. However, when one or more surrogate models were much more inaccurate than others, the predictions using the PWS model were only reasonably close to the accurate surrogate model(s). We also observed that both the best PRESS model and the PWS model were able to significantly reduce the errors compared to the worst surrogate. This suggests that using an ensemble of surrogate models, we can protect against poor choice of a surrogate.

The PWS model generally yielded lower RMS errors than the best PRESS model. Relatively poor performance of the PWS model (compared to the best PRESS model) for six-variable Hartman problem and the radial turbine problem was attributed to accurate modeling of the response by one surrogate or inaccuracy in the representation of weights (see section on the role of generalized cross-validation errors).

Maximum absolute errors

Figure 4-9 shows the maximum absolute error for 1000 DOEs using different surrogate models on different test problems. As was observed for RMS errors, the PWS model and the best PRESS model performed reasonably for all test problems, though individual surrogate models performed better for different test problems.

Numerical quantification of the results is given in Table 4-13. The maximum absolute error obtained using the PWS model and the best PRESS model were comparable to the maximum absolute error obtained using the best surrogate model for that test problem. For most cases, the PWS model also delivered a lower maximum absolute error than the best PRESS model. Relatively poor performance of the PWS model for the Goldstein-Price test problem was attributed to the poor performance of one of the surrogate models (RBNN) on the prediction points.

The results presented in this section suggest that the strategy of using an ensemble of surrogate models potentially yields robust approximation (good correlation, low RMS, and maximum errors) for problems of varying complexities and dimensions, and the results are less sensitive to the choice of DOE. The PWS model may have an advantage compared to the best PRESS model.

Studying the role of generalized cross-validation errors

We observed that the PWS model did not perform well for Camelback and Goldstein-Price function where the RBNN model noticeably yielded large variations. To investigate the underlying issue, we studied the weights and, hence, the role of PRESS error, which is used to determine the weights. Our initial assumption was that the PRESS error is a good estimate of the actual RMS errors for all surrogate models. To validate this assumption, we computed the ratio of actual RMS errors and PRESS for different surrogate models over 1000 DOEs. The results are

summarized in Figure 4-10, and the corresponding mean and standard deviation (based on 1000 DOEs) are given in Table 4-14.

It is observed from the results that PRESS (generalized cross-validation error), on average, underestimated actual RMS errors for polynomial response surface approximation but overestimated RMS error in kriging and RBNN. For Goldstein-Price, the mean was skewed for RBNN because of three simulations, which gave a very large ratio of RMS error and PRESS (the median is 0.44). The implication of this underestimation/overestimation was that the weights associated with polynomial response surface model were overestimated, and weights for kriging and radial basis neural network were underestimated. Noticeably, there were a large number of instances for Camelback and Goldstein-Price functions where PRESS underestimated the RMS errors for RBNN (see long tail of points, with RMS error to PRESS ratio greater than two). This indicated wrong emphasis of RBNN model for these models compared to other more accurate surrogates; hence, a relatively poor performance of the parametric weighted surrogate model was observed. This anomaly in accurately representing the actual errors or developing measures to correct the weight to account for the underestimation/overestimation is a scope of future research.

Effect of sampling density

Often an initial DOE identifies regions of interest, and then the DOE is refined in these regions. At other times, the initial DOE is found insufficient for good approximation, so that it must be refined. The refinement of the DOE can be carried out in two ways: (i) increasing the number of points in the original design space, and (ii) reducing the size of design space. The refinement of the DOE may change the identity of the best surrogate model, so that even if a single surrogate model is used, it may be useful to switch surrogates. In addition, the choice between the best PRESS and the PWS model may depend on sampling density. To investigate

these issues, we study two representative problems: Branin-Hoo function and Camelback function, which were not adequately approximated by different surrogate models (low correlations). Both problems are now modeled with increased number of points (31 points were used for Branin-Hoo function and Camelback function was modeled with 40 points) such that all regions were adequately modeled. We used a cubic polynomial to model Branin-Hoo function and a quartic polynomial to model Camelback function. All other parameters were kept the same. The results obtained for the increased number of points were compared with the previously presented results for smaller number of points in Table 4-15 and Table 4-16.

As can be seen from Table 4-15 and Table 4-16, the predictions improved with increasing number of points. The improvement in kriging (which models the local behavior better) was significantly more than the other two surrogates. The performance of both the best PRESS model and the PWS model was comparable to the best individual surrogate model and significantly better than the worst surrogate model. For the problems considered here, the best PRESS model outperformed the PWS model. This result is expected because of much improved modeling of the objective function by one or more of the surrogates. The results corroborate our earlier findings: (1) if we *a priori* know the best surrogate model for a given problem, that surrogate should be used for approximation; and (2) ensemble of surrogates protects us against the worst surrogate model. These results were evident irrespective of the number of points used to model the response. However, we also note that even if a single surrogate is used, its choice depends on sampling density. For Branin-Hoo function with 12 points, the polynomial response surface approximation had the best correlation and lowest maximum error. Its mean RMS error is slightly higher than kriging but standard deviation is much better. With 31 points, kriging is the best surrogate.

Sensitivity analysis of PWS parameters

To study the effect of variation in the parameters α and β (see Equation (4.4)), we constructed the PWS model for the Goldstein-Price function with different values of α and β . This problem was selected because of significant differences in the performance of different surrogate models. All other parameters were kept the same. The comparison of correlation coefficient and errors based on 1000 DOE samples is given in Table 4-17. To eliminate the skewness of the data due to a few spurious results, we show median, first, and third quartile data for all cases.

When we increased α keeping β constant, we observed modest decrease in errors. This was expected because by increasing α we reduced the importance of individual surrogates and assigned more importance to the averaging, which helped in reducing the effect of bad surrogates. However, it is noteworthy that a few designs, which gave poor performance of one surrogate, deteriorated the performance of the PWS model for respective cases. By increasing β keeping α constant, we emphasized the importance of individual surrogates more than the averaging. For this case, the overall effect was the deterioration of correlation and increase in errors. The effect of variation in β on the results was more pronounced than the effect of variation in α . The above results indicated that the parameters α and β should be chosen according to the performance of the individual surrogates.

Conclusions

In this chapter, we presented a case to simultaneously use multiple surrogates (1) to identify regions of high uncertainty in predictions, and (2) to develop a robust approximation strategy. The main findings can be summarized as follows.

- Regions of high standard deviation in the predicted response of the surrogates correspond to high errors in the predictions of the surrogates. However, we caution the user not to

interpret the regions of low standard deviation (uncertainty) as regions of low error. The standard deviation of responses usually underestimates the error.

- Simultaneous use of multiple surrogate models can improve robustness of the predictions by reducing the impact of a poor surrogate model (which may be an artifact of choice of design of experiment or the inherent unsuitability of the surrogate to the problem). Two suggested ways of using an ensemble of surrogates are to construct PRESS-based weighted average surrogate model or to select the surrogate model that has the least PRESS error among all considered surrogate models.
- The proposed PRESS error based selection of multiple surrogates performed at par with the best individual surrogate model for all test problems, and showed relatively low sensitivity to the choice of DOE, sampling density, and dimensionality of the problem.
- The PRESS-based weighted surrogate model yielded best correlation between actual and predicted response for different test problems.
- While different surrogates performed the best for reducing error (RMS and maximum absolute error) in different test problems, the performance of surrogate models was influenced by the selection of DOE. Ensemble of surrogates (*via* the PRESS-based weighted surrogate (PWS) and the best PRESS model) performed at par with the corresponding best surrogate model for all test problems. The PWS model in general outperformed the surrogate model with best PRESS error.
- It was also observed that PRESS in general underestimated the actual RMS error for polynomial response surface approximation and overestimated the actual RMS error for kriging and radial basis neural network. The correction in weights to account for the underestimation/overestimation of RMS errors by PRESS is a scope of future research.
- Though the best individual surrogate can change with increase in sampling density, the ensemble of surrogates performs comparably with the best surrogate.

We conclude that for most practical problems, where the best surrogate is not known beforehand, use of an ensemble of surrogates may prove a robust approximation method.

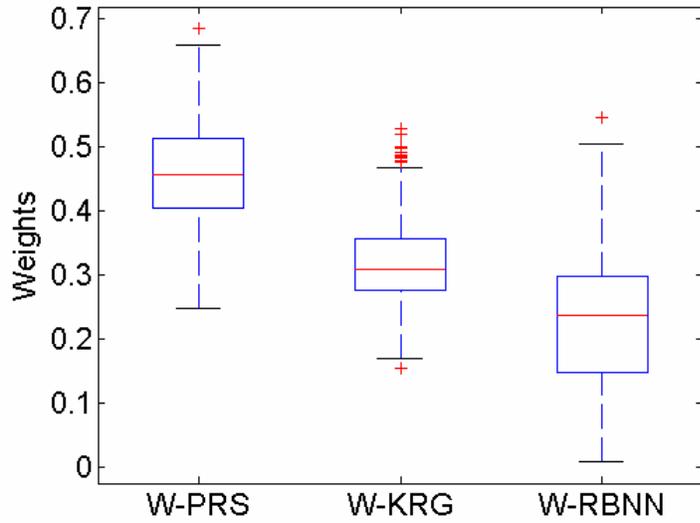


Figure 4-1. Boxplots of weights for 1000 DOE instances (Camelback function). W-PRS, W-KRG and W-RBNN are weights associated with polynomial response surface approximation, kriging, and radial basis neural network models, respectively.

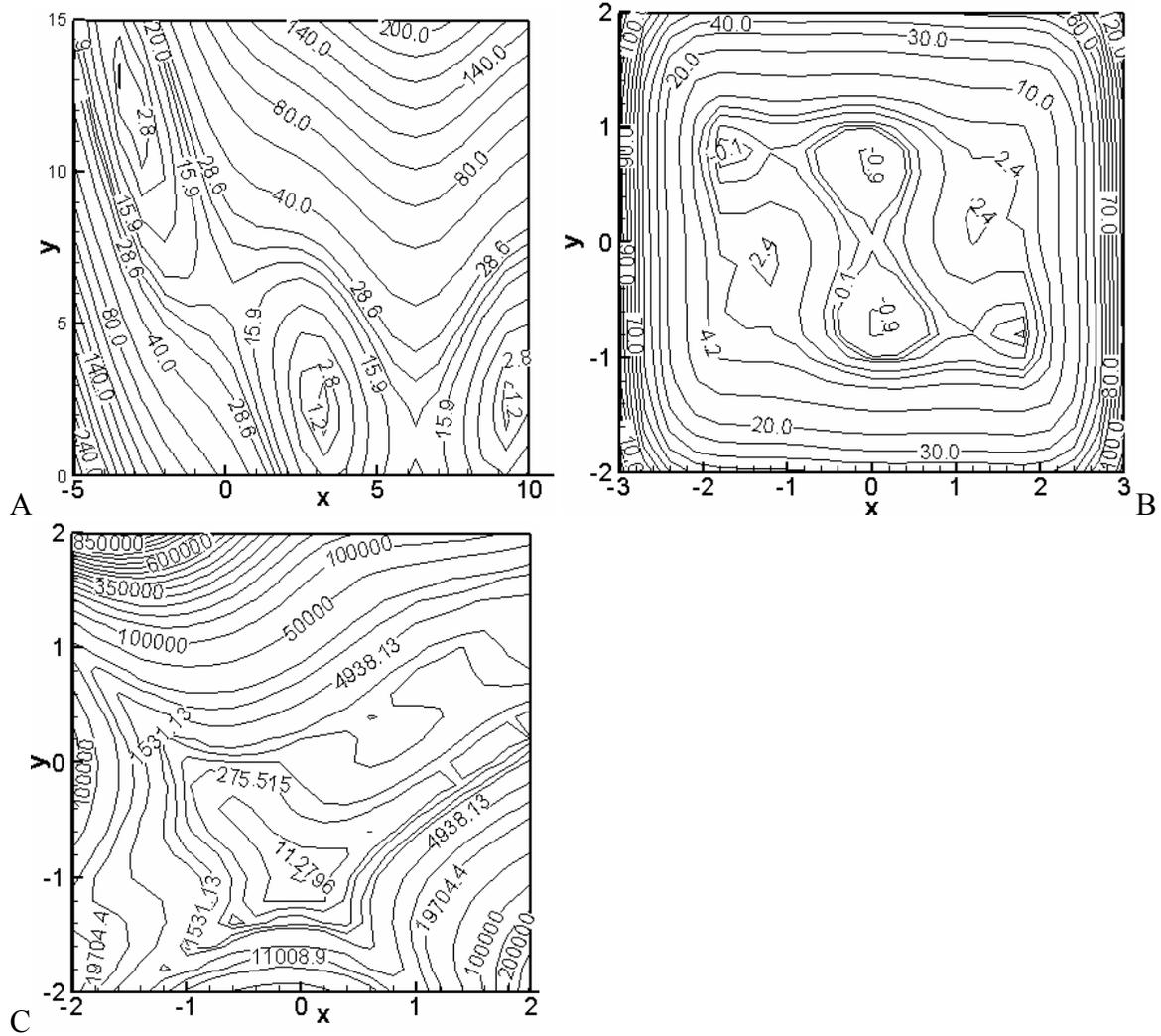


Figure 4-2. Contour plots of two variable test functions. A) Branin-Hoo. B) Camelback. C) Goldstein-Price.

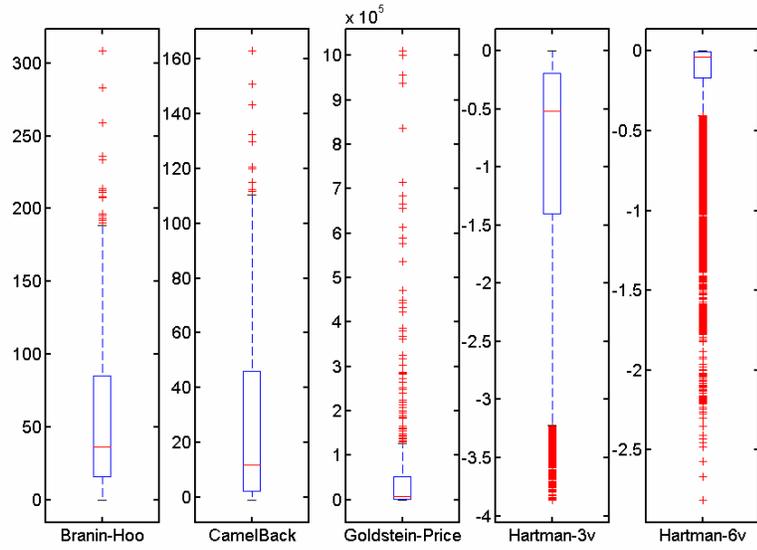


Figure 4-3. Boxplots of function values of different analytical functions.

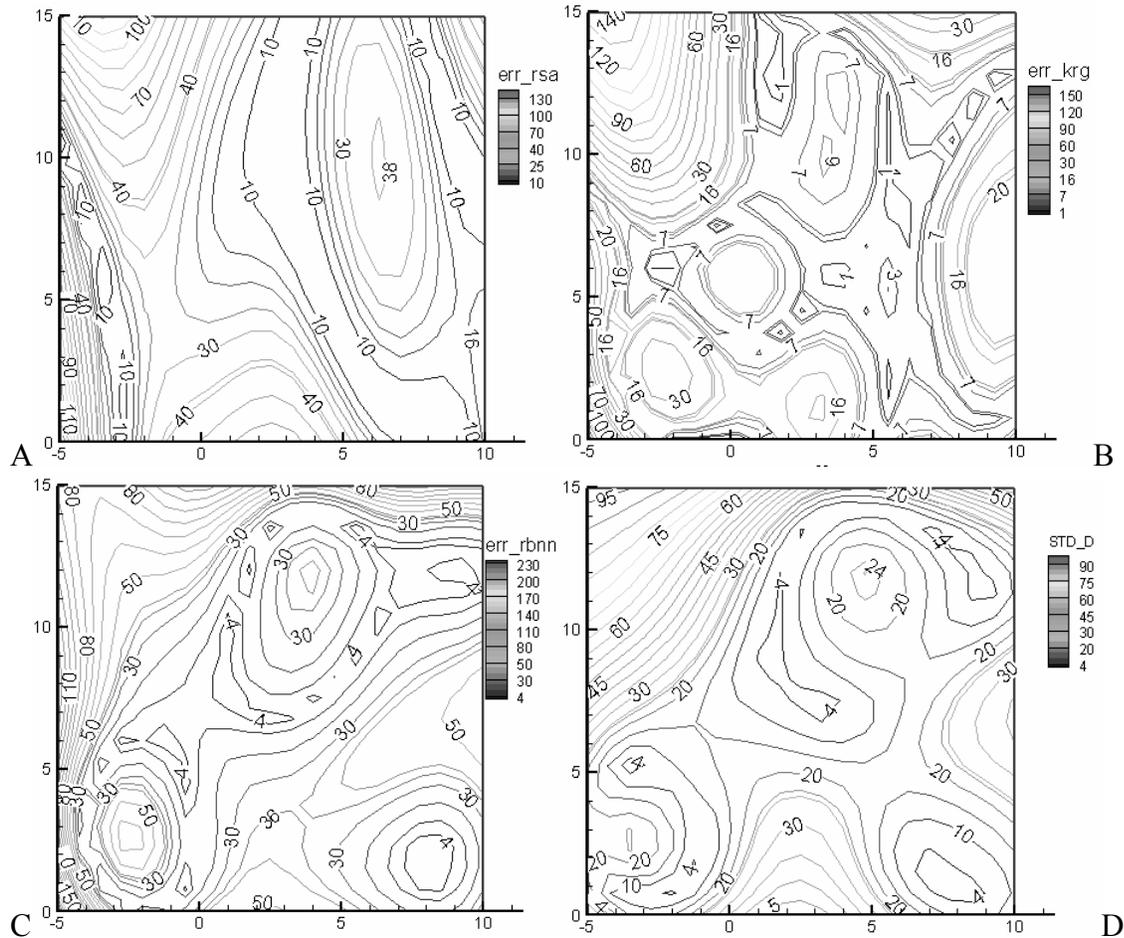


Figure 4-4. Contour plots of errors and standard deviation of predictions considering PRS, KRG, and RBNN surrogate models for Branin-Hoo function.

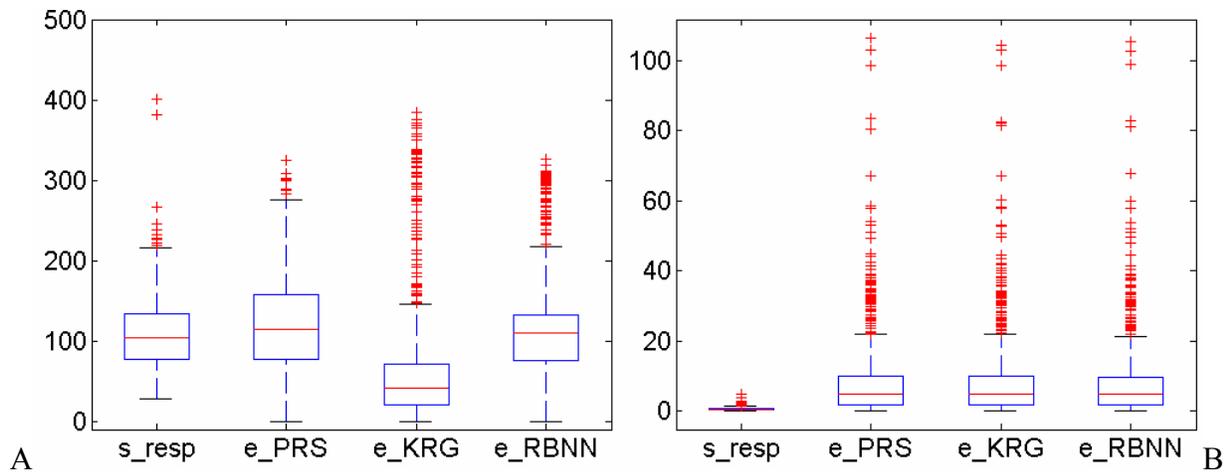


Figure 4-5. Boxplots of standard deviation of responses and actual errors in prediction of different surrogates at corresponding locations (based on 1000 DOEs using Branin-Hoo function). s_{resp} is standard deviation of responses, e_PRS, e_KRG, e_RBNN are actual errors in PRS, KRG and RBNN. A) Maximum standard deviation of responses and corresponding actual errors in approximations. B) Minimum standard deviation of responses and corresponding actual errors in approximations.

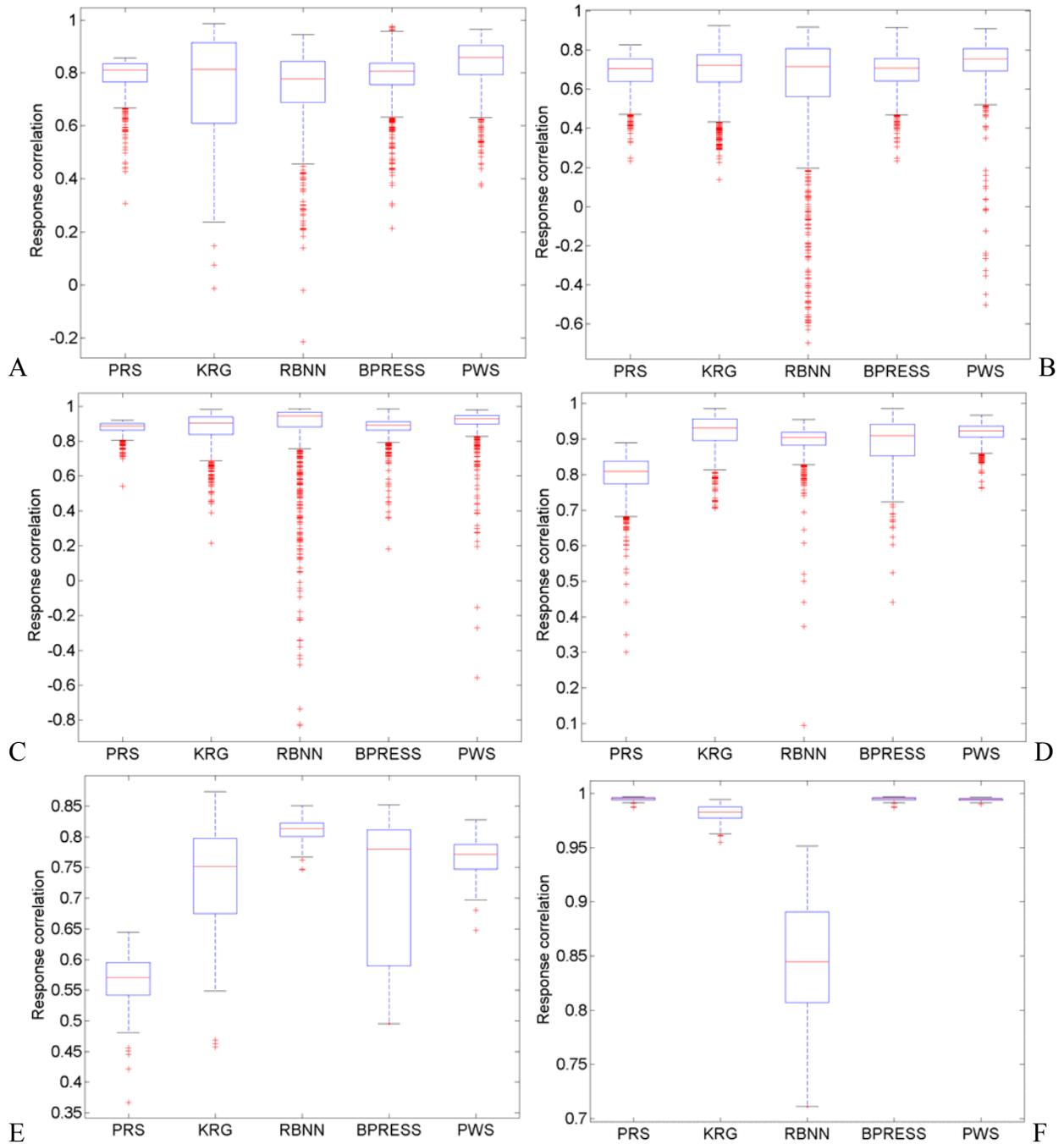


Figure 4-6. Correlations between actual and predicted response for different test problems. 1000 instances of DOEs were considered for all test problems, except Hartman-6 and radial turbine design problem for which we show results based on 100 samples. The center line of each boxplot shows the median value and the box encompasses the 25th- and 75th-percentile of the data. The leader lines (horizontal lines) are plotted at a distance of 1.5 times the inter-quartile range in each direction or the limit of the data (if the limit of the data falls within 1.5 times the inter-quartile range). A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem.

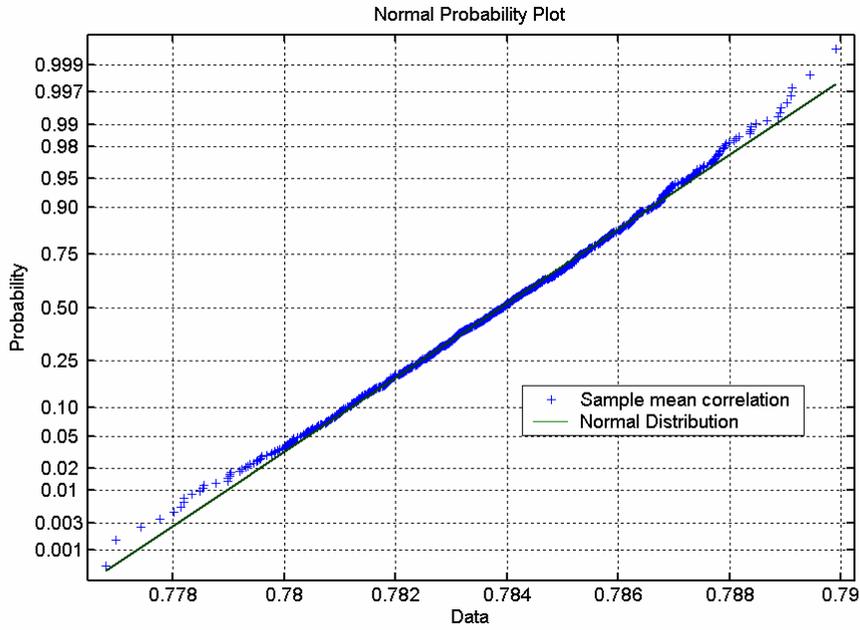


Figure 4-7. Normal distribution approximation of the sample mean correlation coefficient data obtained using 1000 bootstrap samples (kriging, Branin-Hoo function).

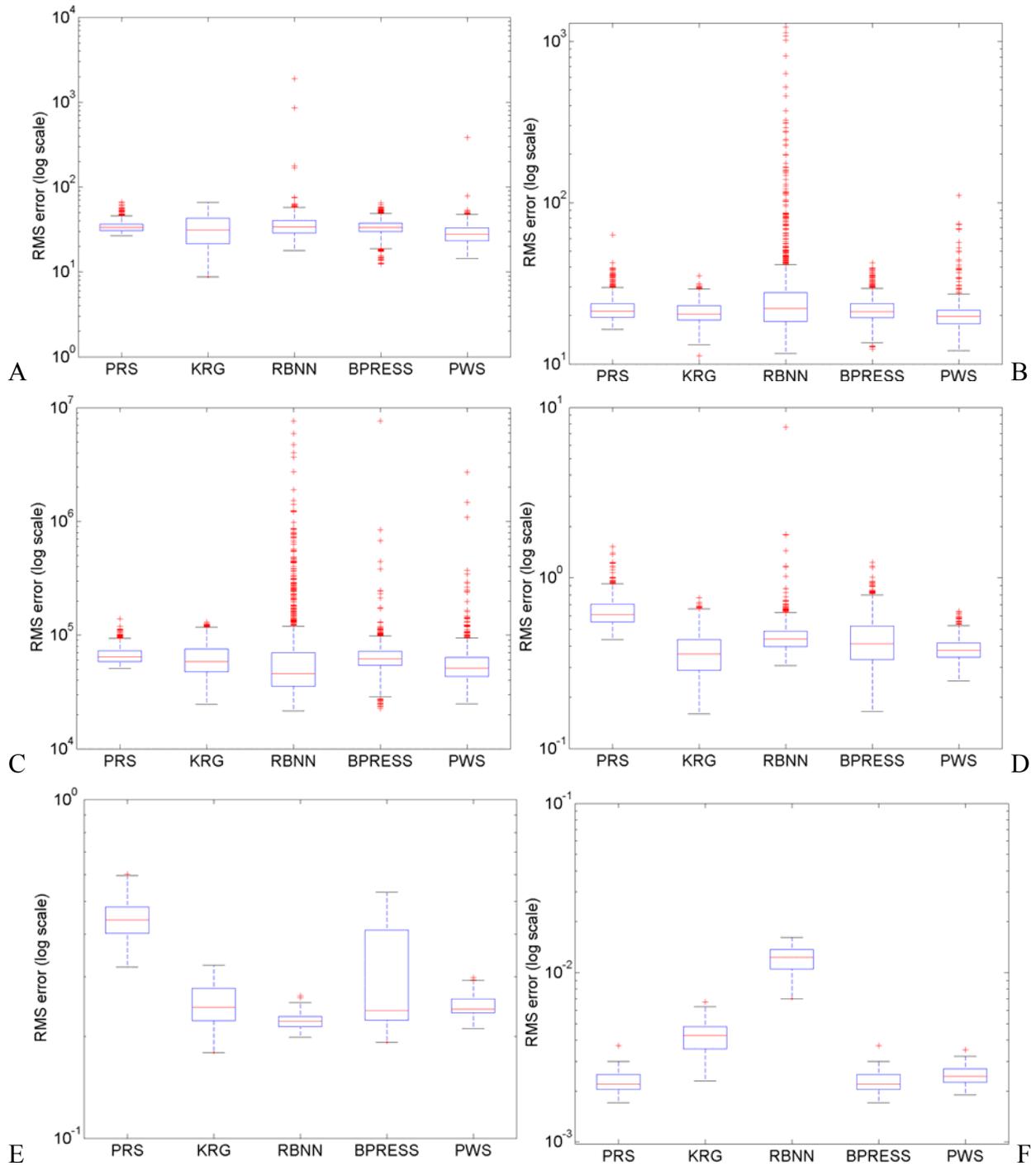


Figure 4-8. RMS errors in design space for different surrogate models. 1000 instances of DOEs were considered for all test problems, except Hartman-6 and radial turbine design problem for which we show results based on 100 samples. A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem.

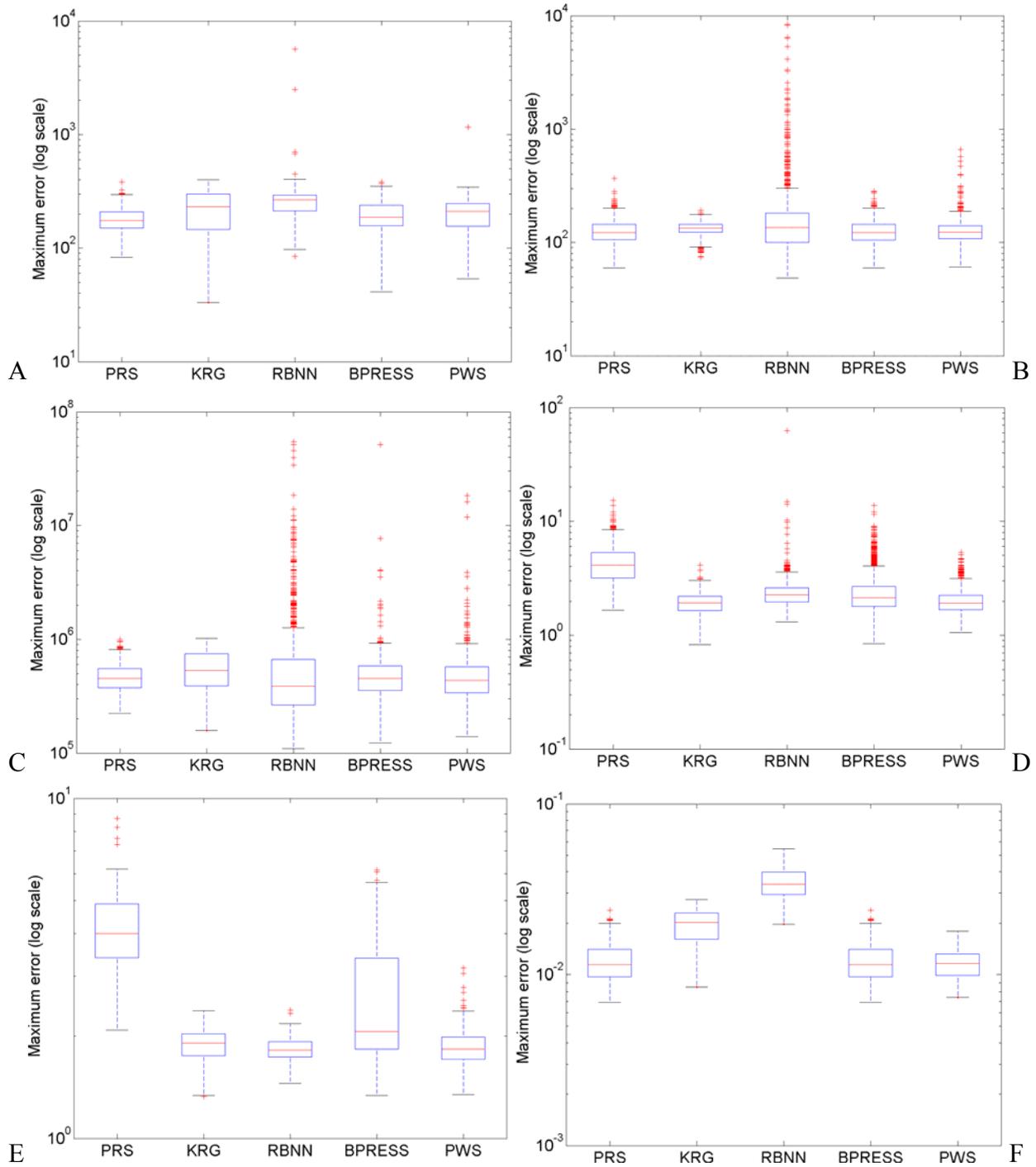


Figure 4-9. Maximum absolute error in design space for different surrogate models. 1000 instances of DOEs were considered for all test problems, except Hartman-6 and radial turbine design problem for which we show results based on 100 samples. A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem.

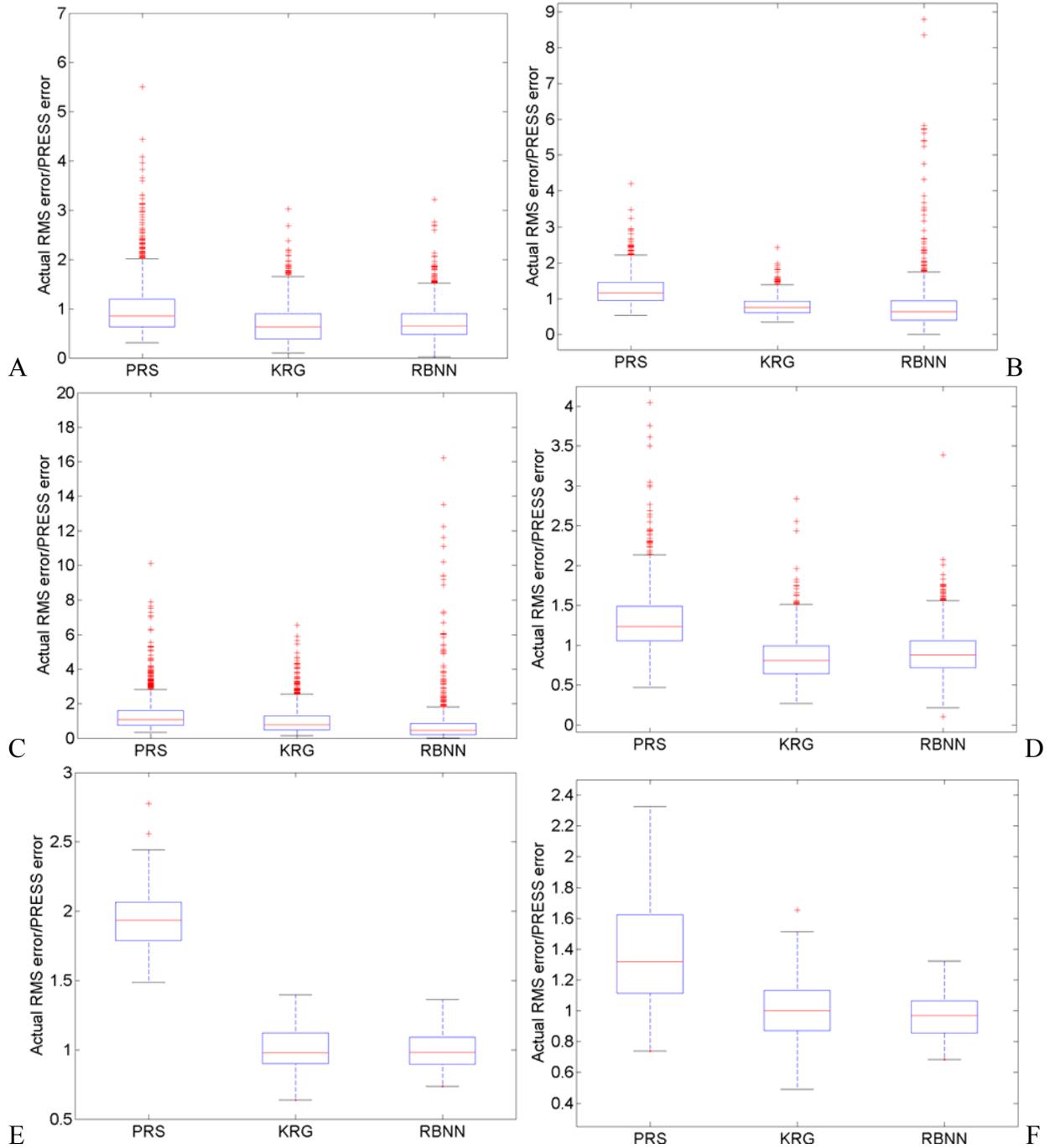


Figure 4-10. Boxplots of ratio of RMS error and PRESS over 1000 DOEs for different problems.

*For Branin-Hoo function, one simulation yielded RMSE/PRESS ratio $\sim O(20)$ for PRS, ** For Goldstein-Price problem, three simulations yielded high ratio of RMS error and PRESS error (20-80) for RBNN. A) Branin-Hoo *. B) Camelback. C) Goldstein-Price **. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem.

Table 4-1. Parameters used in Hartman function with three variables.

| <i>i</i> | <i>a_{ij}</i> | | | <i>c_i</i> | <i>p_{ij}</i> | | |
|----------|-----------------------|------|------|----------------------|-----------------------|--------|--------|
| 1 | 3.0 | 10.0 | 30.0 | 1.0 | 0.3689 | 0.1170 | 0.2673 |
| 2 | 0.1 | 10.0 | 35.0 | 1.2 | 0.4699 | 0.4387 | 0.7470 |
| 3 | 3.0 | 10.0 | 30.0 | 3.0 | 0.1091 | 0.8732 | 0.5547 |
| 4 | 0.1 | 10.0 | 35.0 | 3.2 | 0.03815 | 0.5743 | 0.8828 |

Table 4-2. Parameters used in Hartman function with six variables.

| <i>i</i> | <i>a_{ij}</i> | | | | | | <i>c_i</i> |
|----------|-----------------------|------|------|------|------|------|----------------------|
| 1 | 10.0 | 3.0 | 17.0 | 3.5 | 1.7 | 8.0 | 1.0 |
| 2 | 0.05 | 10.0 | 17.0 | 0.1 | 8.0 | 14.0 | 1.2 |
| 3 | 3.0 | 3.5 | 1.7 | 10.0 | 17.0 | 8.0 | 3.0 |
| 4 | 17.0 | 8.0 | 0.05 | 10.0 | 0.1 | 14.0 | 3.2 |

| <i>i</i> | <i>p_{ij}</i> | | | | | | |
|----------|-----------------------|--------|--------|--------|--------|--------|--|
| 1 | 0.1312 | 0.1696 | 0.5569 | 0.0124 | 0.8283 | 0.5886 | |
| 2 | 0.2329 | 0.4135 | 0.8307 | 0.3736 | 0.1004 | 0.9991 | |
| 3 | 0.2348 | 0.1451 | 0.3522 | 0.2883 | 0.3047 | 0.6650 | |
| 4 | 0.4047 | 0.8828 | 0.8732 | 0.5743 | 0.1091 | 0.0381 | |

Table 4-3. Mean, coefficient of variation (COV), and median of different analytical functions.

| | Branin-Hoo | Camelback | Goldstein-Price | Hartman-3 | Hartman-6 |
|--------|------------|-----------|-----------------|-----------|-----------|
| Mean | 49.5 | 19.1 | 49179 | -0.8 | -0.06 |
| COV | 1.0 | 1.8 | 3.9 | -1.2 | -5.1 |
| Median | 36.7 | 11.8 | 8114 | -0.5 | -0.04 |

Table 4-4. Range of variables for radial turbine design problem.

| Variable | Description | Minimum | Maximum |
|----------|--|---------|---------|
| RPM | Rotational speed | 100000 | 150000 |
| Reaction | Percentage of stage pressure drop across rotor | 0.45 | 0.57 |
| U/Cisen | Isentropic velocity ratio | 0.56 | 0.63 |
| Tip flow | Ratio of flow parameter to a choked flow parameter | 0.30 | 0.53 |
| Dhex% | Exit hub diameter as a % of inlet diameter | 0.1 | 0.4 |
| AN2Frac | Used to calculate annulus area (stress indicator) | 0.68 | 0.85 |

Table 4-5. Numerical setup for the test problems.

| | Branin- Hoo | Camelback | Goldstein- Price | Hartman- 3 | Hartman- 6 | Radial turbine |
|---------------------|----------------|-----------|---------------------|---------------|---------------|-------------------|
| # of variables | 2 | 2 | 2 | 3 | 6 | 6 |
| # of design points | 12 | 20 | 25 | 40 | 150 | 56 |
| # of test points* | 21 | 21 | 21 | 21 | 5 | 254 |
| Order of polynomial | 2 | 3 | 3 | 3 | 3 | 2 |
| <i>Spread</i> | 0.2 | 0.3 | 0.5 | 0.4 | 0.5 | 1 |
| <i>goal</i> | 10 | 10 | 2500 | 0.05 | 0.05 | 0.01 |

*Total number of points is number of points along a direction raised to the power of the number of variables (e.g., 21^3 for Hartman problem with three variables). For the radial turbine problem, 254 indicate total number of test points. ‘spread’ controls the decay rate of radial basis function and ‘goal’ is the desired level of accuracy of the RBNN model on training points.

Table 4-6. Median, 1st, and 3rd quartile of the maximum standard deviation and actual errors in predictions of different surrogates at the location corresponding to maximum standard deviation over 1000 DOEs for different test problems.

| | Branin- Hoo | Camel back | Goldstein- Price | Hartman -3 | Hartman -6 | Radial turbine |
|--|----------------|---------------|---------------------|---------------|---------------|-------------------|
| Median (Max std dev. of response) | 105 | 53 | 2.7e5 | 2.5 | 2.2 | 0.020 |
| Median (Actual error in PRS) | 114 | 61 | 2.9e5 | 3.9 | 3.9 | 0.0016 |
| Median (Actual error in KRG) | 42 | 111 | 3.6e5 | 0.7 | 0.2 | 0.004 |
| Median (Actual error in RBNN) | 110 | 95 | 2.5e5 | 0.6 | 0.1 | 0.033 |
| 1 st /3 rd quartile (Max std dev. of response) | 77/ 134 | 38/ 85 | 1.0e5/ 4.2e5 | 2.0/ 3.2 | 1.9/ 2.7 | 0.017/ 0.022 |
| 1 st /3 rd quartile (Actual error in PRS) | 78/ 158 | 32/ 92 | 1.0e5/ 4.7e5 | 2.8/ 5.2 | 3.3/ 4.9 | 0.0008/ 0.0027 |
| 1 st /3 rd quartile (Actual error in KRG) | 21/ 71 | 66/ 131 | 1.4e5/ 6.5e5 | 0.3/ 1.4 | 0.1/ 0.4 | 0.002/ 0.006 |
| 1 st /3 rd quartile (Actual error in RBNN) | 76/ 132 | 42/ 161 | 1.9e5/ 5.7e5 | 0.3/ 1.1 | 0.1/ 0.3 | 0.028/ 0.038 |

Table 4-7. Median, 1st, and 3rd quartile of the minimum standard deviation and actual errors in predictions of different surrogates at the location corresponding to minimum standard deviation over 1000 DOEs for different test problems.

| | Branin -Hoo | Camel back | Goldstein- Price | Hartman -3 | Hartman -6 | Radial turbine |
|--|----------------|---------------|---------------------|------------------|-------------------|--------------------|
| Median (Min std dev. of response) | 0.41 | 0.26 | 492 | 0.0019 | 0.0011 | 2.1e-4 |
| Median (Actual error in PRS) | 4.7 | 1.7 | 1630 | 0.063 | 0.06 | 1.0-3 |
| Median (Actual error in KRG) | 4.6 | 1.7 | 1513 | 0.062 | 0.07 | 1.1e-3 |
| Median (Actual error in RBNN) | 4.7 | 1.7 | 1510 | 0.064 | 0.07 | 1.0e-3 |
| 1 st /3 rd quartile (Min std dev. of response) | 0.25/ 0.67 | 0.15/ 0.40 | 280/ 770 | 0.0012/ .0029 | 0.0007/ 0.0017 | 1.5e-4/ 3.2e-4 |
| 1 st /3 rd quartile (Actual error in PRS) | 1.7/ 9.8 | 0.7/ 4.4 | 697/ 3854 | 0.025/ 0.143 | 0.03/ 0.11 | 5.02e-4/ 1.9e-3 |
| 1 st /3 rd quartile (Actual error in KRG) | 1.8/ 9.9 | 0.6/ 4.2 | 525/ 3842 | 0.025/ 0.143 | 0.03/ 0.11 | 5.02e-4/ 1.9e-3 |
| 1 st /3 rd quartile (Actual error in RBNN) | 1.8/ 9.7 | 0.6/ 4.2 | 535/ 3871 | 0.024/ 0.142 | 0.03/ 0.11 | 5.02e-4/ 2.1e-3 |

Table 4-8. Median, 1st, and 3rd quartile of the maximum standard deviation and maximum actual errors in predictions of different surrogates over 1000 DOEs for different test problems. (Number after Branin-Hoo and Camelback functions indicates the number of data points used to model the function).

| | Branin- Hoo12 | Branin- Hoo31 | Camel back20 | Camel back40 | Goldstein -Price | Hartman -3 | Hartman- 6 | Radial turbine |
|--|------------------|------------------|-----------------|-----------------|---------------------|---------------|---------------|-------------------|
| Median (Min std dev. of response) | 105 | 88 | 53 | 42 | 2.7E+05 | 2.5 | 2.2 | 0.020 |
| Median (Actual error in PRS) | 175 | 32 | 122 | 37 | 4.5E+05 | 4.1 | 4.0 | 0.087 |
| Median (Actual error in KRG) | 232 | 25 | 135 | 37 | 5.3E+05 | 1.9 | 1.9 | 0.087 |
| Median (Actual error in RBNN) | 268 | 173 | 135 | 80 | 3.9E+05 | 2.3 | 1.8 | 0.082 |
| 1 st /3 rd quartile (Min std dev. of response) | 77/ 134 | 61/ 116 | 38/ 85 | 31/ 58 | 1.0e5/ 4.2e5 | 2.0/ 3.2 | 1.9/ 2.7 | 0.017/ 0.022 |
| 1 st /3 rd quartile (Actual error in PRS) | 150/ 209 | 27/ 39 | 106/ 127 | 31/ 44 | 3.7e5/ 5.5e5 | 3.2/ 5.3 | 3.4/ 4.9 | 0.082/ 0.093 |
| 1 st /3 rd quartile (Actual error in KRG) | 146/ 298 | 16/ 38 | 123/ 145 | 26/ 59 | 3.9e5/ 7.5e5 | 1.7/ 2.2 | 1.7/ 2.0 | 0.082/ 0.093 |
| 1 st /3 rd quartile (Actual error in RBNN) | 214/ 294 | 119/ 233 | 100/ 181 | 61/ 107 | 2.7e5/ 6.7e5 | 2.0/ 2.6 | 1.7/ 1.9 | 0.077/ 0.087 |

Table 4-9. Effect of design of experiment: Number of cases when an individual surrogate model yielded the least PRESS error (based on 1000 DOEs).

| | PRS | KRG | RBNN |
|-----------------|------|-----|------|
| Branin-Hoo | 715 | 131 | 154 |
| Camelback | 880 | 59 | 61 |
| Goldstein-Price | 659 | 143 | 198 |
| Hartman-3 | 229 | 511 | 260 |
| Hartman-6 | 400 | 119 | 481 |
| Radial turbine | 1000 | 0 | 0 |

Table 4-10. Opportunities of improvement *via* PWS: Number of points when individual surrogates yield errors of opposite signs (based on 1000 DOEs).

| | Mean | COV | Total # of points |
|-----------------|------|------|-------------------|
| Branin-Hoo | 218 | 0.23 | 441 |
| Camelback | 227 | 0.26 | 441 |
| Goldstein-Price | 278 | 0.17 | 441 |
| Hartman-3 | 5216 | 0.13 | 9261 |
| Hartman-6 | 9710 | 0.09 | 15625 |
| Radial turbine | 37 | 0.41 | 254 |

Table 4-11. Mean and coefficient of variation (in parenthesis) of correlation coefficient between actual and predicted response (based on 1000 DOEs) for different surrogate models.

| | PRS | KRG | RBNN | Best PRESS | PWS |
|-----------------|--------------------|--------------------|-------------------|--------------------|--------------------|
| Branin-Hoo | 0.79 (0.08) | 0.76 (0.25) | 0.75 (0.18) | 0.78 (0.12) | 0.83 (0.11) |
| Camelback | 0.69 (0.13) | 0.69 (0.19) | 0.61 (0.50) | 0.69 (0.14) | 0.73 (0.20) |
| Goldstein-Price | 0.88 (0.041) | 0.87 (0.11) | 0.86 (0.28) | 0.88 (0.083) | 0.91 (0.12) |
| Hartman-3 | 0.80 (0.073) | 0.92 (0.052) | 0.89 (0.059) | 0.89 (0.074) | 0.92 (0.028) |
| Hartman-6 | 0.56 (0.084) | 0.73 (0.12) | 0.81 (0.023) | 0.71 (0.17) | 0.77 (0.042) |
| Radial turbine | 0.9951 (0.0015) | 0.9814 (0.0088) | 0.8495 (0.062) | 0.9951 (0.0015) | 0.9946 (0.0013) |

Table 4-12. Mean and coefficient of variation (in parenthesis) of RMS errors in design space (based on 1000 instances of DOEs) for different surrogate models.

| | PRS | KRG | RBNN | Best PRESS | PWS |
|-----------------|------------------|------------------|------------------|------------------|------------------|
| Branin-Hoo | 34.4 (0.15) | 32.3 (0.38) | 37.9 (1.70) | 34.1 (0.20) | 29.1 (0.46) |
| Camelback | 22.0 (0.17) | 21.0 (0.16) | 38.0 (2.27) | 21.8 (0.17) | 20.4 (0.30) |
| Goldstein-Price | 6.72e4 (0.17) | 6.31e4 (0.33) | 1.18e5 (3.52) | 7.32e4 (3.32) | 6.28e4 (1.66) |
| Hartman-3 | 0.64 (0.20) | 0.37 (0.28) | 0.47 (0.55) | 0.44 (0.34) | 0.38 (0.16) |
| Hartman-6 | 0.45 (0.14) | 0.25 (0.13) | 0.22 (0.053) | 0.31 (0.34) | 0.25 (0.074) |
| Radial turbine | 0.0023 (0.15) | 0.0043 (0.23) | 0.0120 (0.18) | 0.0023 (0.15) | 0.0025 (0.13) |

Table 4-13. Mean and coefficient of variation (in parenthesis) of maximum absolute error in design space (based on 1000 instances of DOEs).

| | PRS | KRG | RBNN | Best PRESS | PWS |
|-----------------|------------------|------------------|------------------|------------------|------------------|
| Branin-Hoo | 182 (0.25) | 222 (0.41) | 258 (0.75) | 199 (0.29) | 202 (0.35) |
| Camelback | 127 (0.24) | 133 (0.12) | 236 (2.41) | 126 (0.23) | 128 (0.33) |
| Goldstein-Price | 4.74e5 (0.28) | 5.63e5 (0.37) | 1.08e6 (3.55) | 5.56e5 (2.96) | 5.31e5 (1.64) |
| Hartman-3 | 4.40 (0.38) | 1.94 (0.21) | 2.47 (0.86) | 2.59 (0.54) | 2.05 (0.28) |
| Hartman-6 | 4.24 (0.29) | 1.89 (0.11) | 1.84 (0.092) | 2.62 (0.43) | 1.90 (0.17) |
| Radial turbine | 0.0120 (0.28) | 0.0196 (0.21) | 0.0346 (0.22) | 0.0120 (0.28) | 0.0118 (0.20) |

Table 4-14. Mean and coefficient of variation of the ratio of RMS error and PRESS over 1000 DOEs.

| | PRS | KRG | RBNN |
|------------------|-------------|-------------|-------------|
| Branin-Hoo* | 1.02 (0.57) | 0.70 (0.60) | 0.76 (1.07) |
| Camelback | 1.26 (0.34) | 0.80 (0.31) | 0.77 (0.98) |
| Goldstein-Price* | 1.38 (0.75) | 1.04 (0.83) | 0.93 (3.33) |
| Hartman-3 | 1.31 (0.31) | 0.84 (0.33) | 0.92 (0.32) |
| Hartman-6 | 1.95 (0.12) | 1.00 (0.17) | 0.99 (0.14) |
| Radial turbine | 1.39 (0.25) | 1.02 (0.21) | 0.97 (0.14) |

*Branin-Hoo and Goldstein-Price functions had significant difference in the mean and median values of RBNN.

Table 4-15. The impact of sampling density (modeling high gradients) in approximation of Branin-Hoo function (Branin-Hoo12 is the case when we used 12 points for modeling response and Branin-Hoo31 is the case when we used 31 points to approximate function). We used 1000 DOEs samples to get mean and COV.

| | | PRS | KRG | RBNN | Best PRESS | PWS |
|--------------|--------------|------------------|------------------|-----------------|------------------|------------------|
| Correlations | Branin-Hoo12 | 0.79 (0.08) | 0.76 (0.24) | 0.75 (0.18) | 0.78 (0.12) | 0.83 (0.11) |
| | Branin-Hoo31 | 0.988 (0.003) | 0.999 (0.001) | 0.93 (0.076) | 0.998 (0.003) | 0.996 (0.014) |
| RMS Error | Branin-Hoo12 | 34 (0.15) | 32 (0.38) | 38 (1.70) | 34 (0.20) | 29 (0.46) |
| | Branin-Hoo31 | 7.9 (0.11) | 2.4 (0.53) | 20.3 (1.27) | 2.7 (0.64) | 4.3 (0.54) |
| Max Error | Branin-Hoo12 | 182 (0.25) | 222 (0.41) | 258 (0.75) | 199 (0.29) | 202 (0.35) |
| | Branin-Hoo31 | 34 (0.31) | 30 (0.63) | 183 (0.80) | 31 (0.60) | 41 (0.53) |

Table 4-16. The impact of sampling density (modeling high gradients) in approximation of Camelback function (Camelback20 is the case when we used 20 points for modeling response and Camelback40 is the case when we used 40 points to approximate function). We used 1000 DOEs to get mean and COV.

| | | PRS | KRG | RBNN | Best PRESS | PWS |
|--------------|-------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Correlations | Camelback20 | 0.69 (0.13) | 0.69 (0.19) | 0.61 (0.50) | 0.69 (0.14) | 0.73 (0.20) |
| | Camelback40 | 0.97 (0.010) | 0.98 (0.039) | 0.92 (0.080) | 0.98 (0.015) | 0.98 (0.010) |
| RMS Error | Camelback20 | 22 (0.17) | 21 (0.16) | 38 (2.27) | 22 (0.17) | 20 (0.30) |
| | Camelback40 | 7.3 (0.15) | 4.9 (0.74) | 12 (0.35) | 5.3 (0.42) | 5.7 (0.27) |
| Max Error | Camelback20 | 127 (0.24) | 133 (0.12) | 236 (2.41) | 126 (0.23) | 128 (0.33) |
| | Camelback40 | 39 (0.34) | 48 (0.64) | 90 (0.52) | 40 (0.47) | 43 (0.39) |

Table 4-17. Effect of parameters in parametric surrogate filter used for PWS. Three settings of parameters α and β were selected. We show median, 1st, and 3rd quartile data based on 1000 DOEs for Goldstein-Price problem.

| | | PRS | KRG | RBNN | Best PRESS | PWS, $\alpha=0.05,$ $\beta=-1$ | PWS, $\alpha=0.5,$ $\beta=-1$ | PWS, $\alpha=0.05,$ $\beta=-5$ |
|--------------|--------------------------|--------|--------|--------|------------|--------------------------------------|-------------------------------------|--------------------------------------|
| Correlations | Median | 0.89 | 0.90 | 0.94 | 0.89 | 0.93 | 0.94 | 0.93 |
| | 1 st quartile | 0.86 | 0.84 | 0.88 | 0.86 | 0.90 | 0.91 | 0.91 |
| | 3 rd quartile | 0.90 | 0.94 | 0.97 | 0.91 | 0.95 | 0.96 | 0.94 |
| RMS Error | Median | 6.45e4 | 5.84e4 | 4.57e4 | 6.21e4 | 5.15e4 | 4.88e4 | 5.20e4 |
| | 1 st quartile | 5.85e4 | 4.74e4 | 3.56e4 | 5.44e4 | 4.31e4 | 4.11e4 | 4.55e4 |
| | 3 rd quartile | 7.27e4 | 7.57e4 | 6.98e4 | 7.21e4 | 6.37e4 | 6.14e4 | 6.29e4 |
| Max Error | Median | 4.52e5 | 5.32e5 | 3.88e5 | 4.54e5 | 4.35e5 | 4.32e5 | 4.52e5 |
| | 1 st quartile | 3.74e5 | 3.91e5 | 2.65e5 | 3.56e5 | 3.38e5 | 3.33e5 | 3.47e5 |
| | 3 rd quartile | 5.52e5 | 7.49e5 | 6.68e5 | 5.87e5 | 5.75e5 | 5.73e5 | 5.82e5 |

CHAPTER 5 ACCURACY OF ERROR ESTIMATES FOR SURROGATE APPROXIMATION OF NOISE- FREE FUNCTIONS

Introduction

Surrogate based design and optimization process is very attractive for computationally expensive problems. We construct surrogate models to evaluate the performance of designs using a limited amount of data, and couple surrogate models with optimization methods. Different error measures are used to assess the quality of such surrogate models (Queipo et al., 2005). In addition, these measures are used for determining the sampling locations in many surrogate-based optimization methods like EGO (Jones et al., 1998), and for adaptive sampling. The success of these approaches depends on the ability of these error estimation measures to reflect the true errors.

Error measures can be broadly classified as model based or model independent. Model based error measures are typically based on some statistical assumptions. For example, prediction variance for polynomial response surface approximation is developed assuming that the data used to construct the surrogate contain noise that is normally distributed with zero mean and same variance σ^2 in all data and no correlation. When these statistical assumptions are not satisfied, as is usually the case, the accuracy of the resulting error estimates is questionable. Model-independent error measures are not based on any statistical assumption. However, they may have high computational cost.

Error measures can further be characterized as global or local error estimates. Global error measures provide a single number for the entire design space, for example, process variance (model based) for kriging or PRESS (model independent) estimate for polynomial response surface approximation. Local or pointwise error measures estimate error at any point in the entire design space. Examples of model based pointwise error measures are standard error for

polynomial response surfaces, and mean square errors for kriging, etc. Recently, Goel et al. (2006b) proposed using standard deviation of responses as a model independent pointwise error measure.

The primary goal of this chapter is to appraise the performance of different error measures with the help of a variety of test problems in context of simulation based surrogate approximation. We account for the influence of experimental designs on error estimates by considering a large number of samples obtained using Latin hypercube sampling and D-optimality criterion. Noting the difficulties in identifying regions of high uncertainty using a single error estimation measure, we explore the idea of simultaneously using multiple error measures that is motivated from our previous work on ensemble of surrogates (Goel et al., 2006b). Here, we examine the ideas of: (1) combining different error measures to better estimate actual errors, (2) simultaneously using multiple error measures to increase the probability of identifying regions of high uncertainty, and (3) identifying the appropriate error measure for a given problem.

The chapter is organized as follows. We briefly describe relevant error estimation measures in next section. Description of different test problems and numerical procedures followed are delineated next. Then, we present the results of appraisal of different error estimators for different test problems. Subsequently, we demonstrate the concept of simultaneous application of multiple error estimators. Finally, we recapitulate the major conclusions.

Error Estimation Measures

Error in approximation is defined as the difference between actual and predicted response. We can compute the error at each design point, but the quality of different approximations is meaningfully compared by looking at the global prediction metrics, for example, average error, maximum error, or root mean square error in the entire design space. The choice of appropriate

measure depends on the application. Maximum error may be more important for design of critical components but for most applications root mean squared error in the entire design space serves as a good measure of accuracy of approximation.

The main issue in computation of errors is that the actual response at any point is not known (that is the primary reason of developing surrogates). Consequently, we cannot assess actual errors but we use different error measures to estimate either pointwise bounds or root mean square of actual errors (local error measures) or to assess the root mean square error in the entire design space (global error measures). Generally, root mean square error estimation measures are more popular and more practical than the bounds on the errors. These error measures are usually based on certain assumptions on the actual functions, for example noise in data or statistical distribution of coefficient vector for polynomial response surface approximations, Gaussian process for kriging, etc. Now, error measures that do not make any assumption on the data or surrogate model are also being developed.

In this section, we briefly describe the formulation of relevant global and local error estimation measures for popular surrogate models (based on assumptions on data or surrogate model) as well as some model independent error measures.

Error Measures for Polynomial Response Surface Approximation

Polynomial response surface approximation (PRS, Myers and Montgomery 1995) is the most popular surrogate model among practitioners. The observed response $y(\mathbf{x})$ of a function at point \mathbf{x} is represented as a linear combination of basis function vector $\mathbf{f}(\mathbf{x})$ (mostly monomials are selected as basis functions), and true coefficient vector $\boldsymbol{\beta}$, and error ε . Error in the approximation ε is assumed to be uncorrelated, and normally distributed with zero mean and σ^2 variance. That is,

$$y(\mathbf{x}) = \mathbf{f}^T \boldsymbol{\beta} + \varepsilon; \quad E(\varepsilon) = 0, \quad V(\varepsilon) = \sigma^2. \quad (5.1)$$

The vector $\mathbf{f}(\mathbf{x})$ has two components: $\mathbf{f}^{(1)}(\mathbf{x})$ is the vector of basis functions used in the polynomial response surface model, and $\mathbf{f}^{(2)}(\mathbf{x})$ is the vector of additional basis functions that are missing in the linear regression model. Similarly, the coefficient vector $\boldsymbol{\beta}$ can be written as a combination of vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ that represent the true coefficients associated with the basis function vectors $\mathbf{f}^{(1)}(\mathbf{x})$ and $\mathbf{f}^{(2)}(\mathbf{x})$, respectively. Then,

$$y(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}^{(2)} + \varepsilon. \quad (5.2)$$

The polynomial response surface approximation of the observed response $y(\mathbf{x})$ is,

$$\hat{y}(\mathbf{x}) = \sum_j b_j f_j^{(1)}(\mathbf{x}), \quad (5.3)$$

where b_j is the estimated value of the coefficient associated with the j^{th} basis function $f_j^{(1)}(\mathbf{x})$.

Then, the error in approximation at i^{th} design point is given as $e(\mathbf{x}) = y_i - \hat{y}_i$. The coefficient

vector \mathbf{b} can be obtained by minimizing a loss function L , defined as

$$L = \sum_{i=1}^{N_s} |e_i|^p, \quad (5.4)$$

where p is the order of loss function, and N_s is the number of sampled design points.

We use the conventional quadratic loss function ($p = 2$) that minimizes the variance of the error in approximation, unless specified otherwise. This loss function is most popular because we can obtain coefficient vector \mathbf{b} from the solution of a linear system of algebraic equations,

$$\mathbf{b} = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} \mathbf{y}, \quad (5.5)$$

where $X^{(1)}$ is the Gramian design matrix constructed using basis functions $\mathbf{f}^{(1)}(\mathbf{x})$, and \mathbf{y} is the vector of responses at N_s design points. The estimated variance of noise (or root mean square error) in approximation is given by,

$$\sigma_a^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}})}{(N_s - N_{\beta_1})}, \quad (5.6)$$

where N_{β_1} is the number of coefficients in polynomial response surface model. For more details on polynomial response surface approximation, we refer the reader to the texts by Myers and Montgomery (1995), and Khuri and Cornell (1996). The square root of estimated variance of noise is often used as an estimate of root mean square of actual error in approximation in the entire design space (global error measure).

The absolute error in PRS approximation at a point \mathbf{x} is given as $e_{prs}(\mathbf{x}) = |y(\mathbf{x}) - \hat{y}(\mathbf{x})|$.

Two models to predict error in PRS approximation are explained as follows.

Estimated standard error (ESE, Myers and Montgomery, 1995)

When the response surface model $\mathbf{f}^{(1)}(\mathbf{x})$ and the true function $\mathbf{f}(\mathbf{x})$ are the same (Equation (5.2)), the standard error $e_{es}(\mathbf{x})$ is used to characterize approximation error due to noise. This root mean square error estimate is given as,

$$e_{es}(\mathbf{x}) = \sqrt{\text{Var}[\hat{y}(\mathbf{x})]} = \sqrt{\sigma_a^2 \mathbf{f}^{(1)T}(\mathbf{x}) (X^{(1)T} X^{(1)})^{-1} \mathbf{f}^{(1)}(\mathbf{x})}, \quad (5.7)$$

where σ_a^2 is the estimated variance of noise, $\mathbf{f}^{(1)}(\mathbf{x})$ is the vector of basis functions used in approximation, and $X^{(1)}$ is the matrix of linear equations constructed using $\mathbf{f}^{(1)}(\mathbf{x})$.

Root mean square bias error (RMSBE, Goel et al., 2006c, Appendix A)

When noise is small ($\varepsilon \rightarrow 0$, Equation (5.2)), the bias error or modeling error $e_b(\mathbf{x})$ appears due to approximating a higher order polynomial by a lower order function. The bias error is given by,

$$e_b(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} - (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}, \quad (5.8)$$

where $\mathbf{f}(\mathbf{x})$ is the vector of basis functions in assumed true function, and $\boldsymbol{\beta}$ is the coefficient vector associated with $\mathbf{f}(\mathbf{x})$. Since, the coefficient vector $\boldsymbol{\beta}$ is unknown, we estimate the root mean square of the bias error $e_b^{rms}(\mathbf{x})$ by assuming appropriate statistical distribution of the coefficient vector $\boldsymbol{\beta}$, as follows.

$$e_b^{rms}(\mathbf{x}) = \sqrt{E_{\beta}(e_b^2(\mathbf{x}))}, \quad (5.9)$$

$$\begin{aligned} E_{\beta}(e_b^2(\mathbf{x})) &= (\mathbf{f}^{(1)}(\mathbf{x}))^T E_{\beta} \left((\boldsymbol{\beta}^{(1)} - \mathbf{b})(\boldsymbol{\beta}^{(1)} - \mathbf{b})^T \right) (\mathbf{f}^{(1)}(\mathbf{x})) + \\ & (\mathbf{f}^{(2)}(\mathbf{x}))^T E_{\beta} \left(\boldsymbol{\beta}^{(2)} (\boldsymbol{\beta}^{(1)} - \mathbf{b})^T \right) (\mathbf{f}^{(1)}(\mathbf{x})) + \\ & (\mathbf{f}^{(1)}(\mathbf{x}))^T E_{\beta} \left((\boldsymbol{\beta}^{(1)} - \mathbf{b}) \boldsymbol{\beta}^{(2)T} \right) (\mathbf{f}^{(2)}(\mathbf{x})) + (\mathbf{f}^{(2)}(\mathbf{x}))^T E_{\beta} \left(\boldsymbol{\beta}^{(2)} \boldsymbol{\beta}^{(2)T} \right) (\mathbf{f}^{(2)}(\mathbf{x})), \end{aligned} \quad (5.10)$$

$$E_{\beta}(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) = \beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \|\boldsymbol{\beta}_{\varepsilon}\|^2 \left(\frac{\gamma^2}{3} \right) \sum_{p=1}^{N_e} V_{pj}^{(1)} V_{pi}^{(2)}, \quad (5.11)$$

where $\|\boldsymbol{\beta}_{\varepsilon}\|$ is the norm of the solution $\boldsymbol{\beta}_{\varepsilon}$ (of system of linear equations $X\boldsymbol{\beta} = \mathbf{y}$) that satisfies the data \mathbf{y} with least deviation, γ is a constant used to define the distribution of the coefficient vector $\boldsymbol{\beta}$, N_e is the number of null eigenvectors of matrix of linear equations X constructed using $\mathbf{f}(\mathbf{x})$, \mathbf{V}_i represents the i^{th} null eigenvector of X , and $E_y(g(\beta))$ is the expected value of $g(\beta)$ with respect to random variable y .

Note that, the two pointwise error estimation measures, described above, characterize pointwise root mean square actual error.

Error Measures for Kriging

Kriging (KRG) is named after the pioneering work of D.G. Krige (a South African mining engineer) and is formally developed by Matheron in 1963. Kriging estimates the value of an objective function $y(\mathbf{x})$ at design point \mathbf{x} as the sum of a polynomial trend model $\sum_j \beta_j f_j(\mathbf{x})$, and

a systematic departure $Z(\mathbf{x})$ representing low (large scale) and high frequency (small scale) variations around the trend model,

$$y(\mathbf{x}) = \hat{y}(\mathbf{x}) = \sum_j \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}). \quad (5.12)$$

The systematic departure components are assumed to be correlated as a function of distance between the locations under consideration. The Gaussian function is the most commonly used correlation function. The Gaussian correlation between point \mathbf{x} and the i^{th} design point $\mathbf{x}^{(i)}$ is given below.

$$r(Z(\mathbf{x}), Z(\mathbf{x}^{(i)}), \boldsymbol{\theta}) = \prod_{j=1}^{N_y} \exp(-\theta_j (x_j - x_j^{(i)})^2). \quad (5.13)$$

Martin and Simpson (2005) compared the maximum likelihood approach and a cross-validation based approach to estimate kriging parameters β_j, θ_j and found that the maximum likelihood approach to estimate kriging parameters was the best. So we adopt the maximum likelihood approach to estimate kriging parameters β_j, θ_j . The predicted response at point \mathbf{x} is given as follows.

$$\hat{y}(\mathbf{x}) = (\mathbf{r}(\mathbf{x}))^T R^{-1} \mathbf{y} - (X^{(1)T} R^{-1} \mathbf{r}(\mathbf{x}) - \mathbf{f}^{(1)}(\mathbf{x}))^T (X^{(1)T} R^{-1} X^{(1)})^{-1} X^{(1)T} R^{-1} \mathbf{y}, \quad (5.14)$$

where $\mathbf{r}(\mathbf{x})$ is the vector of correlations between point \mathbf{x} and design points (whose components are given by Equation (5.13)), R is the matrix of correlations among design points, \mathbf{y} is the vector of predicted responses at design points, $X^{(1)}$ is the Gramian design matrix constructed using basis functions in trend model at design points, and $\mathbf{f}^{(1)}(\mathbf{x})$ is the vector of basis functions in trend model at point \mathbf{x} . Note that, kriging is an interpolating function that reproduces the data but may yield errors at locations not used to construct the surrogate model.

The estimated process variance associated with the kriging approximation is given as,

$$\sigma^2 = \frac{1}{N_s} (\mathbf{y} - X^{(1)} \boldsymbol{\beta}^*)^T R^{-1} (\mathbf{y} - X^{(1)} \boldsymbol{\beta}^*), \quad (5.15)$$

where $\boldsymbol{\beta}^*$ is the approximation of the coefficient vector $\boldsymbol{\beta}$ in Equation (5.12), and is given as $\boldsymbol{\beta}^* = (X^{(1)T} R^{-1} X^{(1)})^{-1} X^{(1)T} R^{-1} \mathbf{y}$. Process variance should be low for a good approximation. We explore if this error measure can be used as a global estimate of root mean square of actual error in the entire design space. The pointwise estimate of actual error in kriging approximation is given by computing the mean squared error $\varphi(\mathbf{x})$ as follows.

$$\varphi(\mathbf{x}) = \sigma^2 \left(\mathbf{1} + \mathbf{u}(\mathbf{x})^T (X^{(1)T} R^{-1} X^{(1)})^{-1} \mathbf{u}(\mathbf{x}) - \mathbf{r}(\mathbf{x})^T R^{-1} \mathbf{r}(\mathbf{x}) \right), \quad (5.16)$$

$$\mathbf{u}(\mathbf{x}) = X^{(1)T} R^{-1} \mathbf{r}(\mathbf{x}) - \mathbf{f}^{(1)}(\mathbf{x}), \quad (5.17)$$

where σ^2 is the process variance (Equation (5.15)), and $\mathbf{1}$ is the vector of ones. As was the case for error estimation measures for polynomial response surface approximation, mean squared error also is a root mean square error estimate of the actual error. So to compare with the actual error, we use the square root of mean square error (MSE).

$$e_{mse}(\mathbf{x}) = \sqrt{\varphi(\mathbf{x})}, \quad (5.18)$$

Model Independent Error Estimation Models

While the error estimation measures discussed so far are suitable for particular surrogate models, we discuss error measures that can be used with any surrogate model (model independent error measures).

Generalized cross-validation error (GCV)

Generalized cross-validation error (GCV), also known as PRESS (predicted residual sum of squares) in the polynomial response surface approximation terminology ($GCV = PRESS / N_s$), is estimated by using data at N_s points as follows. We fit surrogate

models to $N_s - 1$ points by leaving one design point at a time, and predict response at the left out point. Then, GCV is defined as,

$$GCV = \frac{1}{N_s} \sum_{i=1}^{N_s} (y_i - \hat{y}_i^{(-i)})^2, \quad (5.19)$$

where $\hat{y}_i^{(-i)}$ represents the prediction at $\mathbf{x}^{(i)}$ using the surrogate constructed using all design points, except $(\mathbf{x}^{(i)}, y_i)$. We use the square root of generalized cross-validation error to estimate actual root mean square error in approximation and compare its performance with other global error estimation measures. Even though we have shown a global GCV, a local counterpart of the GCV can also be developed.

The analytical estimate of GCV is available for polynomial response surface approximation (Myers and Montgomery, 1995). Mitchell and Morris (1992) and Currin et al. (1998) provided computationally inexpensive expressions to evaluate cross-validation error for kriging using a constant trend model while holding other model parameters constant. Martin (2005) extended this analytical estimate of cross-validation error to account for more complex trend functions while keeping the model parameters constant. However, here we use a first principle based method to estimate GCV for kriging.

Standard deviation of responses (s_{resp})

Goel et al. (2006b) have recently showed that we can estimate uncertainty in predictions for any surrogate model by using an ensemble of surrogates, and that a convex combination of surrogates provides a more robust approximation than the individual surrogates. The main idea is as follows. Let there be N_{SM} surrogate models such that there are N_{SM} predictions (\hat{y}_i) at any test point. We compute the standard deviation of responses as,

$$s_{resp}(\hat{y}(\mathbf{x})) = \sqrt{\frac{1}{N_{SM}-1} \sum_{i=1}^{N_{SM}} (\hat{y}_i(\mathbf{x}) - \bar{y}(\mathbf{x}))^2}, \quad (5.20)$$

where $\bar{y}(\mathbf{x}) = \frac{1}{N_{SM}} \sum_{i=1}^{N_{SM}} \hat{y}_i(\mathbf{x})$.

In this study, we compute s_{resp} using four surrogate models, kriging, radial basis neural network (Orr, 1996), and two polynomial response surface approximations; one with quadratic loss function ($p=2$, Equation (5.4)), and second with sixth order loss function ($p=6$, Equation (5.4)).

The different error estimation measures used in this paper are summarized in Table 5-1.

PRESS-based weighted average surrogate (PWS)

We construct a PRESS-based weighted average surrogate (PWS) using multiple approximations as follows (Goel et al., 2006b). The predicted response $\hat{y}_{pws}(\mathbf{x})$ at a data point using PWS is given as,

$$\hat{y}_{pws}(\mathbf{x}) = \sum_i^{N_{SM}} w_i \hat{y}_i(\mathbf{x}), \quad (5.21)$$

where N_{SM} is the number of surrogate models, $\hat{y}_i(\mathbf{x})$ is the predicted response by the i^{th} surrogate model, and w_i is the weight associated with the i^{th} surrogate model at design point \mathbf{x} .

The weights are determined as follows.

$$w_i^* = \left(\frac{E_i}{E_{avg}} + \alpha \right)^\beta, \quad w_i = \frac{w_i^*}{\sum_i w_i^*}, \quad (5.22)$$

$$E_{avg} = \sum_{i=1}^{N_{SM}} E_i / N_{SM}; \quad \beta < 0, \alpha < 1,$$

where E_i is the square root of generalized cross-validation error (GCV) for i^{th} surrogate model.

We use $\alpha = 0.05$ and $\beta = -1$ in this study. More details about PWS can be found in Chapter 4.

Ensemble of Error Estimation Measures

Finally, we discuss the concept of an ensemble of error measures that is inspired by our previous work on ensemble of surrogates (Goel et al., 2006b). Since there are many error estimation measures for different surrogates, we can simultaneously use the information to estimate actual errors. In this context, we explore three approaches as follows.

Averaging of multiple error measures

A simple way of combining different error measures is to take an arithmetic or geometric average of appropriate error models. To this end, we can combine standard error with root mean square bias error for polynomial response surface approximation, and mean square error with standard deviation of responses for kriging. We furnish details about averaging of error measures in the results section.

Identification of best error measure

We cannot determine the suitability of any error measure *a priori* without testing the performance against actual data. However, we can assess the performance of various error estimation measures for a given problem by using the information obtained from the global cross-validation (GCV). The trends observed in comparing the actual error and the predicted error at the omitted design points are likely to reflect the accuracy of error estimation measure(s) at other points. This GCV based method to identify the appropriate error measure is computationally very attractive since the information required to make the decision does not require any additional simulations. The details of the procedure are as follows.

We fit surrogate models by leaving one design point at a time. We estimate responses, actual and predicted errors, and standard deviation of responses at the design point not used in constructing the surrogate model. Now, we characterize the performance of different error estimation measures by comparing the actual errors with the predicted errors according to an

appropriate criterion. A few relevant criteria are the correlation between predicted and actual error, the ratio of actual and predicted root mean square error, and the ratio of maximum actual and predicted error. The best of all error measures can be used to characterize the actual errors in entire design space.

Simultaneous application of multiple error measures

The primary application of error estimation measures is to identify the locations where predictions may have high errors. So the risk associated with the application of error estimation measures is the inability to identify the high error zones. We propose to simultaneously use different error estimation measures to reduce the probability of missing regions of high errors. This is accomplished by considering a region to have high errors if at least one error estimation model identified it as a high error region.

Global Prediction Metrics

We compare different global and pointwise error metrics with actual errors using test problems described in next section. The global error measures (root mean square error, process variance, and GCV) for different surrogates are compared with the respective actual root mean square (RMS) errors in approximation. Since our goal is to assess the overall capabilities of different pointwise error measures, we use following global metrics for comparison.

Root mean square error

Pointwise error estimates in design space are converted into global measures like root mean square error (RMSE), as follows,

$$RMSE = \sqrt{\frac{1}{V} \int_V e^2 dV}, \quad (5.23)$$

where e is the error estimate. In low dimensional spaces, we used a uniform grid of test points so the root mean square error is numerically evaluated by implementing quadrature⁴ for integration (Ueberhuber, 1997) as,

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N_{test}} \xi_j e_j^2}{N_{test}}}, \quad (5.24)$$

where ξ_j is the weight used for integration, and N_{test} is the number of test points. For high dimensional problems, where using a uniform grid of test points is computationally expensive, we used a quasi-random set of test points so we choose $\xi_j = 1$ to get root mean square error. We compare the predicted and actual RMSE.

Correlation between predicted and actual errors

Correlation between pointwise actual absolute errors and predicted errors at the test points $r(e, \hat{e})$ is given as

$$r(e, \hat{e}) = \frac{\frac{1}{V} \int_V (e - \bar{e})(\hat{e} - \bar{\hat{e}}) dV}{\sigma(e)\sigma(\hat{e})}. \quad (5.25)$$

where e is the actual error, \hat{e} is the predicted error, \bar{e} is the mean of actual error, $\sigma(e)$ is the variance of the actual error, $\bar{\hat{e}}$ is the mean of predicted error, and $\sigma(\hat{e})$ is the variance of predicted error. The correlation coefficient is numerically evaluated from the data for test points by implementing quadrature for integration (Ueberhuber, 1997) as given in Equation (5.26).

⁴ We used trapezoidal rule for integration.

$$\frac{1}{V} \int_V e \hat{e} \, dV = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \xi_i e_i \hat{e}_i; \quad \bar{e} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \xi_i e_i, \quad (5.26)$$

$$\sigma(e) = \sqrt{\frac{1}{V} \int_V (e - \bar{e})^2 \, dV} = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \xi_i (e_i - \bar{e})^2},$$

where ξ_i is the weight associated with trapezoidal integration rule when uniform grid is used as test points, and N_{test} is the number of test points. As for root mean square estimate for five- and six-variable problems we choose $\xi_i = 1$ to estimate correlation using Equation (5.26). For a high quality error prediction measure, the correlation coefficient should be as high as possible. The maximum value of $r(e, \hat{e})$ is one that defines exact linear relationship between the predicted and the actual errors.

Maximum absolute error

Maximum error is computed using the error data at all test points. We compare the magnitude of maximum actual absolute error with maximum predicted error.

Test Problems and Testing Procedure

Test Problems

To test the predictive capabilities of different error estimators, we employ two types of problems ranging from two to six variables: (1) analytical functions (Dixon-Szegö, 1978) that are often used to test the global optimization methods, and (2) engineering problems: a radial turbine design problem (Mack et al., 2006) that is a new concept design, and cantilever beam design problem (Wu et al., 2001) that is extensively used as a test problem in reliability analysis. The details of each test problem are given as follows.

Branin-Hoo function

$$f(x_1, x_2) = \left(x_2 - 5.1x_1^2 / 4\pi^2 + 5x_1 / \pi - 6 \right)^2 + 10 \left(1 - 1/8\pi \right) \cos(x_1) + 10, \quad (5.27)$$

$$x_1 \in [-5, 10], \quad x_2 \in [0, 15].$$

Camelback function

$$f(x_1, x_2) = \left(4 - 2.1x_1^2 + x_1^4 / 3 \right) x_1^2 + x_1 x_2 + (-4 + 4x_2^2) x_2^2, \quad (5.28)$$

$$x_1 \in [-3, 3], \quad x_2 \in [-2, 2].$$

Goldstein-Price function

$$f(x_1, x_2) = \left[1 + (x_1 + x_2 + 1)^2 (19 - 4x_1 + 3x_1^2 - 14x_2 + 6x_1 x_2 + 3x_2^2) \right] \times \\ \left[30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1 x_2 + 27x_2^2) \right], \quad (5.29)$$

$$x_1, x_2 \in [-2, 2].$$

The graphical representation of these two-variable test problems is given in Figure 5-1 illustrating zones of high gradients.

Hartman functions

$$f(\mathbf{x}) = -\sum_{i=1}^m c_i \exp \left\{ -\sum_{j=1}^{N_v} a_{ij} (x_j - p_{ij})^2 \right\}, \quad (5.30)$$

$$\text{where } \mathbf{x} = (x_1, x_2, \dots, x_{N_v}) \quad x_i \in [0, 1].$$

Two instances of this problem are considered based on the number of design variables. For the chosen examples, $m = 4$.

- **Hartman-3**

This problem has three variables. The choice of parameters is given in Table 5-2 (Dixon-Szegö, 1978).

- **Hartman-6**

This instance of the problem has six design variables and the parameters used in the function are tabulated in Table 5-3 (Dixon-Szegö, 1978). For this case, all design variables were in range [0, 0.5] instead of [0, 1].

Radial turbine design problem

As described by Mack et al. (2006), this six-variable problem is a new conceptual design of a compact radial turbine used to drive pumps that deliver liquid hydrogen and liquid oxygen to the combustion chamber of a spacecraft. The objective of the design is to increase the efficiency of a turbine in the liquid rocket expander cycle engine, while keeping the overall weight of the turbine low. Our interest in this problem is to develop accurate surrogate model(s) of the efficiency as a function of six design variables. The description of design variables and their corresponding ranges are given in Table 5-4 (Mack et al., 2006). The objectives of the design were calculated using a one-dimensional flow analysis ‘meanline’ code (Huber, 2001). Mack et al. (2006) identified the appropriate region of interest by iteratively refining the design space. They also identified the most important variables using global sensitivity analysis.

Cantilever beam design problem (Wu et al., 2001)

The cantilever beam design problem (Figure 5-2) that is widely used for reliability analysis is illustrated as follows. The displacement of the beam, G_d , is given as:

$$G_d = D_0 - D = D_0 - \frac{4L^3}{Ewt} \sqrt{\left(\left(\frac{F_y}{t^2}\right)^2 + \left(\frac{F_x}{w^2}\right)^2\right)}, \quad (5.31)$$

where D_0 is the allowable initial deflection of the beam (taken as 2.25”), E is the Young’s modulus, L , w , and, t , are the length, width, and thickness of the beam, respectively. The length of the beam is fixed at 100” and the displacement G_d is approximated using different surrogate models. Based on the number of design variables, two instances of this problem are given as follows.

- **Two variables**

The design variables are the horizontal loading (F_x), and the vertical loading (F_y). In this case, we use the Young modulus $E = 29e06$ psi, width $w = 2.65354''$, and thickness $t = 3.97917''$. The ranges of the design variables are $F_x \in [700, 1300] lbs$ and

$$F_y \in [900, 1300] lbs .$$

- **Five variables**

The displacement is given as a function of the horizontal loading (F_x), vertical loading (F_y), Young's modulus (E), width (w), and thickness (t). The ranges of the design variables are given in Table 5-5.

Testing Procedure

The main steps in the estimation and testing of errors in surrogate modeling are given as follows.

Design of experiments (DOE)

As we showed in Chapter 3 (Goel et al., 2007b), a combination of Latin hypercube sampling (LHS) and D-optimality criterion (Myers and Montgomery, 1995) to construct design of experiments is helpful in reducing noise and bias errors. In this procedure, we generate a large sample of points using LHS, and then pick the most optimal configuration for the desired order of polynomial using D-optimality criterion without using replicate points.

For all analytical and cantilever beam design problems, we adopt this LHS+D-optimality criterion based approach to sample design points for surrogate construction. The number of design points (N_s), order of polynomial for D-optimality criterion, and the number of points used in the large LHS sample (N_{lhs}) are given in Table 5-6. The large optimal LHS sample is obtained using MATLAB routine 'lhsdesign' with 'maximin' criterion (maximize the minimum distance between points). We allocate a maximum of 100 iterations for optimization. This LHS design serves as the initial grid of points for MATLAB routine 'candexch' to select N_s design points

without duplication, using D-optimality criterion. We allow a maximum of 40 iterations for optimization.

For the radial turbine design problem, Mack et al. (2006) sampled 232 designs in the six-dimensional region of interest using a fractional factorial design. Eight designs were found infeasible, and the 224 feasible design points were used to construct, and to test different surrogate models and error estimation measures. In this study, we randomly select 56 points to construct surrogate models, and the remaining 168 points are used for testing.

We note that there is uncertainty associated with the DOE due to random components, and the possibility of convergence to local optima for both LHS and D-optimal design. To reduce the uncertainty due to choice of DOE, we present results based on 1000 instances of design of experiments for all problems.

Test points

For two- and three- dimensional spaces, the surrogate models were tested using a uniform grid of p^{N_v} points where N_v is the number of variables, and p is the number of points along each direction (given in Table 5-6). We used 2000 points, selected *via* optimized LHS, to assess the performance of error estimation measures for Cantilever-5 and Hartman-6 problems. For radial turbine example, we compared error estimation measures using 168 points that were not used to construct surrogate models.

Surrogate construction

The order of polynomial used for polynomial response surface approximation is given in Table 5-6. While we used Equation (5.5) for PRS approximation in least squares sense, we employed MATLAB routine 'lsqnonlin' to construct PRS with sixth order loss function (Equation (5.4)). We used the kriging software developed by Lophaven et al. (2002) with a linear

trend model, and the Gaussian correlation function. The bounds on the parameters governing the Gaussian distribution were taken as [0.01, 200]. Radial basis neural network was constructed using MATLAB routine 'newrb', with the parameter 'spread' specified as 0.5, and the mean square error 'goal' was taken as square of five-percent of the mean value of the response at training points. For all problems, we normalized design variables between zero and one, such that the minimum value of any variable was scaled to zero. The surrogate models were constructed in normalized design space.

Error estimation

Firstly, we computed pointwise actual absolute errors at test points for different surrogate models. Next, we estimated global error measures, namely, estimated root mean square error for PRS (Equation (5.6)), process variance for kriging (Equation (5.15)), and generalized cross-validation errors for each surrogate model using Equation (5.19). Finally, we evaluated pointwise errors namely, standard error (Equation (5.7)), and RMS bias error (Equation (5.9)) for PRS, mean square error (Equation (5.18)) for kriging, and standard deviation of responses (Equation (5.20)) using four surrogates, two instances of polynomial response surfaces with loss functions $p = 2$ and $p = 6$, respectively, kriging, and radial basis neural network. We used $\gamma = 0.2$ for bias error computations, and the true function was assumed to be the lowest order polynomial that had more coefficient than the number of data points (N_s). We computed maximum and RMS values of pointwise error measures as defined earlier.

Results: Accuracy of Error Estimates

Firstly, we normalized actual RMS error in approximation by the range of any individual function in the entire design space, and show results in Table 5-7. A low value of normalized actual RMS error indicates good approximation of the actual function. Polynomial response surface approximation using a quadratic loss function was a good surrogate for all problems,

except the Hartman problem with three variables (10% or more errors in approximation). Kriging performed reasonably well for all problems, except the Camelback function and the Hartman problem with six variables. RBNN was the poorest of all surrogates overall, but it was the best surrogate for Goldstein-Price problem. Polynomial response surface approximation using a six-order loss function was slightly worse compared to the quadratic loss function. For radial turbine design problem, the performance of PRS with six-order loss function was particularly poor and, this resulted in the poor performance of PWS model (PRS with $p=6$, was assigned a high weight due to a good PRESS). For all problems, except radial turbine design problem, PWS model could significantly reduce the influence of the worst surrogate. Also notably, the PRS models yielded much lower coefficient of variation compared to other surrogate models. The results suggested that the PRS model is indeed a more reliable approximation. The higher variability in kriging approximation is attributed to the variability in estimating the model parameters using the maximum likelihood estimates. The variability in RBNN arises from the complex fitting process that results in different number and location of neurons with each DOE.

Next, we compared the accuracy of different types of error estimates on various test problems in following two subsections. First subsection presents results of comparison of global error estimators with actual errors. We compared the performance of local (pointwise) error estimators in second subsection.

Global Error Measures

We assessed the accuracy of global error estimates by computing the ratio of global error estimates and relevant actual root mean square error in approximation. The desired value of this ratio is one. A value of less than one indicates that the actual RMS error is underestimated by the error predictor, and a value greater than one means that the global error estimator overestimates the magnitude of actual RMS error. We summarized the results based on 1000 DOEs in Figure

5-3 using boxplots of different ratios. The box encompasses the 1st and 3rd quartile of the data. Median of the data is shown by the line inside the box. The leader lines on the ends are placed at either the minimum/maximum of the data or at a distance of 1.5 times the width of the box. The data points that fall beyond the range of leader lines are shown by a symbol ‘+’.

The mean and coefficient of variation (COV) for different problems are given in Table 5-8. We observed that the global error estimates for polynomial response surface approximation (PRS) provided more accurate estimates of the actual RMS error than their kriging counterparts. Both generalized cross validation error (GCV) and estimated RMSE for PRS were reliable error estimates for different test problems, as they yielded small variation with the choice of DOE (low COV on individual problems), and the problem (low variation of the mean, and low COV for all problems). We noted that estimated RMSE typically underestimated the actual RMS errors in approximation, and GCV usually overestimated the errors.

GCV was better than the process variance estimate for kriging, as GCV yielded smaller variation with the choice of design of experiment (higher mean) and the nature of the problem (lower COV). GCV for kriging, usually, provided higher overestimate of actual error than GCV for PRS. Besides, the variation of GCV for kriging with different problems, and DOEs was higher compared to GCV for PRS (compare COV for all problems). Process variance for kriging performed the poorest among all global error measures. This error measure showed large variability with the choice of problem, and design of experiments (high COV for all problems). Usually, process variance overestimated the actual errors and the overestimate was much higher for Branin-Hoo function, and the cantilever beam design problem with two variables.

Overall, the conclusions are that the model independent global error measure (GCV, computationally expensive error measure) provides an equal or better estimate of actual root

mean square error than the model based error measures (computationally inexpensive error measure), namely prediction variance for PRS and process variance for kriging. The variability with the choice of DOE and the nature of the problem was also lower for GCV. Relatively, the performance of error estimation measures for PRS was better than their kriging counterparts. This is remarkable because the error in models considered here is primarily bias error rather than noise error, where kriging is supposed to perform better.

Pointwise Error Measures

Root mean square errors

We computed the ratio of predicted root mean square error (obtained *via* different pointwise error estimation measures) and actual RMS errors in the entire design space. The results for 1000 DOEs are summarized in Figure 5-4 with the help of boxplots, and corresponding mean and COV are given in Table 5-9. Although there was a single estimate of standard deviation of responses, we computed the ratio of predicted RMS and actual RMS error in PRS, KRG, and PWS, separately due to the differences in the actual errors. As before, the desired value of any ratio is one.

The standard error consistently underestimated the actual RMS error for all test problems. However, this error estimate yielded the least variation in the results with the choice of DOE (low COV for individual problems) and the nature of problem (lowest COV for all problems among all error measures). Thus, one can get a fairly good estimate of actual RMS error by inflating the root mean square of the estimated standard error by a factor of 1.5 (at least that is the case with the examples considered in this study). Root mean square of the RMS bias error overestimated the actual RMS error (except Hartman-6 problem) but provided a reasonable estimate of the actual error. The RMS bias error measure significantly overestimated actual errors for a few DOEs for Branin-Hoo function and radial turbine design problem, and

consequently the mean value was influenced (also large COV). The median values of the ratio for Branin-Hoo function and radial turbine design problem were 2.13 and 1.66, respectively. Overall, this error estimate usually had low variability with the choice of design of experiments and test problems. As expected, this error estimate performed the best when the assumed true function and the actual true function were very close (Camelback and Goldstein-Price examples). However, we note that RMS bias error measure involves two assumptions: (1) order of the assumed true polynomial, and (2) magnitude of parameter γ , that might influence its performance. The standard deviation of responses, in general, performed well in characterizing the actual RMS error and often, the predicted RMS error was close to the actual RMS error, except for the cantilever beam design with two variables example. The standard deviation of responses on an average (with respect to test problems) provided a reasonable estimate (median ratio is closed to one) of actual errors in PRS, but this estimate had significant variation in the results both with the choice of test problems (high COV, when all problems were simultaneously considered) and the choice of design of experiment.

An interesting observation from Table 5-7, Table 5-8, and Table 5-9 is that the variation in the actual RMS error in PRS approximation with respect to DOE is lower than the variation in its error estimates (estimated root mean square error, and standard error). This is because the fitting process is not influenced by the statistical assumptions while the error estimate is governed by assumptions that do not apply to the data.

The performance of the predicted root mean square error for kriging in characterizing the actual RMS errors, as denoted by the closeness of the ratio of predicted and actual RMS errors to one, was very good in the mean value of the ratio. However, the prediction capabilities of this error measure often depended on the choice of design of experiments (relatively large COV). For

Goldstein-Price problem, this error measure overestimated the actual RMS errors. The performance of standard deviation of responses in predicting kriging errors was similar to that observed for polynomial response surface approximation. For Branin-Hoo and cantilever beam design problem with two variables, the mean and median values of the ratios were significantly different (median values were 1.97 and 80.10, respectively). We had large variability with the choice of design of experiment, and the nature of problem. An interesting observation was that often when mean square error overpredicted the actual error, the standard deviation of responses underpredicted and *vice-versa*.

The root mean square of the standard deviation of responses usually overpredicted the actual RMS error incurred by PRESS-based weighted average surrogate model except for six-dimensional problems. The variation in results with DOE (COV) was high but comparable to other error estimation measures. However, we noted significant variation in the results with the choice of test problem (highest COV, when all problems data was considered simultaneously).

The performance of standard deviation of responses for cantilever beam design problem with two variables was particularly poor because one surrogate model, radial basis neural network, performed very poorly. Since we have information from four surrogate models, we can make a judicious decision to ignore predictions of radial basis neural network model. We computed the standard deviation of responses using remaining three surrogate models. The mean (number given in parenthesis is COV) of the ratio of root mean square of the re-calculated standard deviation of responses and actual RMS error for PRS, KRG, and PWS based on 1000 DOEs is 0.65 (0.31), 3.22 (0.85), and 0.85 (0.41), respectively. Compared to the performance of standard deviation of responses in Table 5-9, we observed significant improvements in the predictions. The improvement in performance explained the anomaly observed previously (very

high s_{resp}), and cautioned us against the possibility of getting poor estimate of s_{resp} , when predictions of one surrogate model were aberrant.

Correlations between actual and predicted errors

Next, we studied the localized performance of pointwise error estimates by computing correlations between pointwise predicted and actual errors. The mean and COV of correlations based on 1000 DOE experiments are given in Table 5-10. The data is summarized in the boxplots shown in Figure 5-5.

In general, all error estimators had only modest correlations with the actual error field in the entire design space, and there was significant variation with the choice of DOE for most test problems. The standard error (ESE) was the worst error estimator for all test problems, except radial turbine design problem. ESE had large variation (COV) with design of experiments and with the choice of the problem (highest COV when all problems were considered). RMS bias error estimates were good in characterizing the actual error field for Camelback, Goldstein-Price, where the assumed true models were close to the true functions, and cantilever beam design problem with two variables. We noted significant variation in the correlations with design of experiment, and the choice of test problems. For all problems except radial turbine design, RMSBE performed better than ESE.

The kriging mean squared predicted error also showed significant variations with design of experiment, and the nature of the test problem. But the correlations were good for the Branin-Hoo function, and the cantilever beam with two variables. The standard deviation of responses performed at par with other error estimation measures for PRS, KRG, and PWS models and, showed relatively smaller variations (COV) when all problems data was considered simultaneously.

The poor characterization of the entire error field was not unexpected because often the assumptions in the surrogate models were not satisfied. Besides, most error measures characterize the average error considering numerous functions, and comparison with a single function was bound to yield errors. The most interesting result was that overall the model independent error measures performed at par with the model based error estimations measures.

Maximum absolute errors

Next, we show the boxplots of ratio of predicted maximum error to the actual absolute maximum error at test points for 1000 DOEs in Figure 5-6. The mean and coefficient of variation (based on 1000 DOEs) of the ratio of predicted maximum and actual absolute maximum error are summarized in Table 5-11. As was observed for RMS errors, the standard error underestimated the maximum absolute actual errors in polynomial response surface approximation. The magnitude of underestimate varied with the choice of test problem, though variation due to DOE was moderate. On the other hand, RMS bias error, typically, overestimated the maximum actual error for all but six-dimensional test problems (Hartman-6 and radial turbine). This error measure had significant variability with the choice of DOE but results were more consistent with the nature of the problem. The standard deviation of responses yielded mixed behavior in characterizing the maximum actual absolute error. For some surrogate models, it underestimated the actual errors and for others it overestimated, though this result is not unexpected since this error measure is dependent on the surrogate predictions. Besides, there was significant variability with DOE and test problem. Mean square error for kriging underestimated the maximum actual absolute error. The underestimate was significantly high for six dimensional problems. For some problems, the variation in the results with DOE was significantly high.

We note that model based global error estimation measures are computationally inexpensive but they do not characterize the actual root mean square error very well. On the

other hand, local error estimation measures give a more reasonable estimate of actual RMS error in the entire design space. This is particularly true for kriging error estimation measures. In addition, we can use local error estimate to identify regions of high errors. We conclude that despite relatively higher computational cost compared to model based global error measures, it is useful to estimate local errors, especially when the cost of error prediction is significantly lower than the cost of acquiring the data.

Ensemble of Multiple Error Estimators

We observed that no single error estimation measure correctly characterized the entire error field for all problems. We proposed using multiple error measures to improve the error prediction capabilities. First, based on our observations in Table 5-8 and Table 5-9, we used an average of different RMS error estimates. Second, for kriging (though the approach can be applied to other surrogates as well), we used GCV-based measure to identify the best error measure for a given problem. Finally, we simultaneously used different error prediction measures to identify regions of high errors. The detailed results from these methods of ensemble of error estimation measures are as follows.

Averaging of Errors

We observed in previous section (Table 5-8 and Table 5-9), that for PRS approximations some error measures overestimated and others underestimated actual errors. This indicated that we can better estimate the magnitude of actual errors by averaging different error estimates. To verify this observation, we compared the actual RMS errors with the geometric mean of GCV and estimated root mean square error (global error measures), as well as with the geometric mean of standard error and RMS bias error (local error measures). The results of ratio of root mean square of averaged error measures and actual RMS error for different problems are summarized in Table 5-12.

A comparison of individual global error measures (Table 5-8) and averaged global error measure (Table 5-12, column 1) for PRS showed that average of errors was relatively more robust with the choice of problem (ratio is closer to one), and DOE (slightly smaller COV). Similarly, as is exemplified by the ratio of predicted and actual root mean square errors (Table 5-12, column 2), and correlation between actual and predicted error (Table 5-12, column 3), pointwise geometric averaging of estimated standard error and root mean square bias error for PRS was a better choice in characterizing the actual errors than the individual error estimates (Table 5-9). The geometric averaged error performed significantly better than the worst of the individual error estimates, and was at par with the better of the two. Besides, the performance was also more consistent for different problems (ratio closer to one and smaller COV when all problems were considered). Similar results were obtained for pointwise geometric averaging of mean square error and standard deviation of responses for kriging (Table 5-12, columns 4 and 5). In fact, the correlations between predicted and actual errors using the geometric average error were much better than individual error measures for kriging. Thus, we can say that averaging of the error estimates induces robustness in predictions.

At the same time, we reiterate that the choice of true polynomial and the distribution of polynomial coefficients were arbitrary for RMS bias error estimate. Further research is required to explore the influence of these parameters on the averaging of the errors. Besides, in future one may also consider different ways of combining the pointwise error estimates e.g., weighted averaging of errors, etc.

Identification of Suitable Error Estimator for Kriging

We observed that sometimes the mean square error characterized the actual error field better than the standard deviation of responses and *vice-versa*, but we cannot pre-determine the suitability of either error measure. In this section, we investigated the applicability of generalized

cross-validation approach to identify the better error measure. That is, whether the mean square error or standard deviation of responses was more appropriate error measure to characterize the actual error field. We fitted surrogate models to $N_s - 1$ data points, where N_s is the total number of samples. We estimated response, actual and predicted errors, and standard deviation of responses at the design point not used in constructing the surrogate model. Since we knew the actual response at the left out point, we could compute actual absolute error in prediction. Now, we characterized the performance of any error measure using the data at N_s points. We compared different error estimation measures using a suitable criterion, and identified the better of the two error measures to characterize the actual error field. In this study, we used the correlations between predicted and actual error, and the ratio of the predicted and actual root mean square error, as two criteria to identify the better error measure. Note that, the choice of error measure may vary with the nature of the problem as well as with design of experiment.

The mean performance of individual and chosen error measures based on 100 DOEs for different test problems is summarized in Table 5-13. We also tabulate the number of times the predicted best error measure failed to identify the correct better error estimation measure. Although the results for individual error estimation measure were obtained with only 100 DOEs to reduce high computational cost, the mean correlations and RMS errors were comparable to the data obtained with 1000 DOEs (Table 5-9 and Table 5-10).

The results clearly indicated that the error measure chosen using the GCV based information was often the better measure for that problem and DOE; and as expected, the mean of the chosen error measure was close to the better error measure. For the cantilever beam design problem with two variables, the standard deviation of responses performed very poorly (due to poor approximation of one surrogate model, RBNN) in characterizing the root mean square

errors and hence, almost always correctly discarded. So the performance of the predicted best error measure was comparable to mean square error measure. We note that the choice of error measure according to GCV-based criterion might be wrong up to 42% (average 25%) of the time, when the selection criterion was the correlation between the actual and predicted error; and up to 50% (average 30%) times, when the criterion was the ratio of predicted and actual root mean square error. Nevertheless, the accuracy in identifying the correct error estimation measure for any problem and DOE was encouraging.

Detection of High Error Regions using Multiple Errors Estimators

As discussed earlier, error measures are often used to identify the locations where predictions may have high errors/uncertainty. Hence, the failure of error estimation measures to detect regions of high error is undesirable. We can reduce the risk of missing regions of high errors by simultaneously using multiple error measures. We compared the performance of different individual error estimators, and combination of error estimators against the actual absolute errors using the suite of analytical and cantilever beam design problems. Different steps in the test procedure are given as follows.

- We scale the actual and predicted errors by the corresponding maximum errors in the entire design space to avoid disparity in the magnitude of different error estimates.
- We subdivide the entire design space in 2^{N_v} orthants.
- We define an orthant as high error orthant if the maximum scaled actual error in that orthant is equal to or greater than 0.7.
- A high error orthant is considered detected by an error measure, if the maximum scaled predicted error in the orthant is greater than 50% of the maximum scaled actual error in this orthant (0.7+).
- If we fail to detect at least one high error orthant, we consider that case as failure to detect high error region.
- If we fail to detect the orthant with the maximum actual error, the case is denoted as failure to detect maximum actual error.

The test set-up for error estimation is the same as described earlier. The number of cases out of 1000 that failed to detect high error regions using different error estimators is given in Table 5-14 for different test problems. We showed the results of individual and combination of multiple error estimators. For all problems, except Goldstein-Price and Hartman-3, the standard error yielded the least number of failures for polynomial response surface approximation. The mean square error for kriging worked well in identifying regions of high error. The standard deviation of responses performed reasonably in characterizing the actual errors in PRESS-based weighted average surrogate, considering the fact that the present formulation is only a naïve attempt to address a large class of problem of combination of multiple surrogates.

Combination of different error estimators was effective in reducing the number of failures. The combination of standard deviation of responses with standard error for polynomial response surface approximation, and with mean square error for kriging, was most effective in predicting high error regions. By combining more error estimation measures, we reduced the chance of missing high error regions. The results for the detection of maximum error using different error estimators, given in Table 5-15, also follow the same conclusions. The standard error for PRS and mean square error for KRG were the best individual error measures in detecting the maximum error regions and combination of error measures were effective in reducing the risk of missing the maximum error regions.

While the combination of multiple error estimators was useful to identify high error regions, they increased the chance of wrongly identifying low error regions as high error. The steps in the detection of false representation of high error regions using different error estimates are as follows.

- We scaled actual and predicted errors by corresponding maximum errors in the entire design space.

- We considered an orthant to be falsely marked as high error region by an error estimator, if the maximum scaled predicted error in the orthant was greater than or equal to 0.9, and the maximum scaled actual error in this orthant was less than 60% of the maximum scaled predicted error (0.9+).
- As before, we marked a case failed, if it wrongly identified at least one low error region as high error.
- We considered a case to falsely represent the maximum high error region, if the orthant wrongly marked as high error region had the maximum scaled predicted error in the entire design space.

The number of cases with wrong representation of high errors for different test problems while using different error estimators is given in Table 5-16. Different error estimators performed better for different test problems, however most often, the standard deviation of responses created the least number of false high error regions for any surrogate approximation. The root mean square bias error for PRS also performed quite well. As expected, combination of multiple error estimators increased the number of cases, for which false high error regions were created. We cannot conclude if one combination was better than others for all problems. The results for wrong representation of maximum errors, shown in Table 5-17, follow the same conclusions.

A high level summary of the performance of all pointwise error estimation measures is given in Table 5-18.

Conclusions

We compared the accuracy of different error estimators using a suite of analytical and industrial test problems. The main findings can be summarized as follows.

Global Error Estimators

- Global error estimators yielded a reasonably accurate estimate of the actual root mean square error for different test problems and DOEs.

- Estimated root mean square error in polynomial response surface approximation predicted the actual root mean square error in design space consistently; however, this measure underpredicted the actual errors.
- Process variance in kriging significantly overpredicted the actual errors.
- Model independent generalized cross-validation error was a good measure of assessing the actual root mean square error in design space for both polynomial response surface and kriging approximations, though GCV usually overpredicted the actual root mean square error.

Pointwise Error Estimation Models

- Pointwise error estimators performed slightly worse compared to the global measures but this was expected due to the constraints on the assumptions.
- Estimated standard error underestimated the actual errors by approximately 50%. Though this error estimator performed the worst in characterizing the entire error field, it had the best capability to detect the regions of high errors among all the error estimators for polynomial response surface approximation. This error estimator was least influenced by the choice of design of experiment, and the nature of the problem.
- RMS bias error resulted in a good approximation of the root mean square of the actual absolute error for all test problems, though it overpredicted the actual errors. When the assumed true model was close to the actual function, the RMS bias error characterized the actual error field quite well. This error estimate was mostly better than the standard error in characterizing the entire error field.
- Mean square error estimator typically underpredicted the actual root mean square errors, and showed low variation with the nature of the problem; but the prediction of entire error field was highly dependent on the function. However, this error estimate performed very accurately in predicting the high error regions.
- Model based local error estimation measures provided a better estimate of actual root mean square errors compared to the model based global error estimation measures.
- The standard deviation of responses, which is a model independent local error measure, usually gave a reasonable estimate of actual RMS errors for different problems, DOEs, and surrogates. However, the performance of this error estimator in characterizing the entire error field was highly problem dependent. Also this error estimate performed poorly, when one constituent surrogate model predictions were very poor. It should be noted that while model based error measures have been developed over many years, there is a need to develop model independent local error measures.

Simultaneous Application of Multiple Error Measures

- We showed that a geometric averaging of different error measures for PRS (standard error and root mean square bias error) provided a robust estimate (with respect to problems, and choice of DOE) of magnitude of actual RMS errors. The same concept worked well for global measures (GCV and estimated RMS error) as well. We obtained encouraging results for kriging (mean squared error and standard deviation of responses) too.
- We can use the errors and predictions at the design points while constructing the surrogate model using generalized cross validation procedure to identify the suitable error measure for any problem and design of experiment. The predicted best error measure based on GCV was correct about 75% of the simulations, and the predictions of the selected error estimation measure were at par with the best individual error measure for any test problem.
- We further showed that the combination of multiple error estimators improved the ability to detect high error regions for all problems. The combination of standard deviation of responses with the standard error for polynomial response surface approximation, and with mean square error for kriging was the most effective combination.
- More error estimators increased the probability of detecting the high error region. However, we noted that the combination of multiple error estimators increased the chances of wrongly representing a low error region as high error region. Nevertheless, we still benefit against the risk of failing to identify high error regions.

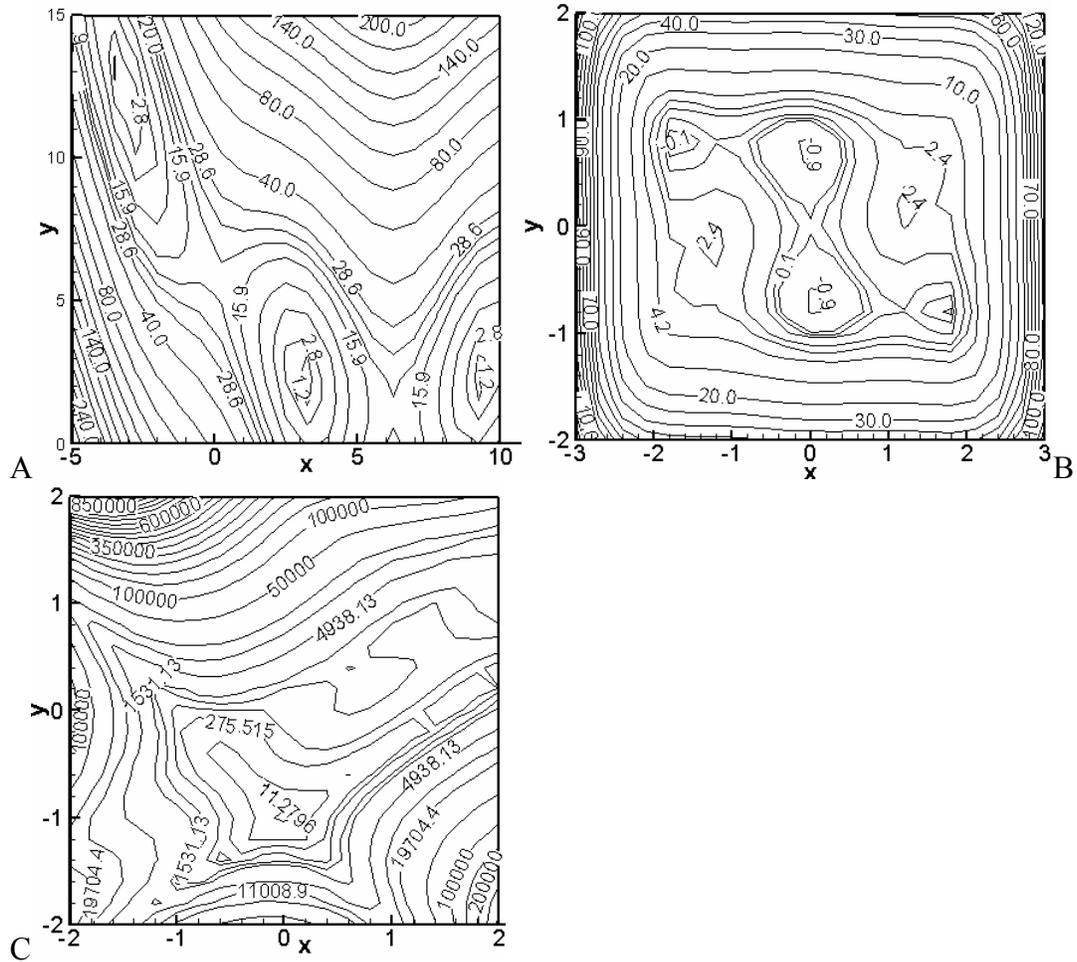


Figure 5-1. Contour plots of two variable analytical functions. A) Branin-Hoo. B) Camelback. C) Goldstein-Price.

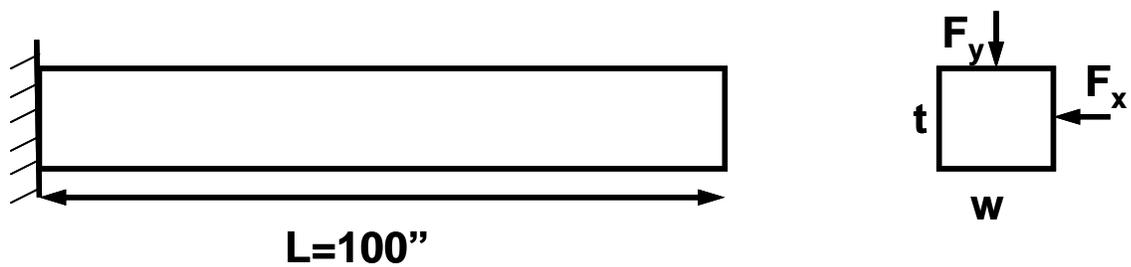
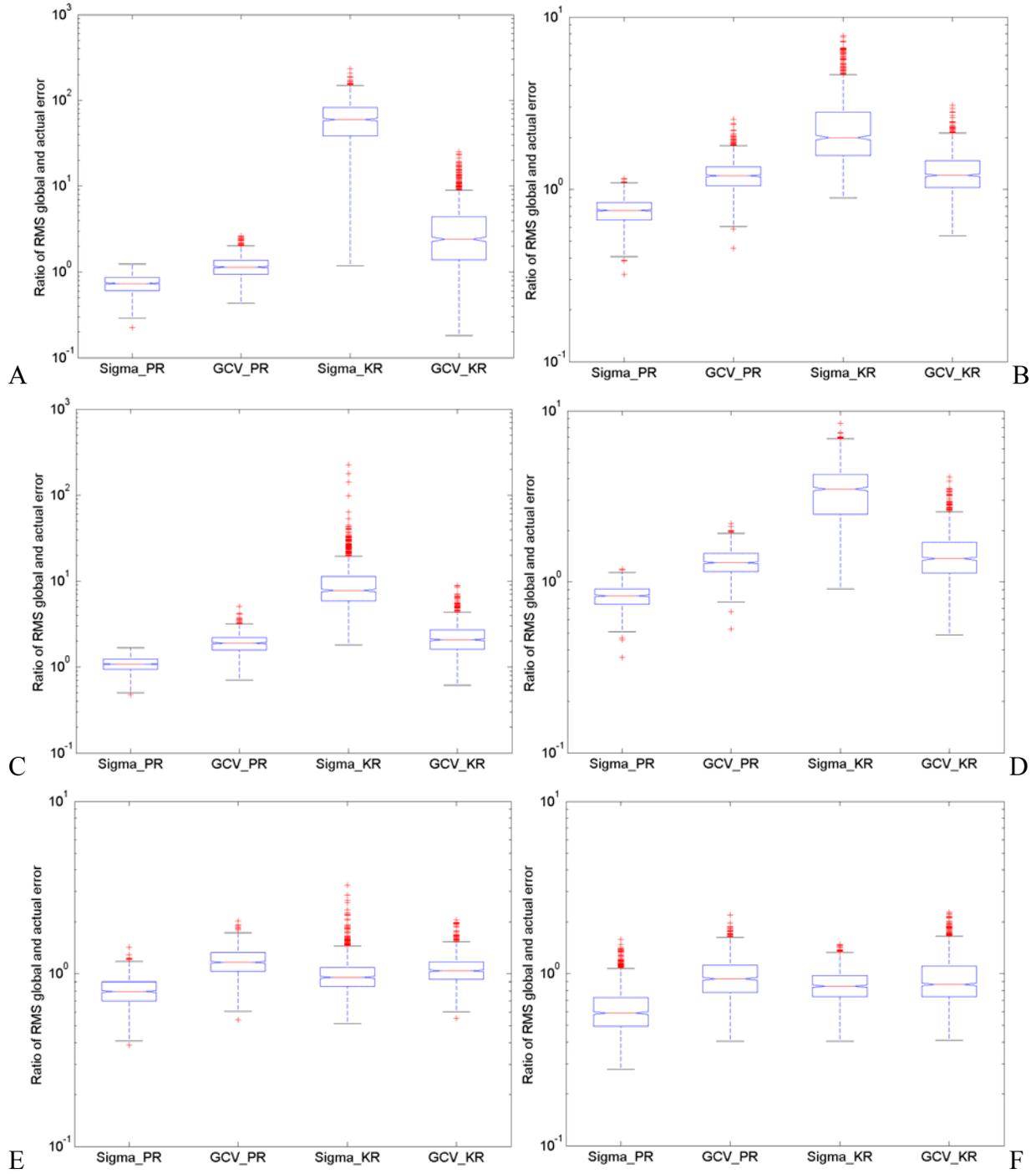


Figure 5-2. Cantilever beam subjected to horizontal and vertical random loads.



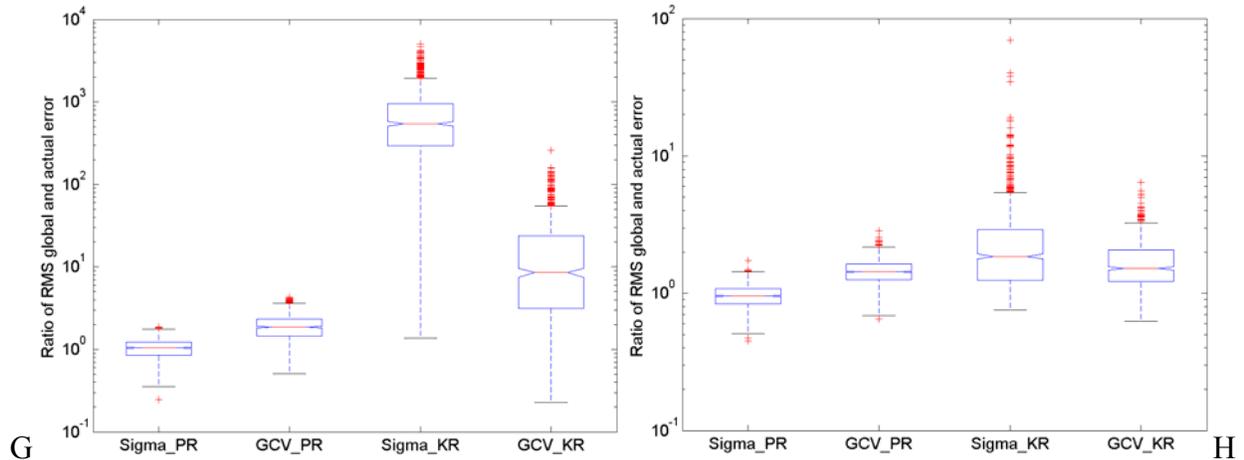
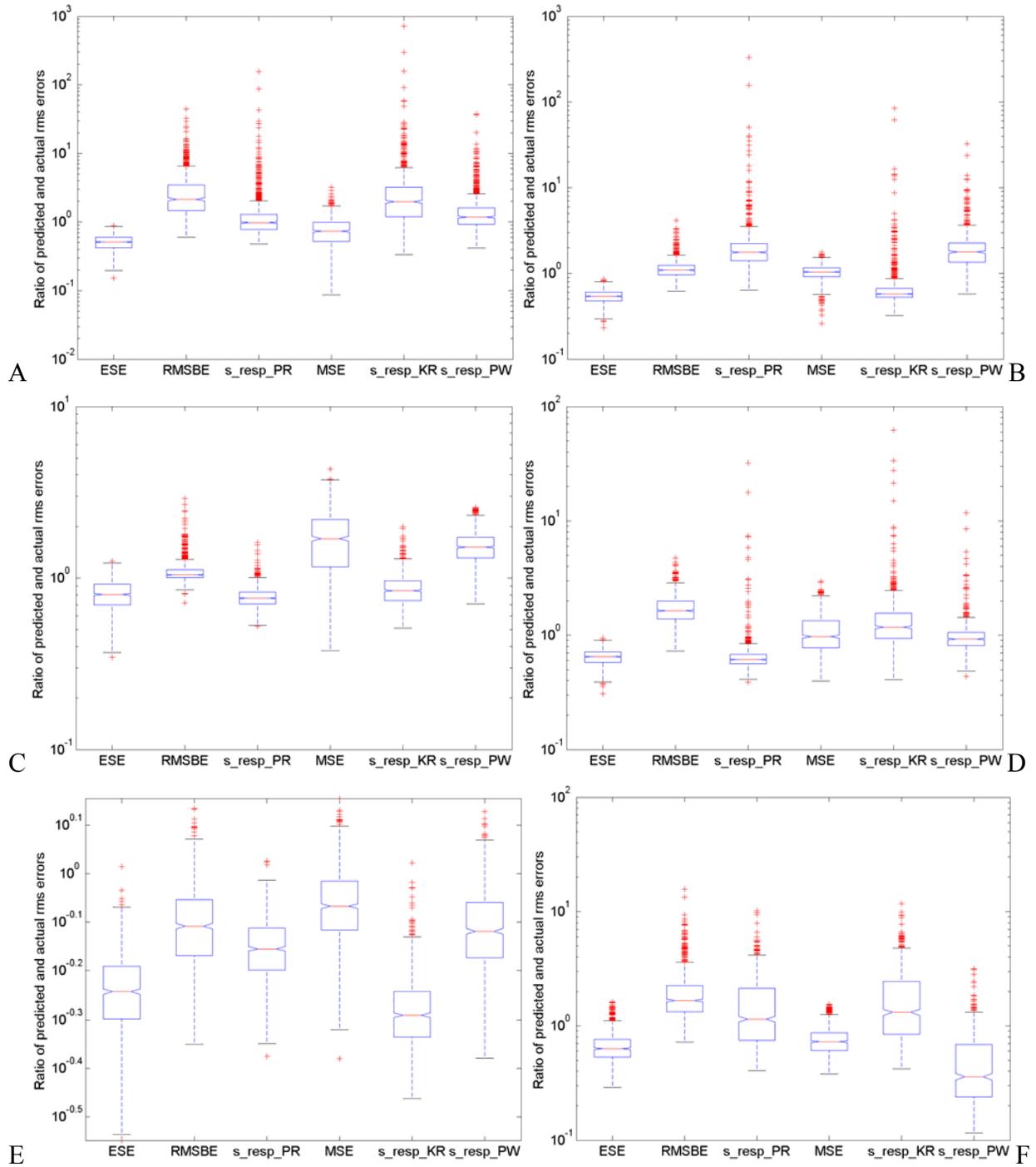


Figure 5-3. Ratio of global error measures and relevant actual RMS error (logscale). GCV_PR: ratio of square root of generalized cross-validation error and actual RMS error for PRS, GCV_KR: ratio of square root of generalized cross-validation error and actual RMS error for KRG, Sigma_PR: ratio of square root of estimated root mean square error and actual RMS error for PRS, Sigma_KR: ratio of square root of process variance and actual RMS error for kriging. A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem. G) Cantilever-2. H) Cantilever-5.



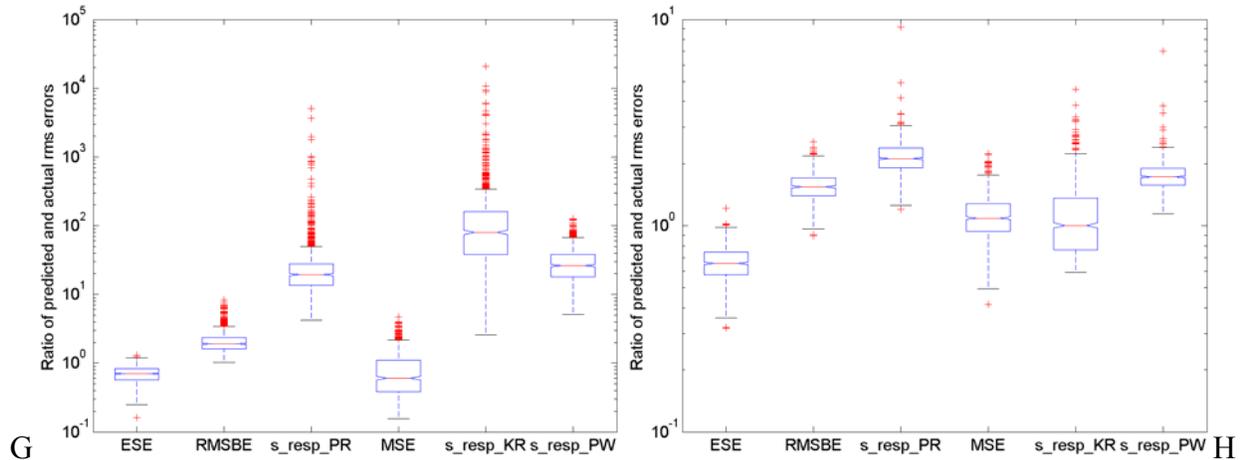
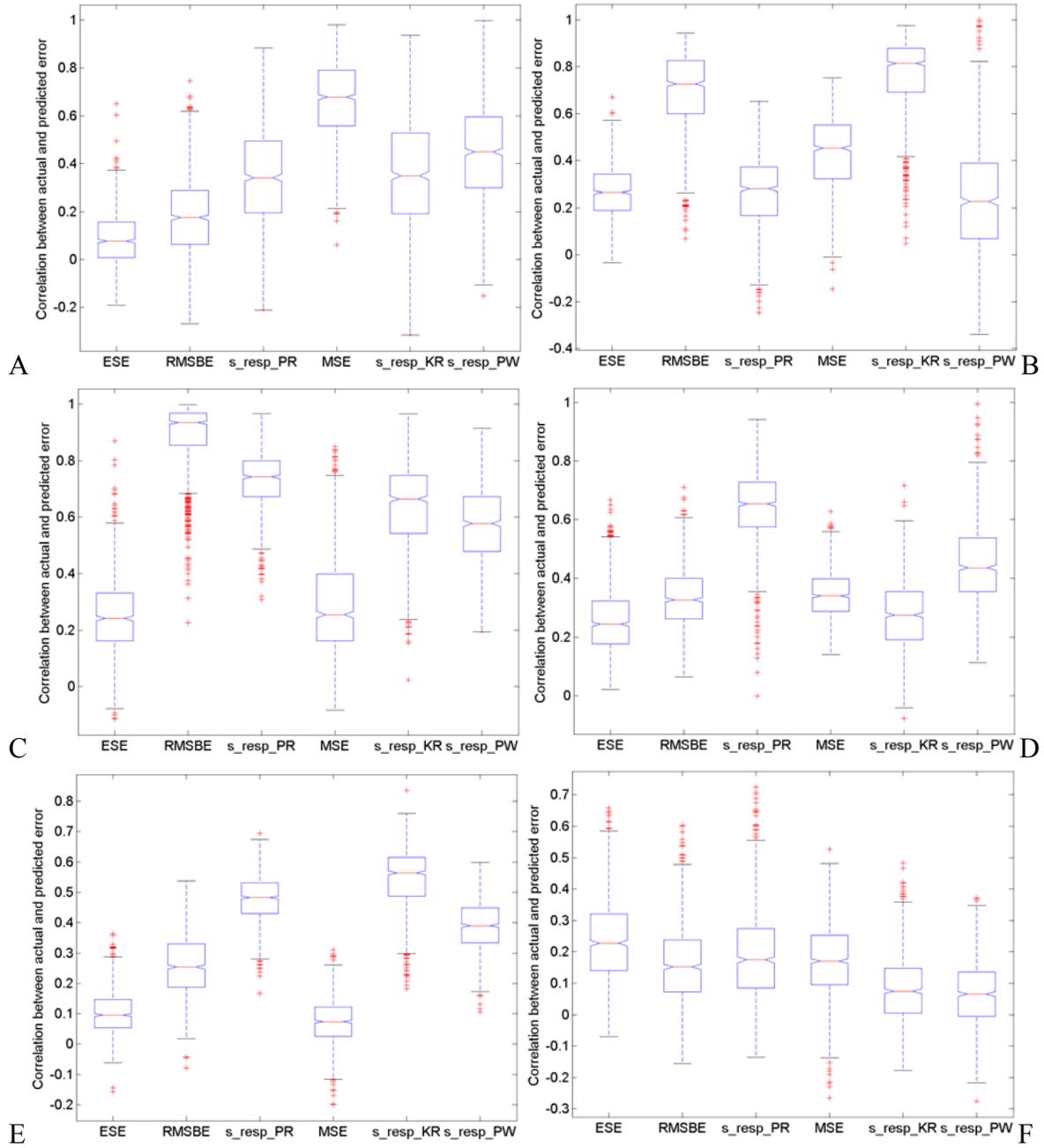


Figure 5-4. Ratio of root mean square values of pointwise predicted and actual errors for different problems, as denoted by predicted error measure. ESE: Estimated standard error, RMSBE: RMS bias error, s_resp_PR: ratio of root mean square of standard deviation of responses and actual RMS error in PRS, MSE: square root of mean square error, s_resp_KR: ratio of root mean square of standard deviation of responses and actual RMS error in kriging and. A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem. G) Cantilever-2. H) Cantilever-6.



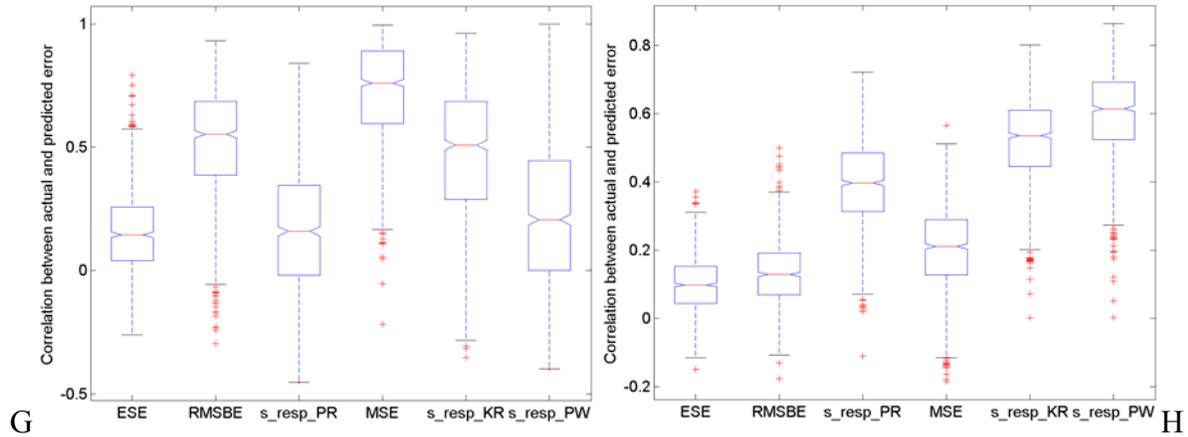
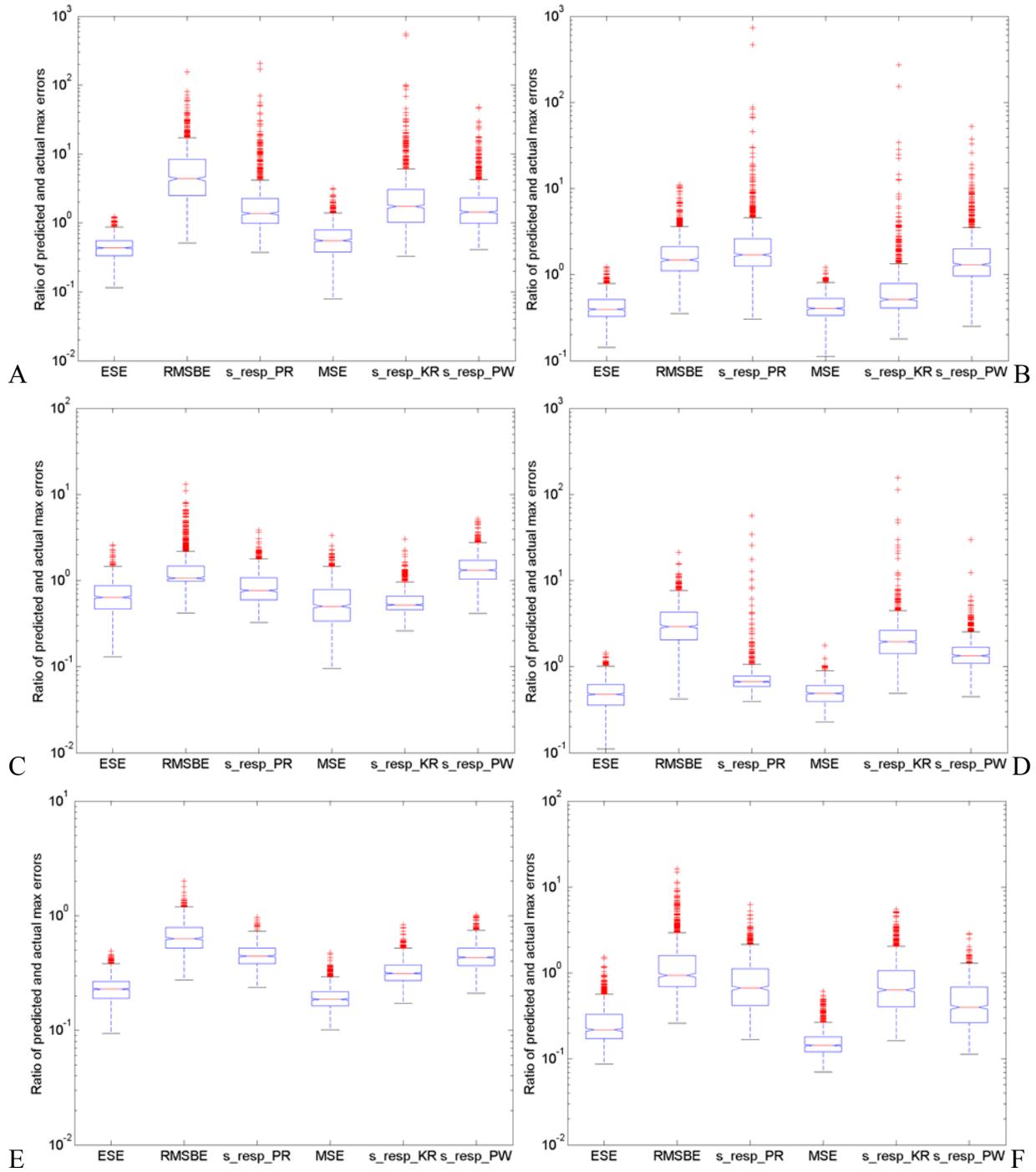


Figure 5-5. Correlation between actual and predicted error measures for different problems, denoted by predicted error measure. ESE: Estimated standard error, RMSBE: RMS bias error, s_resp_PR: correlation between actual error in PRS and standard deviation of responses, MSE: square root of mean square error, s_resp_KR: correlation between actual error in kriging and standard deviation of responses, s_resp_PW: correlation between actual error in PWS and standard deviation of responses. A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem. G) Cantilever-2. H) Camelback-6.



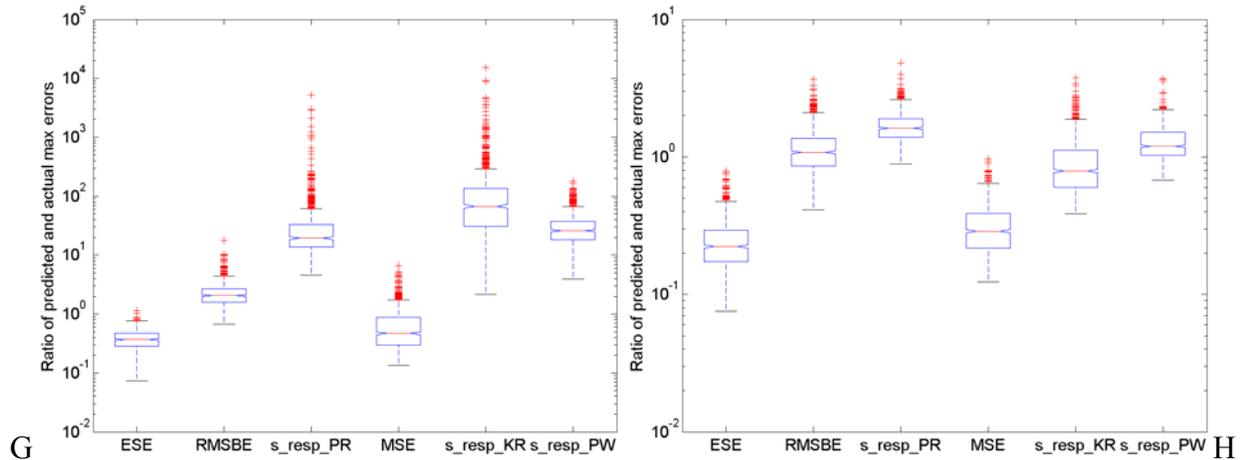


Figure 5-6. Ratio of maximum predicted and actual absolute errors in design space for different problems, denoted by predicted error measure. ESE: Estimated standard error, RMSBE: RMS bias error, s_resp_PR: correlation between actual error in PRS and standard deviation of responses, MSE: square root of mean square error, s_resp_KR: correlation between actual error in kriging and standard deviation of responses. A) Branin-Hoo. B) Camelback. C) Goldstein-Price. D) Hartman-3. E) Hartman-6. F) Radial turbine design problem. G) Cantilever-2. H) Cantilever-6.

Table 5-1. Summary of different error measures used in this study.

| Error measure | Description | Type | Nature | For surrogate |
|------------------|--|-------------------|--------|-----------------------------|
| σ_a (PRS) | Root mean square error, Estimate of variance of noise | Model based | Global | Polynomial response surface |
| σ (KRG) | (square root of) Process variance | Model based | Global | Kriging |
| GCV | Leave-one out error measure | Model independent | Global | All surrogates |
| e_{es} | Standard error characterizes to noise | Model based | Local | Polynomial response surface |
| e_b^{rms} | RMS bias error characterizes modeling error | Model based | Local | Polynomial response surface |
| e_{mse} | Mean squared error characterizes approximation error | Model based | Local | Kriging |
| s_{resp} | Standard deviation of responses characterize approximation uncertainty | Model independent | Local | All surrogates |

Table 5-2. Parameters used in Hartman function with three variables.

| I | a_{ij} | | c_i | p_{ij} | | | |
|-----|----------|------|-------|----------|---------|--------|--------|
| 1 | 3.0 | 10.0 | 30.0 | 1.0 | 0.3689 | 0.1170 | 0.2673 |
| 2 | 0.1 | 10.0 | 35.0 | 1.2 | 0.4699 | 0.4387 | 0.7470 |
| 3 | 3.0 | 10.0 | 30.0 | 3.0 | 0.1091 | 0.8732 | 0.5547 |
| 4 | 0.1 | 10.0 | 35.0 | 3.2 | 0.03815 | 0.5743 | 0.8828 |

Table 5-3. Parameters used in Hartman function with six variables.

| i | a_{ij} | | | c_i | | | |
|-----|----------|------|------|-------|------|------|-----|
| 1 | 10.0 | 3.0 | 17.0 | 3.5 | 1.7 | 8.0 | 1.0 |
| 2 | 0.05 | 10.0 | 17.0 | 0.1 | 8.0 | 14.0 | 1.2 |
| 3 | 3.0 | 3.5 | 1.7 | 10.0 | 17.0 | 8.0 | 3.0 |
| 4 | 17.0 | 8.0 | 0.05 | 10.0 | 0.1 | 14.0 | 3.2 |

| i | p_{ij} | | | | | |
|-----|----------|--------|--------|--------|--------|--------|
| 1 | 0.1312 | 0.1696 | 0.5569 | 0.0124 | 0.8283 | 0.5886 |
| 2 | 0.2329 | 0.4135 | 0.8307 | 0.3736 | 0.1004 | 0.9991 |
| 3 | 0.2348 | 0.1451 | 0.3522 | 0.2883 | 0.3047 | 0.6650 |
| 4 | 0.4047 | 0.8828 | 0.8732 | 0.5743 | 0.1091 | 0.0381 |

Table 5-4. Range of variables for radial turbine design problem.

| Variable | Description | Minimum | Maximum |
|----------|--|---------|---------|
| RPM | Rotational speed | 80000 | 150000 |
| Reaction | Percentage of stage pressure drop across rotor | 0.45 | 0.68 |
| U/Cisen | Isentropic velocity ratio | 0.50 | 0.63 |
| Tip Flow | Ratio of flow parameter to a choked flow parameter | 0.30 | 0.65 |
| Dhex% | Exit hub diameter as a % of inlet diameter | 0.1 | 0.4 |
| AN2Frac | Used to calculate annulus area (stress indicator) | 0.50 | 0.85 |

Table 5-5. Ranges of variables for cantilever beam design problem (five variables case).

| Variable | Minimum | Maximum | Units |
|----------------|---------|---------|-------|
| F _x | 700 | 1300 | lbs |
| F _y | 900 | 1300 | lbs |
| E | 20e6 | 35e6 | psi |
| W | 2.0 | 3.0 | inch |
| T | 3.0 | 5.0 | inch |

Table 5-6. Numerical setup for different test problems. N_v : number of variables, N_s : number of training points, N_{test} is number of test points, for analytical and Cantilever-2 problems $N_{test} = p^{N_v}$, where p is the number of points along each direction. For Hartman-6, Cantilever-5, and radial turbine problem, we specify the number of test points N_{test} in table. N_{lhs} : the number of points in the large LHS sample.

| Problem | N_v | N_s | p or N_{test} | Order of polynomial | N_{lhs} |
|-----------------|-------|-------|-------------------|---------------------|-----------|
| Branin-Hoo | 2 | 20 | 16 | 3 | 100 |
| Camelback | 2 | 30 | 16 | 4 | 150 |
| Goldstein-Price | 2 | 42 | 16 | 5 | 200 |
| Hartman-3 | 3 | 70 | 11 | 4 | 250 |
| Hartman-6 | 6 | 56 | 2000 | 2 | 200 |
| Radial Turbine | 6 | 56 | 168 | 2 | NA |
| Cantilever-2 | 2 | 12 | 16 | 2 | 60 |
| Cantilever-5 | 5 | 42 | 2000 | 2 | 210 |

Table 5-7. Mean and coefficient of variation (COV) (based on 1000 design of experiments) of normalized actual RMS error in the entire design space. We use the range of different functions to normalize the respective actual RMS errors. PRS: polynomial response surface approximation; the value in parenthesis is the order of loss function used to estimate coefficients, KRG: kriging, PWS: PRESS-based weighted surrogate model (using PRS with $p=2$, and $p=6$ loss functions, kriging, and radial basis neural network).

| | | PRS ($p=2$) | KRG | RBNN | PRS ($p=6$) | PWS |
|-----------------|------|---------------|---------|---------|---------------|---------|
| Branin-Hoo | Mean | 0.028 | 0.019 | 0.098 | 0.029 | 0.025 |
| | COV | 0.075 | 0.744 | 3.43 | 0.088 | 0.643 |
| Camelback | Mean | 0.041 | 0.122 | 0.162 | 0.041 | 0.045 |
| | COV | 0.050 | 0.264 | 5.92 | 0.073 | 1.035 |
| Goldstein-Price | Mean | 0.024 | 0.022 | 0.0093 | 0.025 | 0.012 |
| | COV | 0.096 | 0.290 | 0.473 | 0.094 | 0.197 |
| Hartman-3 | Mean | 0.109 | 0.060 | 0.121 | 0.118 | 0.075 |
| | COV | 0.065 | 0.357 | 2.42 | 0.077 | 0.496 |
| Hartman-6 | Mean | 0.073 | 0.098 | 0.068 | 0.076 | 0.066 |
| | COV | 0.072 | 0.129 | 0.110 | 0.097 | 0.101 |
| Radial turbine | Mean | 0.083 | 0.072 | 0.349 | 0.267 | 0.257 |
| | COV | 0.159 | 0.114 | 0.463 | 0.070 | 0.055 |
| Cantilever-2 | Mean | 1.03e-3 | 4.26e-4 | 9.71e-2 | 1.06e-3 | 1.22e-3 |
| | COV | 0.117 | 1.17 | 5.186 | 0.112 | 3.251 |
| Cantilever-5 | Mean | 0.015 | 0.032 | 0.061 | 0.016 | 0.018 |
| | COV | 0.149 | 0.393 | 0.162 | 0.186 | 0.147 |

Table 5-8. Mean and coefficient of variation (COV) (based on 1000 design of experiments) of ratio of global error measures and corresponding actual RMS error in design space. σ_a^{PRS} : estimated RMS error for PRS, GCV_{PRS} : square root of generalized cross-validation error for polynomial response surface approximation (PRS), σ_{KRG} : square root of process variance for kriging, GCV_{KRG} : square root of generalized cross-validation error for kriging. (All problems: indicate the mean and COV for all problems. i.e., 8000 designs of experiments). *Median is 3.08.

| | | σ_a^{PRS} | GCV_{PRS} | σ_{KRG} | GCV_{KRG} |
|-----------------|------|------------------|-------------|----------------|-------------|
| Branin-Hoo | Mean | 0.74 | 1.17 | 62.09 | 3.63 |
| | COV | 0.24 | 0.29 | 0.58 | 0.96 |
| Camelback | Mean | 0.75 | 1.22 | 2.32 | 1.27 |
| | COV | 0.17 | 0.21 | 0.47 | 0.28 |
| Goldstein-Price | Mean | 1.09 | 1.93 | 10.74 | 2.30 |
| | COV | 0.20 | 0.26 | 1.14 | 0.46 |
| Hartman-3 | Mean | 0.82 | 1.30 | 3.46 | 1.48 |
| | COV | 0.15 | 0.18 | 0.35 | 0.36 |
| Hartman-6 | Mean | 0.80 | 1.18 | 1.00 | 1.07 |
| | COV | 0.18 | 0.19 | 0.27 | 0.21 |
| Radial turbine | Mean | 0.64 | 0.97 | 0.86 | 0.96 |
| | COV | 0.32 | 0.28 | 0.21 | 0.34 |
| Cantilever-2 | Mean | 1.04 | 1.94 | 744.2 | 18.32 |
| | COV | 0.26 | 0.34 | 1.01 | 1.38 |
| Cantilever-5 | Mean | 0.97 | 1.45 | 2.69 | 1.73 |
| | COV | 0.18 | 0.20 | 1.37 | 0.42 |
| All problems | Mean | 0.86 | 1.40 | 103.42* | 3.84 |
| | COV | 0.28 | 0.36 | 3.49 | 2.75 |

Table 5-9. Mean and COV (based on 1000 design of experiments) of ratio of root mean squared predicted and actual errors for different test problems, denoted by predicted error measure. e_{es} : ratio of root mean square of estimated standard error and actual RMS error in PRS approximation, e_b^{rms} : ratio of root mean square of RMS bias error and actual RMS error in PRS approximation, s_{resp}^{PRS} : ratio of root mean square of standard deviation of responses and actual RMS error in PRS approximation, e_{mse} : ratio of RMS of square root of mean square error and actual RMS error in kriging approximation, s_{resp}^{KRG} : ratio of root mean square of standard deviation of responses and actual RMS error in kriging, s_{resp}^{PWS} : ratio of root mean square of standard deviation of responses and actual RMS error obtained using PRESS-based weighted average surrogate (PWS). We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses. (All problems: indicate the mean and COV for all problems. i.e., 8000 designs of experiments). * Median is 1.03. ** Median is 0.98. #Median is 1.31.

| | | e_{es} | e_b^{rms} | s_{resp}^{PRS} | e_{mse} | s_{resp}^{KRG} | s_{resp}^{PWS} |
|-----------------|------|----------|-------------|------------------|-----------|------------------|------------------|
| Branin-Hoo | Mean | 0.51 | 3.04 | 1.75 | 0.79 | 4.21 | 1.68 |
| | COV | 0.24 | 1.04 | 3.45 | 0.48 | 6.00 | 1.31 |
| Camelback | Mean | 0.54 | 1.15 | 2.73 | 1.04 | 0.91 | 2.03 |
| | COV | 0.17 | 0.27 | 4.36 | 0.20 | 3.81 | 0.82 |
| Goldstein-Price | Mean | 0.81 | 1.10 | 0.78 | 1.71 | 0.87 | 1.53 |
| | COV | 0.20 | 0.17 | 0.15 | 0.42 | 0.21 | 0.21 |
| Hartman-3 | Mean | 0.64 | 1.74 | 0.73 | 1.10 | 1.50 | 1.00 |
| | COV | 0.15 | 0.29 | 1.67 | 0.38 | 1.71 | 0.54 |
| Hartman-6 | Mean | 0.58 | 0.78 | 0.70 | 0.87 | 0.52 | 0.78 |
| | COV | 0.18 | 0.20 | 0.14 | 0.17 | 0.17 | 0.20 |
| Radial turbine | Mean | 0.68 | 2.03 | 1.56 | 0.76 | 1.82 | 0.50 |
| | COV | 0.32 | 0.62 | 0.74 | 0.28 | 0.75 | 0.74 |
| Cantilever-2 | Mean | 0.71 | 2.14 | 47.29 | 0.83 | 230.7 | 30.57 |
| | COV | 0.26 | 0.40 | 4.77 | 0.77 | 4.09 | 0.58 |
| Cantilever-5 | Mean | 0.66 | 1.56 | 2.16 | 1.12 | 1.13 | 1.75 |
| | COV | 0.18 | 0.16 | 0.20 | 0.24 | 0.42 | 0.18 |
| All problems | Mean | 0.64 | 1.69 | 7.21* | 1.03 | 30.20** | 4.98# |
| | COV | 0.26 | 0.85 | 11.28 | 0.50 | 11.34 | 2.33 |

Table 5-10. Mean and COV (over 1000 design of experiments) of correlations between actual and predicted errors for different test problems. e_{act}^{PRS} : actual error in polynomial response surface (PRS) approximation, e_{es} : estimated standard error, e_b^{rms} : root mean square bias error, e_{act}^{KRG} : actual error in kriging (KRG), e_{mse} : square root of mean square error, s_{resp} : standard deviation of responses, e_{act}^{PWS} : actual error in PRESS-based weighted average surrogate (PWS) model. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses. (All problems: indicate the mean and COV for all problems. i.e., 8000 designs of experiments).

| Error measure → | Correlation between e_{act}^{PRS} and | | | Correlation between e_{act}^{KRG} and | | Correlation between e_{act}^{PWS} and | |
|---|---|-------------|------------|---|------------|---|-------|
| | e_{es} | e_b^{rms} | s_{resp} | e_{mse} | s_{resp} | s_{resp} | |
| Branin-Hoo | Mean | 0.083 | 0.18 | 0.35 | 0.66 | 0.36 | 0.45 |
| | COV | 1.32 | 0.90 | 0.61 | 0.25 | 0.66 | 0.46 |
| Camelback | Mean | 0.27 | 0.70 | 0.27 | 0.43 | 0.77 | 0.24 |
| | COV | 0.41 | 0.24 | 0.59 | 0.37 | 0.20 | 0.95 |
| Goldstein-Price | Mean | 0.25 | 0.89 | 0.73 | 0.29 | 0.64 | 0.57 |
| | COV | 0.54 | 0.13 | 0.14 | 0.60 | 0.24 | 0.23 |
| Hartman-3 | Mean | 0.26 | 0.33 | 0.64 | 0.34 | 0.28 | 0.45 |
| | COV | 0.42 | 0.31 | 0.19 | 0.24 | 0.44 | 0.31 |
| Hartman-6 | Mean | 0.10 | 0.26 | 0.48 | 0.073 | 0.54 | 0.39 |
| | COV | 0.71 | 0.39 | 0.16 | 1.06 | 0.19 | 0.21 |
| Radial turbine | Mean | 0.24 | 0.16 | 0.19 | 0.17 | 0.084 | 0.066 |
| | COV | 0.55 | 0.75 | 0.75 | 0.70 | 1.32 | 1.53 |
| Cantilever-2 | Mean | 0.16 | 0.52 | 0.17 | 0.72 | 0.47 | 0.24 |
| | COV | 1.05 | 0.44 | 1.45 | 0.29 | 0.57 | 1.27 |
| Cantilever-5 | Mean | 0.10 | 0.14 | 0.39 | 0.20 | 0.52 | 0.60 |
| | COV | 0.80 | 0.69 | 0.33 | 0.60 | 0.23 | 0.22 |
| # of problems with mean correlation > 0.5 | | 0 | 3 | 2 | 2 | 4 | 2 |
| All problems | Mean | 0.18 | 0.40 | 0.40 | 0.36 | 0.46 | 0.38 |
| | COV | 0.76 | 0.75 | 0.61 | 0.72 | 0.57 | 0.66 |

Table 5-11. Mean and COV (based on 1000 design of experiments) of ratio of maximum predicted and actual errors for different test problems, denoted as e_{es} : ratio of maximum estimated standard error and maximum actual error in PRS approximation, e_b^{rms} : ratio of maximum RMS bias error and maximum actual error in PRS approximation, s_{resp}^{PRS} : ratio of maximum standard deviation of responses and maximum actual error in PRS approximation, e_{mse} : ratio of maximum square root of mean square error and maximum actual error in kriging approximation, s_{resp}^{KRG} : Ratio of maximum standard deviation of responses and maximum actual error in kriging, s_{resp}^{PWS} : Ratio of maximum standard deviation of responses and maximum actual error obtained using PRESS-based weighted average surrogate (PWS). We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses. (All problems: indicate the mean and COV for all problems. i.e., 8000 designs of experiments).

| | | e_{es} | e_b^{rms} | s_{resp}^{PRS} | e_{mse} | s_{resp}^{KRG} | s_{resp}^{PWS} |
|-----------------|------|----------|-------------|------------------|-----------|------------------|------------------|
| Branin-Hoo | Mean | 0.46 | 7.35 | 3.02 | 0.65 | 4.53 | 2.42 |
| | COV | 0.37 | 1.32 | 3.22 | 0.61 | 5.56 | 1.49 |
| Camelback | Mean | 0.43 | 1.86 | 4.18 | 0.44 | 1.34 | 2.05 |
| | COV | 0.36 | 0.73 | 6.72 | 0.33 | 7.50 | 1.46 |
| Goldstein-Price | Mean | 0.70 | 1.49 | 0.89 | 0.61 | 0.60 | 1.46 |
| | COV | 0.47 | 0.75 | 0.47 | 0.63 | 0.40 | 0.44 |
| Hartman-3 | Mean | 0.51 | 3.45 | 0.93 | 0.52 | 2.69 | 1.51 |
| | COV | 0.40 | 0.60 | 2.56 | 0.31 | 2.49 | 0.77 |
| Hartman-6 | Mean | 0.24 | 0.67 | 0.46 | 0.20 | 0.33 | 0.46 |
| | COV | 0.27 | 0.32 | 0.24 | 0.25 | 0.26 | 0.28 |
| Radial Turbine | Mean | 0.29 | 1.45 | 0.89 | 0.17 | 0.87 | 0.51 |
| | COV | 0.63 | 1.02 | 0.78 | 0.46 | 0.85 | 0.67 |
| Cantilever-2 | Mean | 0.39 | 2.34 | 51.27 | 0.71 | 193.0 | 31.26 |
| | COV | 0.37 | 0.54 | 4.62 | 0.96 | 3.78 | 0.67 |
| Cantilever-5 | Mean | 0.25 | 1.16 | 1.69 | 0.32 | 0.93 | 1.30 |
| | COV | 0.42 | 0.38 | 0.25 | 0.41 | 0.50 | 0.30 |
| All problems | Mean | 0.41 | 2.47 | 7.92 | 0.45 | 25.54 | 5.12 |
| | COV | 0.58 | 1.67 | 10.86 | 0.84 | 10.41 | 2.44 |

Table 5-12. Mean and COV (based on 1000 design of experiments) of ratio of root mean square average error and actual RMS errors for different test problems. GCV: square root of generalized cross validation error, σ_a^{PRS} : square root of prediction variance, e_{es} : estimated standard error (PRS), e_b^{rms} : root mean square bias error (PRS), e_{mse} : square root of mean squared error (KRG), s_{resp} : standard deviation of responses. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to estimate standard deviation of responses. (All problems: indicate the mean and COV for all problems. i.e., 8000 designs of experiments). * Median is 0.90.

| | $\sqrt{GCV \times \sigma_a^{PRS}}$ | | $\sqrt{e_{es} \times e_b^{rms}}$ (RMSE) | | $\sqrt{e_{es} \times e_b^{rms}}$ Correlation | | $\sqrt{e_{mse} \times s_{resp}}$ (RMSE) | | $\sqrt{e_{mse} \times s_{resp}}$ Correlation | |
|-----------------|------------------------------------|------|--|------|---|------|--|------|---|------|
| | Mean | COV | Mean | COV | Mean | COV | Mean | COV | Mean | COV |
| Branin-Hoo | 0.93 | 0.25 | 1.16 | 0.43 | 0.19 | 0.80 | 1.47 | 1.09 | 0.59 | 0.29 |
| Camelback | 0.96 | 0.18 | 0.78 | 0.16 | 0.70 | 0.21 | 0.85 | 0.54 | 0.71 | 0.20 |
| Goldstein-Price | 1.45 | 0.21 | 0.94 | 0.14 | 0.83 | 0.13 | 1.19 | 0.25 | 0.62 | 0.18 |
| Hartman-3 | 1.04 | 0.16 | 1.05 | 0.18 | 0.35 | 0.28 | 1.22 | 0.48 | 0.38 | 0.29 |
| Hartman-6 | 0.97 | 0.18 | 0.67 | 0.18 | 0.24 | 0.37 | 0.67 | 0.16 | 0.51 | 0.17 |
| Radial turbine | 0.79 | 0.30 | 1.16 | 0.43 | 0.21 | 0.59 | 1.12 | 0.44 | 0.13 | 0.86 |
| Cantilever-2 | 1.41 | 0.28 | 1.20 | 0.24 | 0.52 | 0.39 | 10.53 | 1.22 | 0.63 | 0.35 |
| Cantilever-5 | 1.18 | 0.19 | 1.01 | 0.14 | 0.14 | 0.61 | 1.10 | 0.26 | 0.54 | 0.18 |
| All problems | 1.09 | 0.31 | 0.91 | 0.34 | 0.40 | 0.69 | 2.07* | 2.49 | 0.51 | 0.43 |

Table 5-13. Comparison of performance of individual error measures and GCV chosen error measure for kriging. We show mean of errors based on 100 DOEs. For any DOE, predicted best measure is the one that yields better performance on the selected criterion. e_{mse} : square root of mean squared error in kriging, s_{resp} : standard deviation of responses. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to estimate standard deviation of responses. We consider a case ‘failure’ if the GCV chosen error measure performs poorly in the entire design space compared to other error measure. * Median is 0.69.

| Criterion Error model | Correlations | | | | Root mean square error | | | |
|--------------------------|--------------|------------|----------------|--------------|------------------------|------------|----------------|--------------|
| | e_{mse} | s_{resp} | Predicted best | Times failed | e_{mse} | s_{resp} | Predicted best | Times failed |
| Branin-Hoo | 0.68 | 0.35 | 0.59 | 33 | 0.79 | 3.83 | 1.33 | 50 |
| Camelback | 0.42 | 0.77 | 0.65 | 35 | 1.03 | 0.80 | 0.99 | 21 |
| Goldstein-Price | 0.30 | 0.62 | 0.62 | 14 | 1.79 | 0.90 | 1.28 | 38 |
| Hartman-3 | 0.35 | 0.28 | 0.35 | 30 | 1.03 | 1.45 | 1.07 | 47 |
| Hartman-6 | 0.083 | 0.54 | 0.50 | 9 | 0.87 | 0.53 | 0.87 | 5 |
| Radial turbine | 0.16 | 0.10 | 0.15 | 42 | 0.78 | 2.03 | 0.94 | 37 |
| Cantilever-2 | 0.73 | 0.46 | 0.68 | 29 | 0.80 | 491.7 | 8.02* | 5 |
| Cantilever-5 | 0.21 | 0.51 | 0.50 | 7 | 1.19 | 1.16 | 1.21 | 40 |

Table 5-14. Number of cases out of 1000 for which error estimators failed to detect high error regions. e_{es} : estimated standard error, e_b^{rms} : root mean square bias error, s_{resp}^{PRS} : standard deviation of responses as an error measure for PRS, e_{mse} : mean square error (kriging), s_{resp}^{KRG} : standard deviation of responses as an error measure for kriging, s_{resp}^{PWS} : standard deviation of responses as an error measure for PWS, ‘+’ indicates that multiple error estimators are combined. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses.

| Error measure(s) | Branin-Hoo | Camel back | Goldstein -Price | Hartman -3 | Hartman -6 | Cantilever-2 | Cantilever-5 |
|---------------------------------------|------------|------------|------------------|------------|------------|--------------|--------------|
| e_{es} | 31 | 52 | 252 | 197 | 11 | 4 | 23 |
| e_b^{rms} | 641 | 410 | 128 | 561 | 681 | 245 | 508 |
| s_{resp}^{PRS} | 357 | 243 | 226 | 88 | 119 | 501 | 157 |
| $e_{es} + s_{resp}^{PRS}$ | 19 | 37 | 62 | 28 | 1 | 4 | 2 |
| $e_b^{rms} + s_{resp}^{PRS}$ | 222 | 110 | 46 | 57 | 54 | 170 | 59 |
| $e_{es} + e_b^{rms}$ | 22 | 32 | 61 | 146 | 6 | 4 | 14 |
| $e_{es} + e_b^{rms} + s_{resp}^{PRS}$ | 14 | 22 | 22 | 21 | 1 | 4 | 0 |
| e_{mse} | 99 | 2 | 15 | 31 | 0 | 87 | 11 |
| s_{resp}^{KRG} | 272 | 145 | 31 | 634 | 69 | 302 | 84 |
| $e_{mse} + s_{resp}^{KRG}$ | 30 | 1 | 3 | 13 | 0 | 49 | 1 |
| s_{resp}^{PWS} | 237 | 186 | 168 | 399 | 106 | 433 | 33 |

Table 5-15. Number of cases out of 1000 for which error estimators failed to detect maximum error regions. e_{es} : estimated standard error, e_b^{rms} : root mean square bias error, s_{resp}^{PRS} : standard deviation of responses as an error measure for PRS, e_{mse} : mean square error (kriging), s_{resp}^{KRG} : standard deviation of responses as an error measure for kriging, s_{resp}^{PWS} : standard deviation of responses as an error measure for PWS, ‘+’ indicates that multiple error estimators are combined. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses.

| Error measure(s) | Branin-Hoo | Camel back | Goldstein-Price | Hartman-3 | Hartman-6 | Cantilever-2 | Cantilever-5 |
|---------------------------------------|------------|------------|-----------------|-----------|-----------|--------------|--------------|
| e_{es} | 17 | 28 | 199 | 120 | 6 | 3 | 20 |
| e_b^{rms} | 400 | 212 | 72 | 379 | 346 | 130 | 436 |
| s_{resp}^{PRS} | 202 | 141 | 56 | 34 | 30 | 357 | 54 |
| $e_{es} + s_{resp}^{PRS}$ | 11 | 22 | 26 | 11 | 0 | 3 | 0 |
| $e_b^{rms} + s_{resp}^{PRS}$ | 120 | 53 | 9 | 19 | 11 | 95 | 19 |
| $e_{es} + e_b^{rms}$ | 12 | 14 | 44 | 80 | 2 | 3 | 13 |
| $e_{es} + e_b^{rms} + s_{resp}^{PRS}$ | 7 | 11 | 7 | 7 | 0 | 3 | 0 |
| e_{mse} | 66 | 1 | 13 | 19 | 0 | 63 | 6 |
| s_{resp}^{KRG} | 189 | 69 | 16 | 339 | 16 | 235 | 20 |
| $e_{mse} + s_{resp}^{KRG}$ | 20 | 1 | 2 | 8 | 0 | 40 | 0 |
| s_{resp}^{PWS} | 139 | 101 | 65 | 189 | 34 | 279 | 11 |

Table 5-16. Number of cases out of 1000 for which different error estimators wrongly marked low error regions as high error. e_{es} : estimated standard error, e_b^{rms} : root mean square bias error, s_{resp}^{PRS} : standard deviation of responses as an error measure for PRS, e_{mse} : mean square error (kriging), s_{resp}^{KRG} : standard deviation of responses as an error measure for kriging, s_{resp}^{PWS} : standard deviation of responses as an error measure for PWS, '+' indicates that multiple error estimators are combined. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses.

| Error measure(s) | Branin-Hoo | Camel back | Goldstein-Price | Hartman-3 | Hartman-6 | Cantilever-2 | Cantilever-5 |
|---------------------------------------|------------|------------|-----------------|-----------|-----------|--------------|--------------|
| e_{es} | 293 | 164 | 479 | 585 | 950 | 497 | 963 |
| e_b^{rms} | 270 | 172 | 175 | 545 | 844 | 238 | 871 |
| s_{resp}^{PRS} | 186 | 220 | 240 | 124 | 554 | 433 | 529 |
| $e_{es} + s_{resp}^{PRS}$ | 358 | 271 | 542 | 618 | 973 | 628 | 975 |
| $e_b^{rms} + s_{resp}^{PRS}$ | 345 | 315 | 336 | 577 | 922 | 494 | 924 |
| $e_{es} + e_b^{rms}$ | 386 | 271 | 517 | 729 | 988 | 550 | 983 |
| $e_{es} + e_b^{rms} + s_{resp}^{PRS}$ | 425 | 350 | 572 | 742 | 993 | 649 | 989 |
| e_{mse} | 253 | 579 | 797 | 739 | 999 | 262 | 975 |
| s_{resp}^{KRG} | 270 | 78 | 110 | 269 | 509 | 405 | 279 |
| $e_{mse} + s_{resp}^{KRG}$ | 418 | 596 | 812 | 772 | 1000 | 486 | 981 |
| s_{resp}^{PWS} | 128 | 189 | 171 | 167 | 643 | 244 | 319 |

Table 5-17. Number of cases out of 1000 for which different error estimators wrongly marked low error regions as the maximum error region. e_{es} : estimated standard error, e_b^{rms} : root mean square bias error, s_{resp}^{PRS} : standard deviation of responses as an error measure for PRS, e_{mse} : mean square error (kriging), s_{resp}^{KRG} : standard deviation of responses as an error measure for kriging, s_{resp}^{PWS} : standard deviation of responses as an error measure for PWS, '+' indicates that multiple error estimators are combined. We used four surrogates, PRS with $p=2$ and $p=6$ loss functions, KRG, and RBNN models, to construct PWS, and to estimate standard deviation of responses.

| Error measure(s) | Branin-Hoo | Camel back | Goldstein-Price | Hartman-3 | Hartman-6 | Cantilever-2 | Cantilever-5 |
|---------------------------------------|------------|------------|-----------------|-----------|-----------|--------------|--------------|
| e_{es} | 220 | 118 | 399 | 483 | 774 | 313 | 910 |
| e_b^{rms} | 247 | 146 | 147 | 487 | 728 | 198 | 798 |
| s_{resp}^{PRS} | 151 | 177 | 193 | 88 | 302 | 386 | 394 |
| $e_{es} + s_{resp}^{PRS}$ | 285 | 224 | 461 | 516 | 824 | 505 | 936 |
| $e_b^{rms} + s_{resp}^{PRS}$ | 311 | 264 | 281 | 519 | 790 | 443 | 860 |
| $e_{es} + e_b^{rms}$ | 335 | 214 | 439 | 646 | 928 | 386 | 956 |
| $e_{es} + e_b^{rms} + s_{resp}^{PRS}$ | 373 | 292 | 492 | 662 | 941 | 534 | 969 |
| e_{mse} | 195 | 314 | 573 | 474 | 845 | 207 | 862 |
| s_{resp}^{KRG} | 237 | 58 | 71 | 214 | 275 | 362 | 174 |
| $e_{mse} + s_{resp}^{KRG}$ | 359 | 342 | 599 | 545 | 868 | 424 | 884 |
| s_{resp}^{PWS} | 110 | 146 | 127 | 129 | 372 | 211 | 203 |

Table 5-18. High level summary of performance of different pointwise error estimators.

| Error measure | Relative to actual error | Characterization actual error field | Variation with DOE | Variation with problem | Detect high error region |
|------------------|--------------------------|-------------------------------------|--------------------|------------------------|--------------------------|
| e_{es} | Underestimate | Poor | Low | Low | Good |
| e_b^{rms} | Overestimate | Moderate | Low | Low | Poor |
| e_{mse} | Underestimate | Moderate | High | Moderate | Good |
| s_{resp}^{PRS} | Varies/ underestimate | Moderate | Moderate | High | Moderate |
| s_{resp}^{KRG} | Varies/ overestimate | Moderate | High | High | Moderate |
| s_{resp}^{PWS} | Varies/ underestimate | Poor | Moderate | High | Moderate |

CHAPTER 6 CRYOGENIC CAVITATION MODEL VALIDATION AND SENSITIVITY EVALUATION

Introduction

Code validation and verification is a complex and time-consuming, but essential exercise to ensure accuracy of the predictions of CFD codes (Roache et al., 1986; AIAA editorial board, 1994; ASME editorial board, 1994; AIAA guide, 1998; Roache, 1998; Oberkampf et al., 2004). For computational verification and validation exercises, multiple aspects need to be addressed. First of all, one needs to ensure that the numerical representation of the analytical model approaches the correct solution as the grids and time step sizes approach the limiting values; this is the so-called verification. Verification deals with programming errors, algorithmic insufficiencies, and inconsistencies. The second aspect is to investigate whether and how a particular physical model can reproduce, or at least, satisfactorily approximate the observed phenomena, and reproduce the experimental measurements. That is, one should examine the propriety of the mathematical models and assumptions; this is the so-called validation.

Code validation or model validation is further complicated when the mathematical model involves adjustable parameters because there is a danger of fitting the experimental errors rather than the physical reality. We demonstrate for cryogenic cavitation, how the tools of surrogate modeling and global sensitivity analysis, which are extensively used in design and optimization of computationally expensive problems (Li and Padula, 2005; Queipo et al., 2005), can help with model validation and calibration.

Cavitating Flows: Significance and Previous Computational Efforts

Cavitation is one of the foremost problems observed in the turbomachinery such as inducers, pumps, turbines, marine propellers, nozzles, hydrofoils, etc. due to wide ranging pressure variations in the flow. Cavitation occurs when the local pressure in the flow falls below

the vapor pressure, and consequently, the fluid undergoes a phase change (Batchelor, 1967; Brennen, 1994, 1995). Cavitation induces noise, mechanical vibrations, material erosion, and can severely impact the performance as well as the structural integrity of fluid machinery. The study of cavitating flows is complicated by simultaneous presence of turbulence, multiple timescales, large density variations or phase change, interfacial dynamics etc. Due to its practical importance and rich physics, cavitating flow is a topic of substantial interest and challenge to the computational community.

The study of cavitating flows in cryogenic environment has practical importance for space applications because cryogenics often serve as fuels for space launch vehicles (NASA online facts, 1991). A key design issue related to such liquid rocket fuel and oxidizer pumps is the minimum pressure that the design can tolerate for a given inlet temperature and rotating speed. At low inlet pressure (to reduce tank weight) and high pump rotational speeds (to reduce engine weight), cavitation is prone to appear in the inducer section. To date, there is no established method in industry to estimate the actual loads due to cavitation on inducer blades. There have been methods proposed, each with its limited validity and challenges (Garcia, 2001). Most rocket engine systems designed in the U.S. have experienced issues with cavitating elements in the pump. This includes recent programs such as, alternate turbo pump (ATP) for the space shuttle main engine (SSME), the Fastrac LOX pump, and the RS-68 commercial engine (Garcia, 2001). An integrated framework based on computational modeling and control strategies is desirable to treat this critical and difficult issue. It is clear that the design of efficient turbomachinery components requires understanding and accurate prediction of the cryogenic cavitating flows.

Cavitating flow computations have been conducted using both density-based (Merkle et al., 1998; Kunz et al., 2000; Ahuja et al., 2001; Venkateswaran et al., 2002) and pressure-based

numerical approaches (Athavale and Singhal, 2001; Singhal et al., 2002; Senocak and Shyy, 2002, 2004a-b), with the cavitation models developed based on: (1) Rayleigh-Plesset type of bubble formulation (Kubota et al., 1992), which separates the liquid and the vapor regions based on the force balance notion, and (2) homogeneous fluid approach (Senocak and Shyy, 2002), which treats the cavity as a region consisting of continuous mixture of liquid and vapor phases. In the homogeneous fluid model, the density field is commonly modeled *via* either a generalized equation of state (Edwards et al., 2000, Ventikos and Tzabiras, 2000) or a transport equation of the liquid/vapor phase fraction (Merkle et al., 1998; Kunz et al., 2000; Senocak and Shyy, 2002, 2004a-b; Utturkar et al., 2005b). Recent efforts made in computational and modeling aspects of cavitating flows are discussed by Wang et al. (2001), Ahuja et al. (2001), Preston et al. (2001), Venkateswaran et al. (2002), Senocak and Shyy (2004a-b), Utturkar et al. (2005a-b), and references within.

Influence of Thermal Environment on Cavitation Modeling

To date, the majority of the cavitation modeling efforts have focused on the assumption that cavitation occurs with negligible energy interactions (isothermal condition). This assumption is reasonable for cavitation in non-cryogenic fluids but fails for thermo-sensible fluids, like liquid hydrogen and liquid oxygen (cryogens), due to the differences in material properties (low liquid-vapor density ratio, low thermal conductivities, steep slope of pressure-temperature saturation curves etc.), and the coupling of thermal effects such as, variation in vapor pressure/density with temperature etc. (Utturkar et al., 2005a-b; Utturkar, 2005; Hosangadi and Ahuja, 2005). Figure 6-1 (Lemmon et al., 2002) illustrates the behavior of the physical properties of two representative cryogens, liquid nitrogen and liquid hydrogen, in the liquid-vapor saturation regime. The temperature range in the plots is chosen based on the general operating condition of the fluids, which is close to the critical point. We observe substantial variation in the

material properties with changes in the temperatures. Relatively, the variation in the material properties (vapor pressure, liquid-vapor density ratio, and latent heat of vaporization, etc.) with temperature for liquid hydrogen is higher than that observed for liquid nitrogen.

As summarized by Utturkar (2005), dynamic similarity in case of isothermal cavitation is dictated by the cavitation number σ (Equation (6.1) with constant vapor pressure p_v). In the context of cryogenic cavitation, the actual cavitation number needs to be defined as follows (Brennen, 1994):

$$\sigma = \frac{p_\infty - p_v(T_c)}{0.5\rho_l U_\infty^2}, \quad (6.1)$$

where p_∞ is the reference pressure, U_∞ is the reference velocity, ρ_l is liquid density, and T_c is the temperature in the cavity. The local cavitation number can be related to the far-field cavitation number (based on the vapor pressure there) by the following first-order approximation (Brennen, 1994):

$$\frac{1}{2}\rho_l U_\infty^2(\sigma - \sigma_\infty) = \frac{dp_v}{dT}(T_\infty - T_c). \quad (6.2)$$

Equation (6.2) clearly indicates that the cumulative effect of the aforesaid factors would produce a noticeable rise in the local cavitation number, and subsequently, suppress the intensity of cavitation. Representative values of the vapor pressure gradients ($\frac{dp_v}{dT}$) in the operating temperature regime for liquid nitrogen and hydrogen are 20kPa/K and 37kPa/K, respectively.

The influence of thermal effects on cavitation has been numerically and experimentally investigated as early as 1956. Stahl and Stepanoff (1956) introduced a ‘B-factor’ method to estimate temperature drop in terms of the ratio of vapor volume to liquid volume during vaporization process; and used it to appraise head depression due to thermodynamic effects in

cryogenic cavitation. Gelder et al. (1966), Ruggeri and Moore (1969), and Hord (1973a) simplified, and extended this B-factor theory to account for dynamic effects *via* bubble growth, varying cavity thickness, and convective heat transfer. Holl et al. (1975) presented an entrainment theory to correlate the temperature depression and flow parameters. Cooper (1967) used a non-dimensional vaporization parameter along with a barotropic equation of state to incorporate pressure depression due to thermal effects while numerically simulating liquid hydrogen pumps. Brennen (1994, 1995), and Franc et al. (2003) presented methods of assessing thermodynamic effects on bubble dynamics by incorporating it into Rayleigh-Plesset equation. We refer the reader to the works by Hosangadi and Ahuja (2005), and Utturkar (2005) for more insight in the application regime and pros/cons of these methods.

Experimental and Numerical Modeling of Cryogenic Cavitation

Hord (1973a-b) conducted by far the most comprehensive experiments on cryogenic cavitation with liquid nitrogen and liquid hydrogen, under different sets of inlet velocities and temperature conditions, and employing a variety of geometries (hydrofoil and ogives of varying diameters). Temperature and pressure data in the cavitating region, which have been commonly employed for numerical validation (Hosangadi and Ahuja, 2003), was acquired over the geometries at regular spatial intervals using thermocouples and pressure sensors.

There have been limited computational studies for cryogenic cavitating flows. The key challenges for numerical computations are the presence of strong non-linearity in the energy equation and the temperature dependence of physical properties (Lemmon et al., 2002) such as, vapor pressure and density ratio (as seen from Figure 6-1(A) and Figure 6-1(B)). The main features of a few selected numerical studies (Reboud et al., 1990; Delannoy and Reboud, 1993; Deshpande et al., 1997; Lertnuwat et al., 2001; Tokumasu et al., 2002, 2003; Hosangadi and

Ahuja, 2003; Hosangadi et al., 2003; Rachid, 2003; Rapposelli and Agostino, 2003; Utturkar et al., 2005a-b) are summarized in Table 6-1.

A transport-based cavitation model, proposed by Merkle et al. (1998), has been adopted in multiple efforts for non-cryogenic conditions. The same basic framework can also be used to simulate cryogenic cavitating flows, subject to proper modification of the model parameters to better reflect the transport properties of cryogenic fluids and physical mechanisms of the flow environment. Utturkar et al. (2005b) showed that the accuracy of predictions is affected by the model parameters, and there is a need to calibrate the model parameters to account for cavitation in cryogenic conditions. As discussed earlier, the temperature dependent material properties also play a significant role in the predictions. These material properties are typically obtained from the models developed using the experimental data, and, naturally, contain uncertainties. The numerical approach employed in the present study has been previously tested, and documented against different flow problems (Utturkar et al. 2005a-b; Utturkar, 2005). Our focus here is to address the validation aspect, namely, to what extent a transport-based cavitation model can reproduce the cryogenic cavitation physics, and how can we improve its performance. Furthermore, realizing that the fluid properties and thermal environments add further challenges to cavitation models, the interplay between fluid and flow characteristics will also be probed.

Surrogate Modeling Framework

To facilitate the formulation of a suitable mathematical framework to probe the global sensitivity (Sobol, 1993) of the above-mentioned cavitation model and uncertainties in fluid properties in cryogenic environment, we first construct suitable surrogate models (Queipo et al., 2005). Since the fidelity of surrogate models is critical in determining the success of the sensitivity analysis and model validation, we adopt multiple surrogate models to help ascertain the performance measures. There are alternative surrogate models (for example, polynomial

response surface, kriging etc.) but the model that represents a particular function the best is not known *a priori*. Consequently, the predictions using different surrogate models have a certain amount of uncertainty. Goel et al. (2006b) suggested that the simultaneous use of multiple surrogate models may be beneficial to quantify, and to reduce uncertainties in predictions. They also proposed a cross-validation-error based weighted average surrogate model that was shown to represent a wide variety of test problems very well. The global cross validation error used in this study is also known as predicted residual sum of squares (PRESS) in polynomial response surface approximation terminology. Here, we used four surrogate models, polynomial response surface approximation (PRS), kriging (KRG), radial basis neural network (RBNN), and PRESS-based weighted average surrogate (PWS) model constructed using these three surrogates.

These surrogate models are used to calibrate the model parameters of the present transport based cavitation model (Merkle et al., 1998) in cryogenic conditions. While the surrogate model approach has become popular for fluid device optimization (Li and Padula, 2005; Queipo et al., 2005; Goel et al., 2006d), its application in CFD model validation and improvement has not yet been actively pursued. The present work represents such an endeavor.

Scope and Organization

Specifically, the objectives of this paper are:

- To study the physical aspects of cavitation dynamics in cryogenic environment and perform model (and code) validation,
- To conduct a global sensitivity analysis to assess the sensitivity of the response to temperature dependent material properties and model parameters, and
- To calibrate the parameters of a transport-based cryogenic cavitation model for suitable flow conditions, and to account for different fluid properties.

The organization of this chapter is as follows. The governing equations and the numerical approach followed in this paper are described in next section. Afterwards, we present results of

global sensitivity analysis to measure the relative importance of different model parameters and uncertainties in material properties; and calibration of model parameters. The influence of thermal effects on the cavitation model is studied in detail in a following section. Finally, we summarize the major outcome of this study.

Governing Equations and Numerical Approach

The set of governing equations for cryogenic cavitation under the homogeneous-fluid modeling strategy comprises the conservative form of the Favre-averaged Navier-Stokes equations, the enthalpy equation, the $k - \varepsilon$ two-equation turbulence closure, and a transport equation for the liquid volume fraction. The mass-continuity, momentum, enthalpy, and cavitation model equations are given below:

$$\frac{\partial \rho_m}{\partial t} + \frac{\partial(\rho_m u_j)}{\partial x_j} = 0, \quad (6.3)$$

$$\frac{\partial(\rho_m u_i)}{\partial t} + \frac{\partial(\rho_m u_i u_j)}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} [(\mu + \mu_t) \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3} \frac{\partial u_k}{\partial x_k} \delta_{ij} \right)], \quad (6.4)$$

$$\frac{\partial}{\partial t} [\rho_m (h + f_v L)] + \frac{\partial}{\partial x_j} [\rho_m u_j (h + f_v L)] = \frac{\partial}{\partial x_j} \left[\left(\frac{\mu_m}{\text{Pr}_m} + \frac{\mu_t}{\text{Pr}_t} \right) \frac{\partial h}{\partial x_j} \right], \quad (6.5)$$

$$\frac{\partial \alpha_l}{\partial t} + \frac{\partial(\alpha_l u_j)}{\partial x_j} = \dot{m}^+ + \dot{m}^-, \quad (6.6)$$

where ρ_m is the density of the fluid-vapor mixture, u_j denotes the components of velocity, p is pressure, μ_m, μ_t are mixture laminar and turbulent viscosities, respectively, h is sensible enthalpy, f_v is the vapor mass fraction, L is the latent heat of vaporization, Pr is the Prandtl number, α_l is the fraction of liquid in the mixture, and \dot{m}^+, \dot{m}^- are the source terms for the cavitation model. The subscript ‘ t ’ denotes turbulent properties, ‘ l ’ represents the liquid state, ‘ v ’

represents the vapor state, and ‘ m ’ denotes the mixture properties. The mixture property ϕ_m , sensible enthalpy, and the vapor mass fraction are respectively expressed as

$$\phi_m = \phi_l \alpha_l + \phi_v (1 - \alpha_l), \quad (6.7)$$

$$h = C_{pm} T, \quad (6.8)$$

$$f_v = \frac{\rho_v (1 - \alpha_l)}{\rho_m}. \quad (6.9)$$

For the problems studied here, we neglect the effects of kinetic energy and viscous dissipation terms in the energy equation (6.5) ($O(\text{Re}^{-0.5})$, $\text{Re} \sim O(10^6)$), because the temperature field in cryogenic cavitation is mainly dictated by the phenomenon of evaporative cooling (refer to the section on the effect of thermal environment).

Transport-based Cavitation Model

Physically, the cavitation process is governed by thermodynamics and kinetics of the phase change process. The liquid-vapor conversion associated with the cavitation process is modeled through \dot{m}^+ and \dot{m}^- terms in Equation (6.6), which represent condensation and evaporation, respectively. The particular form of these phase transformation rates, which in case of cryogenic fluids also dictate the heat transfer process, forms the basis of the cavitation model. The liquid-vapor condensation rates for the present transport based cavitation model (Merkle et al., 1998) are:

$$\dot{m}^- = \frac{C_{dest} \rho_l \min(0, p - p_v) \alpha_l}{\rho_v (0.5 \rho_l U_\infty^2) t_\infty}; \quad \dot{m}^+ = \frac{C_{prod} \max(0, p - p_v) (1 - \alpha_l)}{(0.5 \rho_l U_\infty^2) t_\infty}, \quad (6.10)$$

where C_{dest} and C_{prod} are the empirical model parameters controlling the evaporation and condensation rates, p_v is the vapor pressure, ρ_v, ρ_l are the vapor and liquid densities, U_∞ is the reference velocity scale, and t_∞ is the reference time scale, defined as the ratio of the

characteristic length scale D to the reference velocity scale U_∞ ($t_\infty = D/U_\infty$). Merkle et al. (1998) validated this cavitation model with the experimental data for non-cryogenic fluids (e.g., water) and specified $C_{dest} = 1.0$, and $C_{prod} = 80.0$ as optimal model parameters (referred here as ‘original’ parameters). However, Utturkar (2005), and Hosangadi and Ahuja (2005) found that the previously calibrated values of the Merkle et al. (1998) cavitation model ($C_{dest} = 1.0$ and $C_{prod} = 80.0$) are inadequate to provide a good match with the experimental data under the cryogenic condition. Consequently, Utturkar et al. (2005) suggested $C_{dest} = 0.68$ and $C_{prod} = 54.4$ (obtained *via* numerical experimentation) as more appropriate model parameters for liquid nitrogen. However, they noted difficulties in the simultaneous prediction of the temperature and pressure profiles on the surface of the test geometry. The present efforts represent advances in the practice of multi-surrogate model approach for code validation.

Thermodynamic Effects

The evaporation and condensation processes result in absorption and release of the latent heat of vaporization that regulates the thermal effects. Furthermore, there is a significant variation in the physical properties ($\rho_l, \rho_v, p_v, \mu, C_p, K$, and L) with temperature (Lemmon et al., 2002) in the operating range that manifests coupling between different governing equations, and underscores the importance of thermal effects in cryogenic cavitation. As indicated by phase diagram in Figure 6-1(D), the physical properties (liquid and vapor densities) are much stronger functions of temperature than pressure, and one can fairly assume the respective phase values on the liquid-vapor saturation curve at a given temperature.

We illustrate the impact of thermal effects in cryogenic environment due to phase change on temperature predictions, and thermo-sensible material properties on temperature and pressure

predictions by analyzing energy equation and cavitation sources terms. Firstly, we separate the latent heat terms in the energy equation (Equation (6.5)) onto the right-hand-side to obtain temperature-based form of the energy equation as follows,

$$\frac{\partial}{\partial t}[\rho_m C_{pm} T] + \frac{\partial}{\partial x_j}[\rho_m u_j C_{pm} T] = \frac{\partial}{\partial x_j} \left[C_{pm} \left(\frac{\mu_m}{Pr_{Lm}} + \frac{\mu_t}{Pr_t} \right) \frac{\partial T}{\partial x_j} \right] - \underbrace{\left\{ \frac{\partial}{\partial t}[\rho_m (f_v L)] + \frac{\partial}{\partial x_j}[\rho_m u_j (f_v L)] \right\}}_{\text{energy source/sink term}}. \quad (6.11)$$

As can be seen from Equation (6.11), the ‘lumped’ latent heat terms manifest as a non-linear source term into the energy equation, and physically represent the latent heat transfer rate or the influence of phase change during evaporation and condensation. The spatial variation of thermodynamic properties and the evaporative cooling effect are intrinsically embedded into this transport-based source term causing a coupling of all the governing equations.

The influence of thermal effects due to thermo-sensible material properties can be further analyzed by studying the cavitation source terms (Equation (6.10)) more closely. Firstly, we consider a case when $p - p_v < 0$, i.e., $\dot{m}^+ = 0$, and the evaporation source term can be written as,

$$\dot{m}^- = \frac{C_{dest} \alpha_l}{t_\infty} \left(\frac{\rho_l}{\rho_v} \right) \left(\frac{p - p_v}{0.5 \rho_l U_\infty^2} \right) = \beta R(T) \sigma(T), \quad (6.12)$$

where $\beta = C_{dest} \alpha_l / t_\infty$, R is the temperature dependent liquid-vapor density ratio, and σ is the cavitation number. Expanding Equation (6.12) using the Taylor’s series and utilizing Equation (6.2), we get,

$$\dot{m}^- = \beta \left(R(T_\infty) + \left. \frac{dR}{dT} \right|_{T_\infty} \Delta T + \dots \right) \left(\sigma(T_\infty) + \left. \frac{d\sigma}{dT} \right|_{T_\infty} \frac{\Delta T}{0.5 \rho_l U_\infty^2} + \dots \right), \quad (6.13)$$

$$\dot{m}^- = \beta \left(R(T_\infty) \sigma(T_\infty) + \sigma(T_\infty) \left. \frac{dR}{dT} \right|_{T_\infty} \Delta T + R(T_\infty) \left. \frac{dp_v}{dT} \right|_{T_\infty} \frac{\Delta T}{0.5 \rho_l U_\infty^2} + O(\Delta T^2) \right). \quad (6.14)$$

Similarly, we can do an analysis of condensation source term for the condition $p - p_v > 0$, such that $\dot{m}^- = 0$. Then,

$$\dot{m}^+ = \frac{C_{prod}(1-\alpha_l)}{t_\infty} \left(\frac{p - p_v}{0.5 \rho_l U_\infty^2} \right) = \gamma \sigma(T), \quad (6.15)$$

where $\gamma = \frac{C_{prod}(1-\alpha_l)}{t_\infty}$. As before, using Taylor's series,

$$\dot{m}^+ = \gamma \left(\sigma(T_\infty) + \left. \frac{dp_v}{dT} \right|_{T_\infty} \frac{\Delta T}{0.5 \rho_l U_\infty^2} + O(\Delta T^2) \right). \quad (6.16)$$

It can be concluded from Equations (6.14) and (6.16) that the thermal effects influence the cavitation source terms in two ways, (1) thermal rate of change of liquid-vapor density ratio $\left. \frac{dR}{dT} \right|_{T_\infty}$, which is negative (Figure 6-1(B)), and (2) thermal rate of change of vapor pressure $\left. \frac{dp_v}{dT} \right|_{T_\infty}$, which is positive (Figure 6-1(A)), thus illustrating competing influences of thermal effects. It is obvious that the degree of influence of thermal effects depends on the choice of operating fluid, and the operating conditions (T_∞, p_∞) due to the non-linear variation of material properties with temperature.

Speed of Sound (SoS) Model

Numerical modeling of sound propagation is a very important factor in accurate prediction of cavitation in liquid-vapor multiphase mixture. The speed of sound affects the numerical calculation *via* the pressure correction equation by conditionally endowing it with a convective-diffusive form in the mixture region. Past studies (Senocak, 2002; Senocak and Shyy, 2002; Wu

et al., 2003; Senocak and Shyy, 2004a) discuss in detail the modeling options, their impact and issues. The SoS model used here is outlined below,

$$\text{SoS} = C_\rho = C(1 - \alpha_l). \quad (6.17)$$

The density correction term in the continuity equation is thus coupled to the pressure correction term as shown below.

$$\rho' = C_\rho p'. \quad (6.18)$$

In the pure liquid region, we recover the diffusive form of the pressure equation. Senocak and Shyy (2002, 2004a-b) suggested an $O(1)$ value for the constant C to expedite the convergence of the iterative computational algorithm. However, their recommendation is valid under normalized values for inlet velocity and liquid density. Since we employ dimensional form of equations for cryogenic fluids, we suggest an $O(1/U_\infty^2)$ value for C (Utturkar, 2005), which is consistent with the above suggestion in terms of the Mach number regime.

Turbulence Model

The $k - \varepsilon$ two-equation turbulence model with wall functions is presented as follows (Launder and Spalding, 1974):

$$\frac{\partial(\rho_m k)}{\partial t} + \frac{\partial(\rho_m u_j k)}{\partial x_j} = P_t - \rho_m \varepsilon + \frac{\partial}{\partial x_j} \left[\left(\mu + \frac{\mu_t}{\sigma_k} \right) \frac{\partial k}{\partial x_j} \right], \quad (6.19)$$

$$\frac{\partial(\rho_m \varepsilon)}{\partial t} + \frac{\partial(\rho_m u_j \varepsilon)}{\partial x_j} = C_{\varepsilon_1} \frac{\varepsilon}{k} P_t - C_{\varepsilon_2} \rho_m \frac{\varepsilon^2}{k} + \frac{\partial}{\partial x_j} \left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon} \right) \frac{\partial \varepsilon}{\partial x_j} \right]. \quad (6.20)$$

The turbulence production term (P_t) and the Reynolds stress tensor are defined as:

$$P_t = \tau_{ij} \frac{\partial u_i}{\partial x_j}; \tau_{ij} = -\overline{\rho_m u'_i u'_j}, \quad (6.21)$$

$$\rho_m u'_i u'_j = \frac{2\rho_m k \delta_{ij}}{3} - \mu_t \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right).$$

The parameters for this model, $C_{\varepsilon_1} = 1.44$, $C_{\varepsilon_2} = 1.92$, $\sigma_\varepsilon = 1.3$, $\sigma_k = 1.0$ are adopted from the equilibrium shear flow calibration (Shyy et al., 1997) and the turbulent viscosity is defined as:

$$\mu_t = \frac{\rho_m C_\mu k^2}{\varepsilon}, C_\mu = 0.09. \quad (6.22)$$

Of course, the turbulence closure and the eddy viscosity levels can affect the outcome of the simulated cavitation dynamics, especially, in case of unsteady simulations. For detailed investigations of turbulence modeling on cavitating flow computations, we refer to recent works by Wu (2005), Wu et al. (2003a-c, 2005), and Utturkar et al. (2005). Vaidyanathan et al. (2003) conducted a sensitivity analysis to assess the interplay between turbulence model parameters and the cavitation model parameters in non-cryogenic environment. They observed that multiple combinations of turbulence parameters and cavitation model parameters yield same performance.

To appraise the influence of turbulence modeling on the current problem, we follow the previous investigation by Vaidyanathan et al. (2003); and compare the standard $k - \varepsilon$ turbulence model (Launder and Spalding, 1974) and a non-equilibrium $k - \varepsilon$ turbulence model developed by Shyy et al. (1997) that accounts for absence of equilibrium between the production and destruction of dissipation of turbulent kinetic energy. Both turbulence models offered very similar predictions within the experimental uncertainties (refer to the supplementary results at the end of this chapter). Hence, we restrict the scope of this study to the calibration of cryogenic cavitation model parameters with the standard $k - \varepsilon$ turbulence model.

Numerical Approach

The governing equations are numerically solved using a CFD code STREAM (Thakur et al., 2002) based on a pressure-based algorithm and the finite-volume approach. We use multi-block, structured, curvilinear grids to analyze flow over different geometries in this chapter. The viscous terms are discretized by second-order accurate central differencing while the convective

terms are approximated by the second-order accurate controlled variations scheme (CVS) (Shyy, 1994; Shyy and Thakur, 1994). The use of CVS scheme prevents oscillations under sharp gradients caused by the evaporation source term in the cavitation model while retaining second order of formal accuracy. The pressure-velocity coupling is implemented through the extension of the SIMPLEC (Patankar, 1980; Versteeg and Malalasekara, 1995) type of algorithm cast in a combined Cartesian-contravariant formulation (Thakur et al., 2002) for the dependent and flux variables, respectively, followed by adequate relaxation for each governing equation, to obtain steady-state results. The temperature dependent material properties are updated from the NIST (Lemmon et al., 2002) database at the end of each computational iteration.

Results and Discussion

Test Geometry, Boundary Conditions, and Performance Indicators

We simulate flows over a 2-D hydrofoil and an ogive in cryogenic environment, which serve as the benchmark problems for validating the cryogenic cavitation models. Hord (1973a-b) experimentally investigated the flow over these geometries inside suitably designed wind-tunnels (Figure 6-2(A)). He reported average pressure and temperature data at five probe locations over the body surface for different cases that are referenced alpha-numerically in two reports (Hord 1973a-b). We employ (1) Case ‘290C’ for liquid nitrogen ($Re = 9.1 \times 10^6$, $\sigma_\infty = 1.7$, $T_\infty = 83.06K$, liquid-vapor density ratio = 95, hydrofoil), and (2) Case ‘249D’ for liquid hydrogen ($Re = 2.0 \times 10^7$, $\sigma_\infty = 1.57$, $T_\infty = 20.70K$, liquid-vapor density ratio = 47, hydrofoil), to conduct optimization and sensitivity studies.

A simplified geometry, schematic computational domain, and the boundary conditions for the two test problems are shown in Figure 6-2. The computational grids consists of 320×70 and 340×70 non-uniformly distributed grid points for the hydrofoil and the ogive, respectively, such

that the cavitation regime is adequately resolved, and the deployment of wall functions (Rogallo and Moin, 1984) near the no-slip boundary conditions is allowed (Utturkar et al., 2005a-b). The inlet boundary conditions are implemented by stipulating the values of the velocity components, phase fraction, temperature, and turbulence quantities from the experimental data (Hord, 1973a-b). At the walls, pressure, phase fraction, and turbulence quantities are extrapolated, along with applying the no-slip (in the form of the wall function, Versteeg and Malalasekara, 1995) on velocity, and adiabatic conditions on temperature. Pressure and other variables are extrapolated at the outlet boundaries while enforcing global mass conservation by rectifying the outlet velocity components. In addition, we hold the pressure at the reference point (illustrated in the experimental reports, Hord, 1973a-b) constant at the reference value specified in the experiments. Though the cavitating flows are unsteady in nature, no time-dependent data was reported by Hord. Utturkar (2005) showed that the flows considered here can be modeled as steady state. Furthermore, for sheet cavitating flows, it has been shown by Senocak and Shyy (2002) that steady state computations can well capture the essential flow features, and reach close agreement with the measurements. Consequently, we modeled the flow as steady state.

The quality of the predictions is numerically quantified by computing the L_2 norms of the differences between computed and experimental values of pressure (P_{diff}) and temperature (T_{diff}) at each of the five probe locations on the surface of hydrofoils. These metrics are desired to be low to obtain good prediction quality.

Surrogates-based Global Sensitivity Assessment and Calibration

Since minor changes in flow environment can lead to substantial changes in the predictions in cryogenic environment (Utturkar et al., 2005b), it is imperative to appraise the role of model parameters and uncertainties in material properties on the predictions. In this section, we

characterize the parameters that significantly affect predictions using surrogate-based global sensitivity analysis (GSA), and then calibrate the cryogenic cavitation model parameters. In the following, we present in detail the process of model parameter optimization and sensitivity evaluations based on Case ‘290C’ for liquid nitrogen. A corresponding study based on Case ‘249D’ for liquid hydrogen has also been carried out. To save space, we do not repeat the detailed information and only report the outcome.

Global Sensitivity Assessment

We employ variance-based, non-parametric GSA method, proposed by Sobol (1993) (refer to Appendix C), to evaluate the sensitivity of cryogenic cavitation model with respect to model parameters and material properties, and to get an insight about the factors that influence the accuracy of predictions. We can study the influence of uncertainty in different material properties ρ_v, ρ_l, K, C_p , and L , and model parameters C_{dest}, C_{prod} , and t_∞ . However, to keep Re (Reynolds number based on upstream flow) and σ_∞ (cavitation number based on upstream flow) constant for the given case, and to keep the computational expense reasonable, we select one material property each from energy equation and cavitation transport α -equation. Consequently, we choose $C_{dest}, C_{prod}, \rho_v$, and L as variables. The model parameters, C_{dest} and C_{prod} , are perturbed on either side of the values proposed by Utturkar (2005) ($C_{dest} = 0.68$; $C_{prod} = 54.4$) by 15%, and the material properties, ρ_v and L , are perturbed within 10% of the values they assume from the NIST database (Lemmon et al., 2002) (perturbations are denoted as ρ_v^* and L^*). The ranges of the variables are given in the Table 6-2. The performance of the cryogenic cavitation model is characterized by prediction errors P_{diff} and T_{diff} , defined in the previous section.

To conduct a global sensitivity analysis, the response function is decomposed into additive functions of variables and interactions among variables. This allows the total variance (V) in the response function to be expressed as a combination of the main effect of each variable (V_i), and its interactions with other variables (V_{iZ}). The sensitivity of the response function with respect to any variable is measured by computing its sensitivity indices. The sensitivity indices of main effect (S_i) and total effect (S_i^{total}) of a variable are given as follows:

$$S_i = V_i / V, \quad S_i^{total} = (V_i + V_{iZ}) / V. \quad (6.23)$$

Surrogate construction

In the absence of a closed form solution characterizing the objective functions (P_{diff} and T_{diff}), different components of variances are evaluated using numerical integration. Since direct coupling of CFD simulations with numerical integration schemes is computationally expensive, we use surrogate models of the performance indicators. We evaluate the responses P_{diff} and T_{diff} using CFD simulations at 70 data points (combinations of variables) selected *via* face-centered cubic composite design (FCCD, 25 points), and Latin hypercube sampling (LHS, 45 points) experimental designs. We construct four surrogates: polynomial response surface approximation (PRS), kriging (KRG), radial basis neural network (RBNN), and a PRESS-based weighted average surrogate (PWS) model (refer to Chapters 2 and 4) of both responses in scaled variable space (all variables are scaled between zero and one, such that zero corresponds to the minimum value). We use reduced cubic polynomials for PRS, and specify ‘spread’ value of 0.5 for RBNN. Relevant details of the quality of fit of surrogate models are summarized in Table 6-3. Low PRESS and low root mean square error compared to the range of the function indicate that the two responses are adequately approximated by the surrogate models. For both objectives, RBNN

surrogate is the worst of the three surrogates, and kriging is the best (compare PRESS errors).

The contribution of different surrogate models to the PWS model, given by the weights in Table 6-3, accounts for the poor performance of RBNN by assigning it a low weight. We employ Gauss-quadrature integration scheme, with ten Gaussian points along each direction, to evaluate sensitivity indices. Responses at the Gaussian points are evaluated using surrogate models.

The influence of the choice of surrogate model on the prediction of sensitivity indices is illustrated in Figure 6-3 with the help of main effects for P_{diff} . Since all surrogates predict similar trend about the importance of different variables, we may conclude that the variability in predictions due to the choice of surrogate model is small.

Main and interaction effects of different variables

We show sensitivity indices of main effects and total effects (estimated *via* PWS model) in Figure 6-4 to quantify the relative importance of different parameters on P_{diff} and T_{diff} . The sensitivity indices of the main effects (pie-charts) suggest that C_{dest} is the most influential, and C_{prod} is the least influential parameter within the selected range of variation, i.e., the cavity morphology is more influenced by the evaporation rate term, compared to the condensation term that is influential in determining the pressure recovery rate posterior to cavity closure. The variability due to material properties as indicated by the sensitivity indices associated with vapor density ρ_v and latent heat L , is smaller compared to the model parameters, but is not negligible. The variability in vapor density ρ_v influences both pressure and temperature predictions, but has more significant impact on pressure predictions. On the other hand, the variability in latent heat (L) within the selected uncertainty range affects temperature predictions only. Relatively moderate influence of variation in latent heat on temperature predictions does not lead to significant variation in pressure predictions, because the latter is more significantly influenced by

the parameters that directly appear in cavitation source terms, and have more pronounced effect. The differences between the main and the total sensitivity indices for both P_{diff} and T_{diff} highlight the importance of interaction among parameters. The interaction between C_{dest} and ρ_v is particularly stronger than other parameters.

Validation of global sensitivity analysis

We validate the results of global sensitivity analysis by evaluating the variation in responses P_{diff} and T_{diff} , when only one parameter is changed at a time and remaining parameters are fixed at their mean values (mean of the selected range). We assign six equi-spaced levels for each variable and calculate the variation in responses using Abramowitz and Stegun (1972) six-point numerical integration scheme that has seventh order accuracy. The sensitivity indices of main effects of different parameters on P_{diff} and T_{diff} are shown in Figure 6-5. The results obtained by actual computations are in-sync with the findings of the global sensitivity analysis, that is, the model parameter C_{dest} and uncertainty in vapor pressure ρ_v are the most influential parameters for accurate pressure and temperature predictions, and uncertainty in latent heat L is important only for predicting temperature accurately. The differences in the actual magnitude of sensitivity analysis results can be explained by accounting for, (1) the small number of points used for actual sensitivity computations, (2) neglect of interaction terms, and (3) the errors in surrogate modeling. Nevertheless, the important trends in the results are captured adequately.

The results indicate that the performance of cryogenic cavitation model is more susceptible to the variability in temperature-dependent vapor density ρ_v , compared to the variability in latent heat L . This calls for more attention in developing accurate models of ρ_v . Also, the variables, which appear in cavitation source terms (\dot{m}^- or \dot{m}^+), may tend to register greater influence on

the computed results. Thus, intuitively, reference velocity U_∞ , reference time scale t_∞ , and liquid density ρ_l , which are omitted from the present GSA, are expected to induce large variability in the computation, as compared to other omitted properties such as thermal conductivity K , and specific heat C_p . Furthermore, as depicted by sensitivity indices in Figure 6-4, largely the impact of different parameters is expected to be the same on pressure and temperature due to the tight coupling between various flow variables.

Calibration of Cryogenic Cavitation Model

In the previous section, we observed that one of the model parameters C_{dest} significantly influences the performance of the present cryogenic cavitation model. This information can be used to calibrate the present cavitation model parameters associated with different fluids. Firstly, we optimize the model parameter (C_{dest}) of the present cryogenic cavitation model using the benchmark case of liquid nitrogen flow over a hydrofoil ‘290C’, while fixing the model parameter C_{prod} at 54.4 (minimal influence on predictions) and assuming the temperature dependent material properties obtained from the NIST database (Lemmon et al., 2002) to be accurate.

We observed that increasing C_{dest} , increases P_{diff} and decreases T_{diff} . As shown in Figure 6-6, the parameters, which yield good pressure predictions (low P_{diff}), produce large errors in temperature predictions (high T_{diff}) and *vice-versa* (low T_{diff} but high P_{diff}). So this model calibration/system identification problem is a multi-objective optimization to simultaneously minimize P_{diff} and T_{diff} by varying the model parameter C_{dest} . Since the cavitation dynamics primarily impacts pressure fluctuations, we seek to improve the pressure prediction capabilities of the present cryogenic cavitation model without incurring a significant deterioration of

temperature predictions. Consequently, we allow the model parameter C_{dest} to vary between 0.578 and 0.68.

Surrogate modeling of responses

To represent the responses P_{diff} and T_{diff} using surrogate models, we sample data using CFD simulations at nine locations. The location of points, and corresponding P_{diff} and T_{diff} , shown in Figure 6-7, clearly exhibit the conflicting nature of the two objectives. As before, we construct PRS, KRG, RBNN, and PWS models. We approximate P_{diff} with a reduced cubic PRS and T_{diff} with a reduced quintic PRS. The relevant metrics, depicting the quality of surrogate models, and the weights associated with different surrogates in PWS model are summarized in Table 6-4. Low PRESS and low RMS error indicate that the two responses are well represented by all surrogate models. While no single surrogate model performs the best for both responses, RBNN is the worst of the three surrogates considered here. The weights associated with different surrogates in PWS model also reflect the same. PRESS and maximum error measures indicate that the PWS model obtained by averaging different surrogate models performs significantly better than the worst surrogate, and the performance is comparable to the best surrogate.

Multi-objective optimization

Different methods to solve multi-objective optimization problems can be obtained from different texts (Chankong and Haimes, 1983; Steuer, 1986; Sen and Yang, 1998; Miettinen, 1999; Deb, 2001). We convert the present multi-objective optimization problem into a single objective optimization problem by combining the two performance metrics (P_{diff} and T_{diff}) using weights (weighted sum strategy, Deb 2001); or by treating one performance metric as objective function, and the second performance metric as constraint function (ϵ -constraint strategy, Chankong and Haimes, 1983). We obtain many candidate Pareto optimal solutions by varying

the weights for weighted sum strategy, and constraint values for ε -constraint strategy. After removing dominated and duplicate solutions from the set of candidate solutions, the function space and the variable space illustration of Pareto optimal front (POF) obtained from different surrogate models is shown in Figure 6-8. We observe that different POF obtained by using multiple surrogate models are close to one another in both function and variable space. All surrogate models predict that a small increase in T_{diff} will lead to significant reduction in the P_{diff} (Figure 6-8(A)). However, we note that the pressure fluctuations play more important role in determining the cavity morphology and loadings on turbomachinery. Consequently, accurate pressure prediction is our primary objective.

We select a tradeoff solution on the POF for validation, such that noticeable reduction in P_{diff} can be realized without incurring significant deterioration of T_{diff} . Corresponding C_{dest} (referred as ‘best-compromise’ parameter), computed (*via* CFD simulations), and surrogate predictions of the two responses (P_{diff} and T_{diff}) are given in Table 6-5. The errors in predictions of P_{diff} , and T_{diff} are small for all surrogates, except RBNN. Clearly, PWS model yields the best predictions on both objectives. A graphical comparison of surface pressure and temperature profiles obtained with the original (Merkle et al., 1998) and optimal parameters of present transport based cavitation model is shown in Figure 6-9(A). The calibrated model parameters yield 72% reduction in P_{diff} by allowing 3.8% increase in T_{diff} compared to the original parameters (Merkle et al., 1998). The improvements in the surface pressure prediction, that is the more important criterion to estimate loadings due to cavitation, are obvious whereas the deterioration in the temperature predictions is small.

From cavitation dynamics point of view, the main issue with the predictions using original parameters was the poor prediction of the cavity closure region. The best-compromise model

parameters reduce the evaporation source term by reducing the model parameter C_{dest} . This change brings favorable changes in the cavity closure region by allowing an earlier onset of condensation, and hence, faster recovery of the pressure as was observed in experiments.

Optimization outcome for hydrogen

We repeat the model calibration exercise for liquid hydrogen fluid considering case ‘249D’ (hydrofoil) as the benchmark case. The corresponding ‘best-compromise’ C_{dest} parameter is found to be 0.767. Notably, the ratio of best-compromise and baseline value of C_{dest} for both nitrogen and hydrogen is 0.94. The surface pressure and temperature profiles shown in Figure 6-9(B) clearly demonstrate improvements in pressure predictions with the calibrated parameters compared to the original parameters (Merkle et al., 1998).

Validation of the calibrated cavitation model

The calibrated model parameters of the present cryogenic cavitation model are validated by simulating additional benchmark cases for two geometries (hydrofoil, Hord 1973a; and ogive Hord, 1973b) using different working fluids, liquid nitrogen and liquid hydrogen. The cases considered in this study, along with the corresponding best-compromise model parameters, are enlisted in Table 6-6. We compared the surface pressure and temperature profiles predicted using the cryogenic cavitation model with the calibrated (best-compromise) and the original model parameters (Merkle et al., 1998) in Figure 6-10. The model with best-compromise parameters exhibits substantially more robust performance for different geometries, fluids, and flow environments.

The results presented here clearly spell the merits of employing a systematic methodology to examine the role of cavitation model parameters. In the present case, the implications of the optimization on pressure and thermal fields are inconsistent. While this indicates the merits of

adopting a multi-objective optimization framework, as has been conducted here, it also suggests that there is a need for further investigation of the effect of thermal variations on cryogenic cavitating flows. It should also be reiterated that in terms of practical impact, the pressure prediction is our primary objective because pressure fluctuation is what causes poor performance or even catastrophic situation of fluid machinery. In the following, we offer further assessment of the thermal effect.

Investigation of Thermal Effects and Boundary Conditions

In the previous section, we observed discrepancies in simultaneous predictions of temperature and pressure. To understand the underlying issues related to thermodynamic effects, we study the influence of thermo-sensitive material properties and the role of thermal boundary condition on the hydrofoil wall for Case 290C. We use the best-compromise values of model parameters (liquid N₂) in all cases. Again, we use standard $k - \varepsilon$ turbulence model (refer to supplementary results at the end of this chapter).

Influence of Thermo-sensitive Material Properties

Firstly, we highlight the influence of thermal effects *via* phase change and thermo-sensitive properties on the temperature and pressure predictions in Figure 6-11. The difference of pressure and free stream vapor pressure ($p - p_v(T_\infty)$), and the difference of pressure and the actual vapor pressure (based on temperature, $p - p_v(T)$) are shown in Figure 6-11(A). The cavitation in cryogenic environment differs from non-cryogenic environment in two ways: (1) under-shoot at the leading edge of the hydrofoil indicates slower pressure recovery in cryogenic environment, and the influence of cooling due to heat absorption than that observed in the non-cryogenic environment, and (2) the vapor pressure in the cavity in cryogenic environment is not constant (continuous increase) due to the variation in temperature. This increase in vapor pressure (as

marked by Δp_v in Figure 6-11(A)) is attributed to the variation in temperature (Figure 6-1). The change in vapor pressure affects the cavitation source terms (Equations (6.14) and (6.16)) and resultant liquid-vapor fraction, which impacts the source terms in energy equation to enforce coupling of thermal effects in governing equations. To contrast the thermal effect on the cavitation dynamics, we also show a solution obtained by assigning a zero latent heat in Figure 6-11(A). With zero latent heat and an adiabatic wall condition, the fluid field exhibits a constant temperature throughout, resulting in a constant vapor pressure. This isothermal cavitation case yields a substantially larger cavity with near constant pressure on the surface inside cavity, which is quite different from the experimental measurement.

The temperature on the surface of hydrofoil in cavitating conditions is shown in Figure 6-11(B). The significant drop in temperature near the leading edge of the cavity is explained as follows. The phase change, as modeled, is dictated by the vapor pressure. When the local pressure in the flow falls below the vapor pressure, evaporation begins instantaneously as indicated by the transport model. This results in absorption of the latent heat of vaporization to facilitate the phase change. However, unlike boiling heat transfer where heat is continuously supplied through an external heat source, the heat transfer in cavitating flow largely stems from the convective and conductive heat transfer, and the latent heat release/absorption within the fluid, with external heat source playing minor roles. Consequently, a decrease in fluid temperature is observed in the cavity region. As we approach the cavity closure region, the condensation of fluid releases latent heat, increasing the fluid temperature locally. Furthermore, since the condensation process is dictated by the vapor pressure (with the local temperature effect exerted indirectly *via* the change in vapor pressure in response to the temperature field), the rate of latent heat release can be fast in comparison to the rate of convective and conductive

heat transfer, and consequently, in simulations, we observe an ‘overshoot’ in temperature profile. The experiments also show an increase in temperature of the fluid in the closure region but probably due to the lack of sufficient number of probes on the surface, the existence of the overshoot could not be ascertained.

Overall, the pressure predictions on the hydrofoil surface follow the same trends as observed in experiments. However, we note differences in predictions with experimental data near the closure region of the cavity.

Impact of Boundary Conditions

To investigate the discrepancy between experimental and predicted surface pressure and temperature profiles, we also assess the impact of different thermal boundary conditions on the predictions. While all the walls on the wind tunnel are modeled as adiabatic, the hydrofoil surface is modeled as either adiabatic (Neumann boundary), or specified temperature (Dirichlet boundary) wall. The temperature profile required for implementing Dirichlet boundary condition is obtained by interpolating/extrapolating the experimental temperature at five probe locations on the surface of hydrofoil.

The predicted pressure and temperature profiles on the surface of hydrofoil, obtained with different thermal boundary conditions, are compared with the experimental data (Hord, 1973a) in Figure 6-12. The introduction of heat transfer through the hydrofoil surface by Dirichlet boundary condition has little influence on the pressure distribution. With the given Reynolds number, the heat transfer at the hydrofoil surface is relatively small compared to the impact of latent heat, and subsequently, only minor variations in the vapor pressures are observed. In the cavity closure region, the latent heat released during condensation cannot be redistributed *via* convection and conduction fast enough, resulting in an overshoot in temperature there. The temperature profile on the first computation point above the hydrofoil surface, shown in Figure

6-12(C), also indicates that the effect of heat transfer due to Dirichlet boundary condition is largely restricted to the boundary, and has minimal influence on the flow inside the cavity.

Overall, it can be said that the effect of thermal boundary condition on the hydrofoil surface has little impact on the performance of the present cryogenic cavitation model.

Conclusions

In this chapter, we presented results of model validation and improvement of a transport-based cryogenic cavitation model using benchmark experimental data for 2-D hydrofoils and ogives provided by Hord (1973a-b). We used surrogate-based global sensitivity analysis to study the role of model parameters and uncertainties in temperature dependent material properties. The model parameters, originally used in present transport based cavitation model (Merkle et al., 1998), were calibrated for cryogenic environment using multiple surrogates, and optimization techniques. The main conclusions of this study are as follows.

- Performance of the current cryogenic cavitation model was more influenced by the model parameter associated with the evaporation source term (C_{dest}) than the uncertainty in material properties. The high sensitivity index associated with temperature dependent vapor density indicated significant impact on the accuracy of pressure and temperature predictions. The variations in the latent heat of vaporization influenced the accuracy of temperature predictions only. The model parameter associated with the production source term in the present cryogenic cavitation model (C_{prod}) did not influence predictions.
- The best-compromise model parameters selected for present transport based cavitation model (Merkle et al., 1998) were $C_{dest,LN2} = 0.6392$, $C_{dest,LH2} = 0.767$, and $C_{prod} = 54.4$. The choice of these parameters reduced the importance of evaporation source term, which resulted in earlier onset of the condensation and hence, the cavity closure. Utturkar et al. (2005b) had made adjustment based on trial-and-error ($C_{dest,LN2} = 0.68$, $C_{dest,LH2} = 0.816$, and $C_{prod} = 54.4$) and limited optimization processes. In their approach, there was a lack of probing in regard to the sensitivity and robustness of the outcome. The merits of the present effort lie in a systematic use of the optimization and sensitivity methodology, a detailed assessment of the thermal boundary condition, and a reasonably broad range of fluid and flow cases.
- Simultaneous use of multiple surrogate models evidently helped in increasing confidence in the results of global sensitivity analysis and optimization. The predictions using PRESS-

based weighted average surrogate model were more accurate than individual surrogate models.

- The impact of thermal boundary conditions on the prediction of flow was apparently not significant. However, the thermal effect caused by the phase change (latent heat) clearly affects the cavitation dynamics including the vapor pressure, and consequently, the cavity size. As we have shown here, the thermal effects play a very significant role in the accurate prediction of the pressure *via* phase change, and thermo-sensitive material properties in cryogenic environment, with little impact caused by wall heat transfer.
- The trends of the optimization on pressure and thermal fields follow opposite directions. While this indicates the usefulness of adopting a multi-objective optimization framework, as has been conducted here, it should also be pointed out again that in terms of practical impact, the pressure prediction is our primary objective because pressure fluctuation is what causes poor performance or even catastrophic situation of fluid machinery.

Though advancements in the pressure prediction capabilities of the present cavitation model has been made in this work, further model development at a conceptual level should be pursued to better address the discrepancies between measurements and computations, especially in the thermal field. Clearly, more experimental investigation is also needed to better quantify the measurement uncertainty, and to offer insight into flow structures.

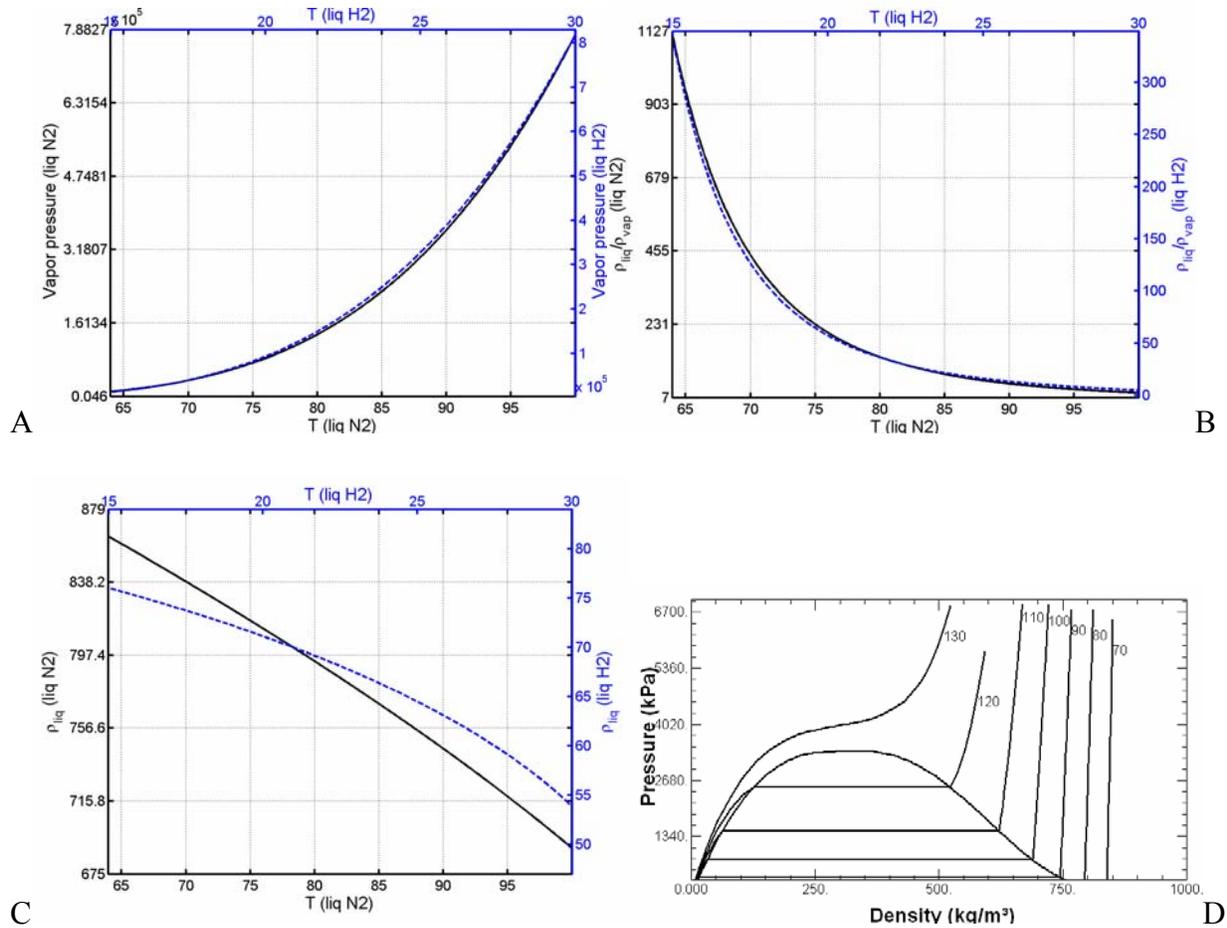


Figure 6-1. Variation of physical properties for liquid nitrogen (solid line, relevant x -axis is on bottom and y -axis is on left), and liquid hydrogen (Lemmon et al., 2002) (dashed line, relevant x -axis is on top and y -axis is on right) with temperature. A) Vapor pressure vs. temperature along saturation line. B) Ratio of liquid density to vapor density vs. temperature along saturation line. C) Liquid density vs. temperature along saturation line. D) Pressure-density chart –lines denote isotherms (liquid N₂).

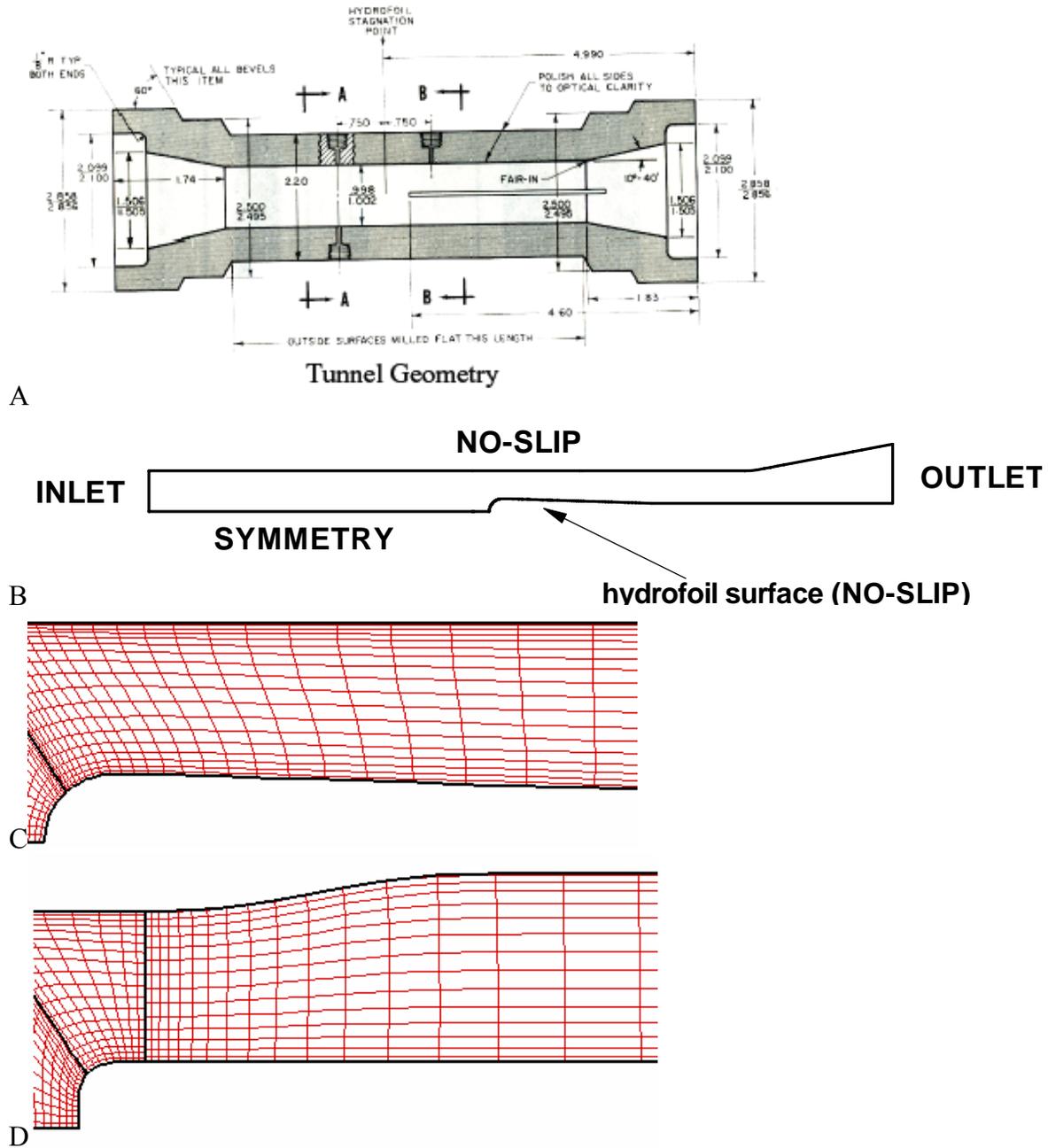


Figure 6-2. Experimental setup and computational geometries. A) Experimental setup used by Hord (1973a-b) to conduct cryogenic cavitation experiments over hydrofoil and ogive geometries. (B) A schematic of the computational setup. (C) The geometry of the adopted hydrofoil. (D) The geometry of the adopted 0.357-inch ogive.

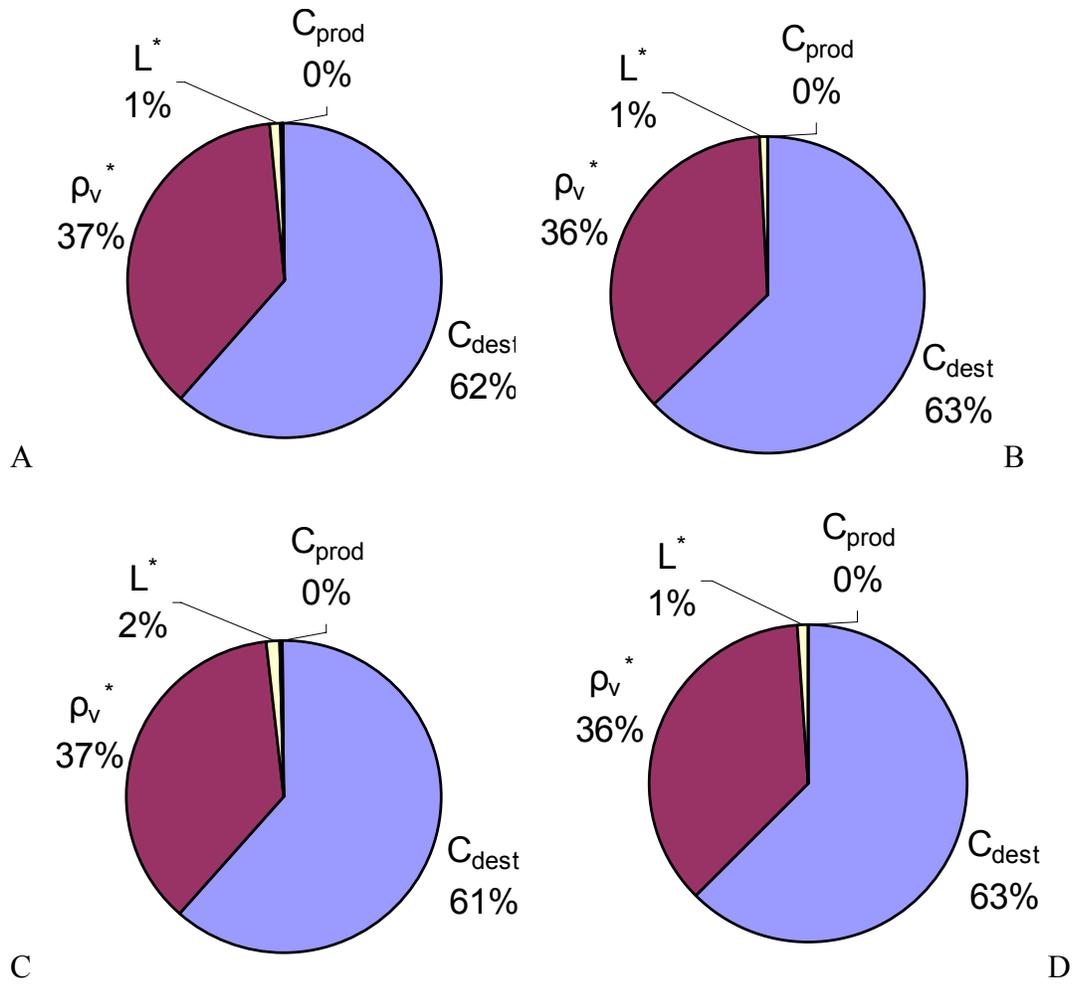


Figure 6-3. Sensitivity indices of main effects using multiple surrogates of prediction metric (liquid N2, Case 290C). A) Polynomial response surface approximation. B) Kriging. C) Radial basis neural network. D) PRESS-based weighted average surrogate.

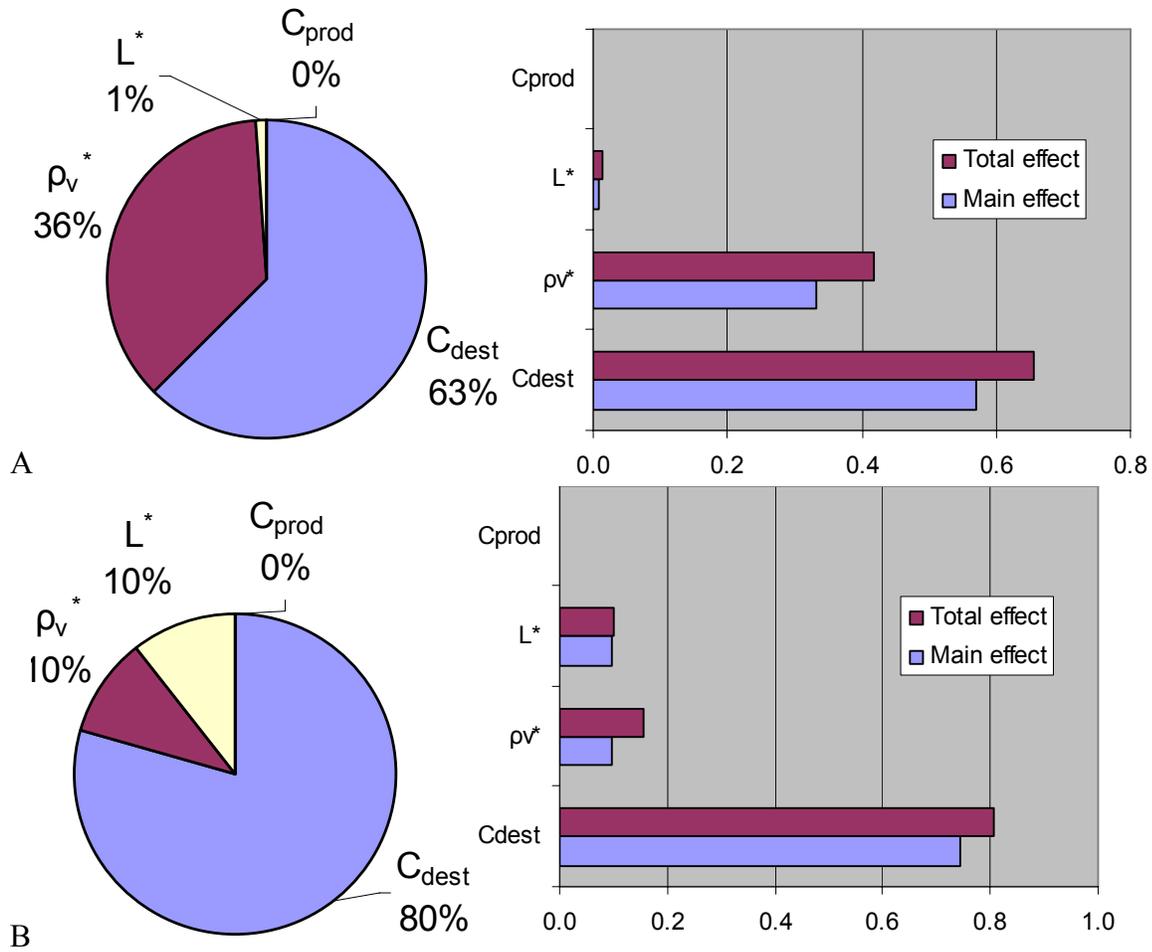


Figure 6-4. Influence of different variables on performance metrics quantified using sensitivity indices of main and total effects. We show results obtained using PWS surrogate (liquid N2, Case 290C). A) Sensitivity indices of main and total effects for P_{diff} . B) Sensitivity indices of main and total effects for T_{diff} .

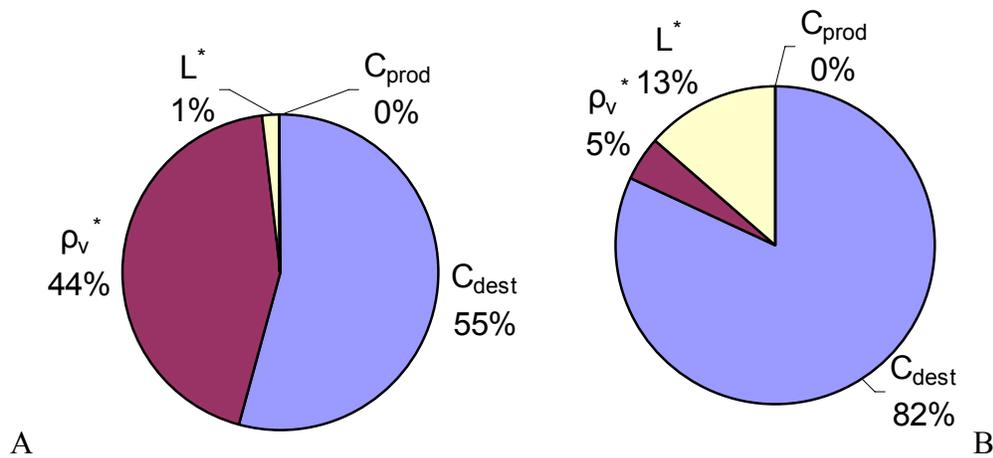


Figure 6-5. Validation of global sensitivity analysis results for main effects of different variables (liquid N2, Case 290C). A) P_{diff} . B) T_{diff} .

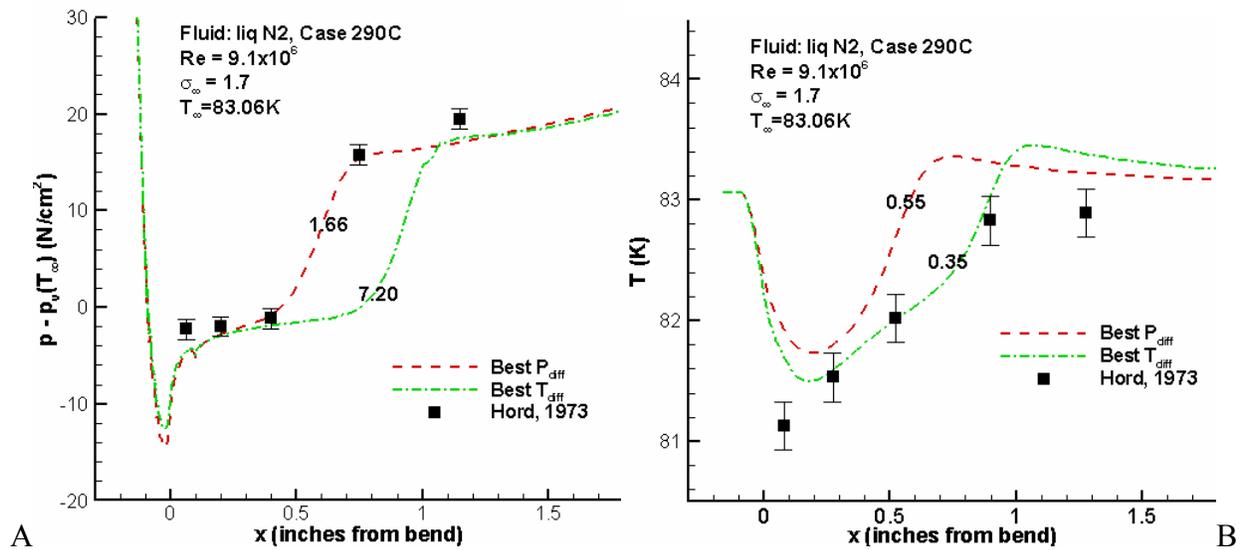


Figure 6-6. Surface pressure and temperature predictions using the model parameters for liquid N2 that minimized P_{diff} and T_{diff} , respectively (Case 290C). The number on each surface pressure or temperature profile represents P_{diff} or T_{diff} value associated with appropriate model parameters. A) Surface pressure. B) Surface temperature.

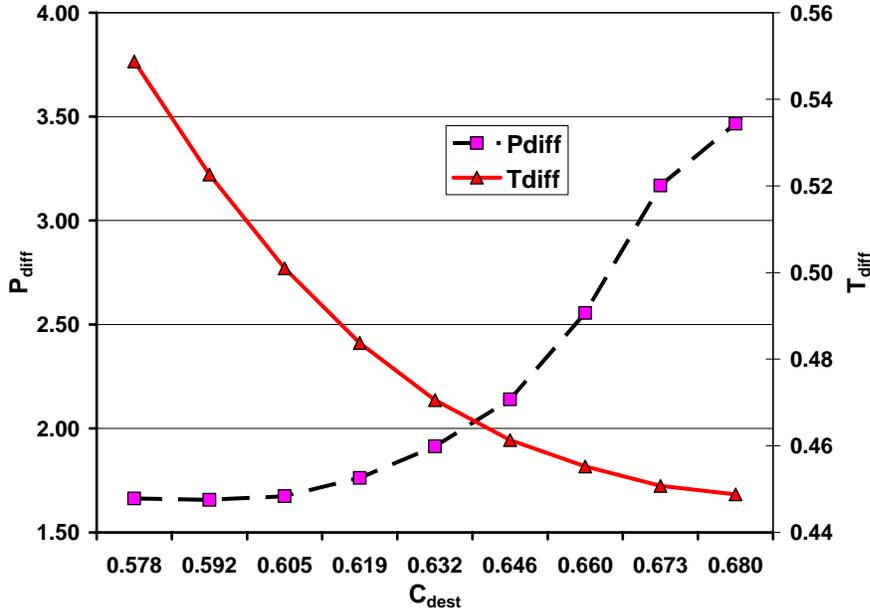


Figure 6-7. Location of points (C_{dest}) and corresponding responses (P_{diff} is shown on the left y-axis, and T_{diff} is shown on the right y-axis) used for calibration of the cryogenic cavitation model (liquid N2, Case 290C).

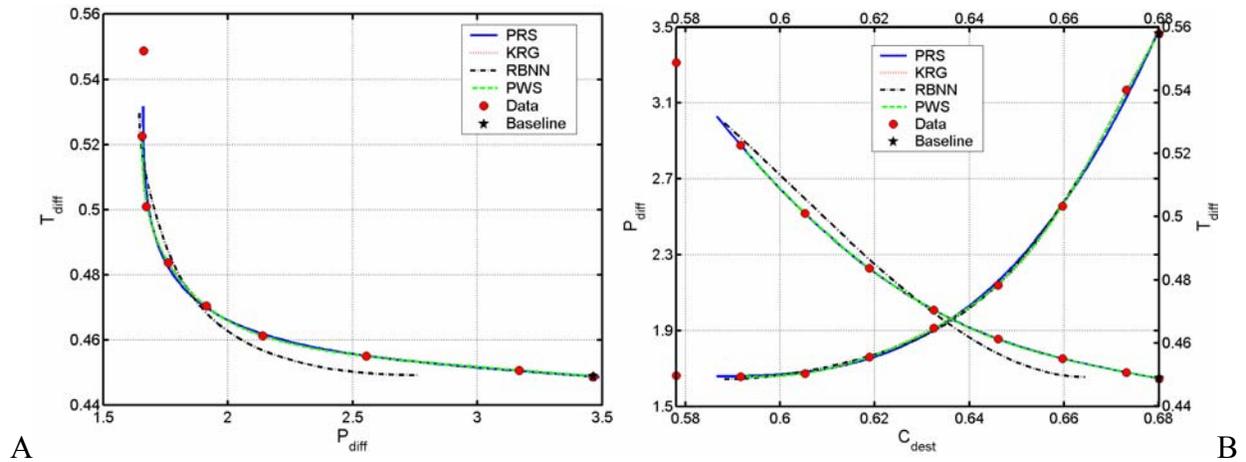


Figure 6-8. Pareto optimal front (POF) and corresponding optimal points for liquid N2 (Case 290C) using different surrogates. PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural network, PWS: PRESS-based weighted average surrogate. A) Function space representation of POF. B) POF in design variable space.

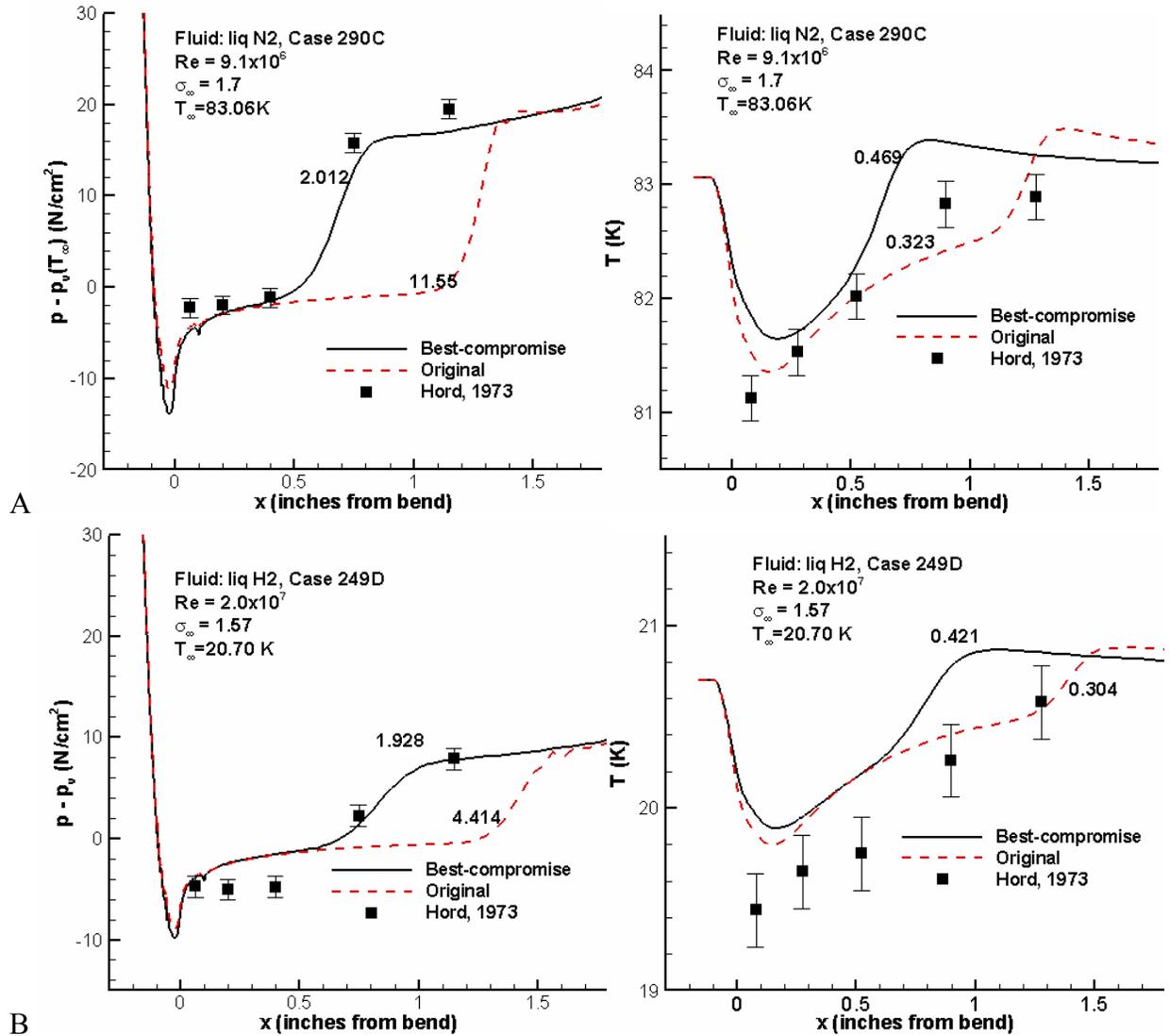
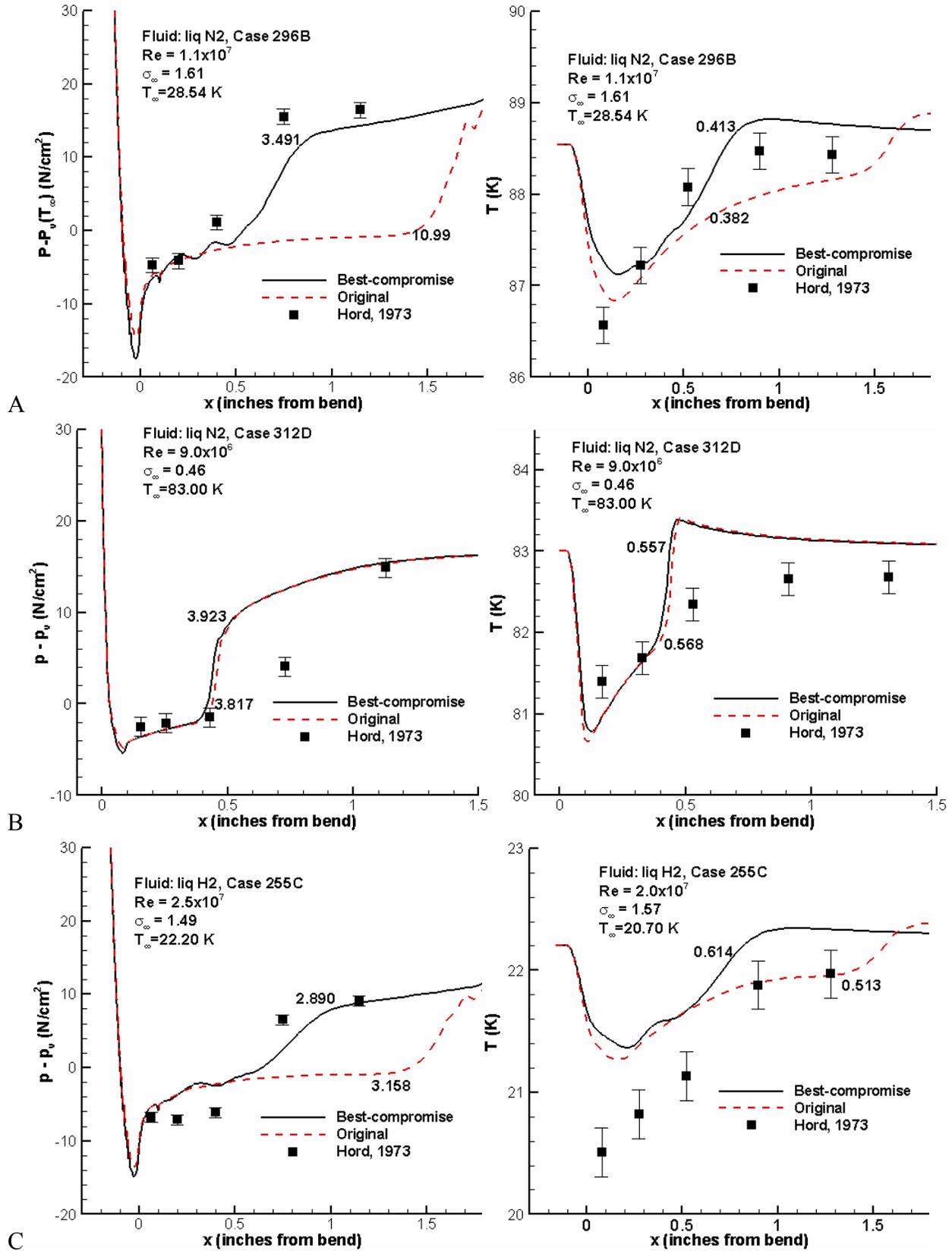
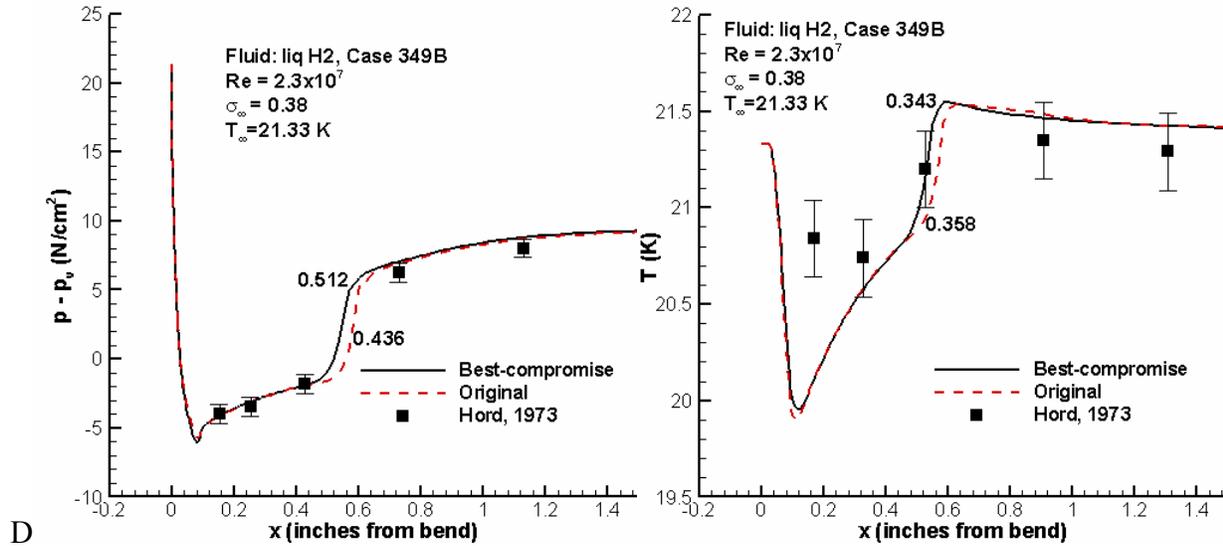


Figure 6-9. Surface pressure and temperature predictions on benchmark test cases using the model parameters corresponding to original and best-compromise values for different fluids. The number on each surface pressure/temperature profile represents P_{diff} or T_{diff} value associated with appropriate model parameters. A) Case 290C, liquid nitrogen, hydrofoil. B) Case 249D, liquid hydrogen, hydrofoil.





D
 Figure 6-10. Surface pressure and temperature predictions using the original parameters ($C_{dest} = 1.0$, $C_{prod} = 80.0$), and best-compromise parameters ($C_{prod} = 54.4$, and $C_{dest,LN2} = 0.6392$; or $C_{dest,LH2} = 0.767$) for a variety of geometries and operating conditions. The number next to each surface pressure or temperature profile represents P_{diff} or T_{diff} value associated with appropriate model parameters. A) Case 296B, liq N2, hydrofoil. B) Case 312D, liq N2, ogive. C) Case 255C, liq H2, hydrofoil. D) Case 349B, liq H2, ogive.

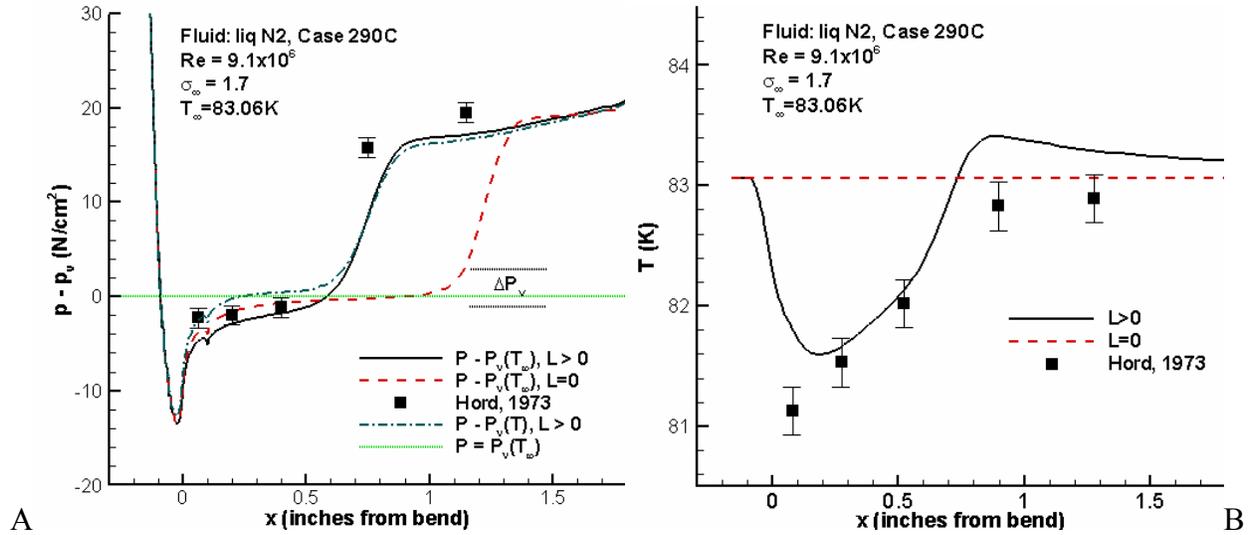


Figure 6-11. Surface pressure and temperature profile on 2D hydrofoil for Case 290C where the cavitation is controlled by, (1) temperature-dependent vapor pressure (designated as $L > 0$), and (2) zero latent heat, and hence, isothermal flow field (designated as $L = 0$). The range indicated by Δp_v shows the level of variations in vapor pressure caused by the temperature variations inside the cavity. (We use best-compromise model parameters $C_{dest, LN_2} = 0.6392$, $C_{prod} = 54.4$ to perform simulations). A) Surface pressure. B) Surface temperature.

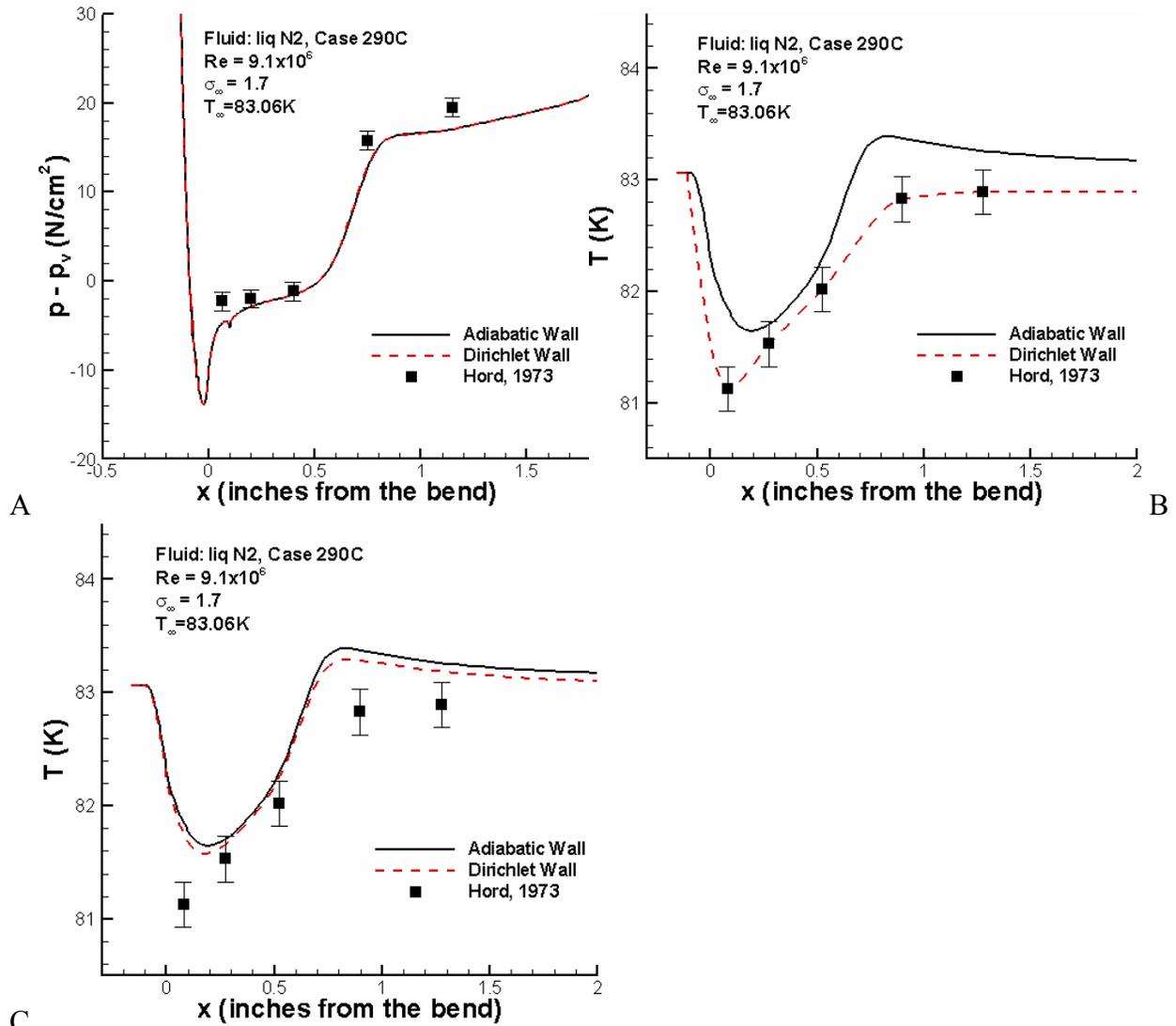


Figure 6-12. Impact of different boundary conditions on surface pressure and temperature profile on 2D hydrofoil (Case 290C, liquid N₂) and predictions on first computational point next to boundary. We use the best-compromise model parameters $C_{dest, LN_2} = 0.6392$, $C_{prod} = 54.4$ for simulations. A) Surface pressure. B) Surface temperature. C) Temperature on the first computational point.

Table 6-1. Summary of a few relevant numerical studies on cryogenic cavitation.

| Reference | Main features |
|-------------------------------|--|
| Reboud et al. (1990) | a) Potential flow equations with semi-empirical formulation |
| Delannoy and Reboud (1993) | b) Simplistic interfacial heat transfer equation (suitable only for sheet cavitation) c) Energy equation not solved |
| Deshpande et al. (1994, 1997) | a) Explicit interface tracking b) Simplistic model for vapor flow inside cavity (suitable only for sheet cavitation) c) Energy equation solved only in the liquid region |
| Lertnuwat et al. (2001) | a) Incorporated energy balance in Rayleigh-Plesset equation to model bubble oscillations b) Good agreement with DNS but deviations under isothermal and adiabatic conditions |
| Tokumasu et al. (2002, 2003) | a) Explicit interface tracking b) Improved model for vapor flow inside the cavity (suitable only for sheet cavitation) c) Energy equation solved only in the liquid region |
| Hosangadi and Ahuja (2003) | a) Solved energy equation in the entire domain with dynamic update of material properties |
| Hosangadi et al. (2003) | b) Some inconsistency with experimental results is noted c) Noticed significant changes in the cavitation model parameters in cryogenic and non-cryogenic conditions |
| Rachid (2003) | a) Theoretical model to account for compressibility effects in a liquid-vapor mixture b) Introduced dissipative effects in phase transformation intermediate between two extreme reversible thermodynamic phenomena |
| Rapposelli & Agostino (2003) | a) Employed thermodynamic relations to extract speed of sound for various liquids b) Captures most features of bubble dynamics well |
| Utturkar et al. (2005a-b) | a) Solved energy equation in the entire domain with dynamic update of material properties b) Test results for different fluids and reference conditions were consistent with the experimental results |

Table 6-2. Ranges of variables for global sensitivity analyses. C_{dest} and C_{prod} are the model parameters associated with the cavitation model source terms for liquid nitrogen, ρ_v^* and L^* are the multiplication factors of vapor density and latent heat obtained from NIST database (Lemmon et al., 2002), respectively.

| Variable | Minimum | Baseline | Maximum |
|------------|---------|----------|---------|
| C_{dest} | 0.578 | 0.68 | 0.782 |
| C_{prod} | 46.24 | 54.4 | 62.56 |
| ρ_v^* | 0.90 | 1.0 | 1.10 |
| L^* | 0.90 | 1.0 | 1.10 |

Table 6-3. Performance indicators and corresponding weights in surrogate approximations of prediction metrics P_{diff} and T_{diff} . PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural network, PWS: PRESS-based weighted surrogate. (PRESS is the square root of predicted residual sum of squares). The test example is Case 290C (liquid nitrogen flow over hydrofoil).

| Surrogate | Parameter | P_{diff} | | T_{diff} | |
|-----------|----------------------|------------|---------|------------|---------|
| | | Values | Weights | Values | Weights |
| | # of training points | 70 | | 70 | |
| | Minimum of data | 1.653 | | 0.334 | |
| | Mean of data | 3.984 | | 0.462 | |
| | Maximum of data | 9.000 | | 0.673 | |
| PRS | # of coefficients | 23 | 0.320 | 19 | 0.335 |
| | R_{adj}^2 | 0.979 | | 0.954 | |
| | PRESS | 0.344 | | 0.0136 | |
| | Maximum error | 0.609 | | 0.0285 | |
| | RMS error | 0.297 | | 0.0121 | |
| KRG | Process variance | 1.277 | 0.603 | 1.67e-3 | 0.597 |
| | PRESS | 0.166 | | 6.92e-3 | |
| RBNN | PRESS | 1.538 | 0.077 | 0.0726 | 0.068 |
| | Maximum error | 0.0905 | | 8.64e-3 | |
| PWS | PRESS | 0.227 | | 9.56e-3 | |
| | Maximum error | 0.199 | | 9.64e-3 | |

Table 6-4. Performance indicators and corresponding weights in surrogate approximations of prediction metrics P_{diff} and T_{diff} in model-parameter space. PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural network, PWS: PRESS-based weighted surrogate. (PRESS is the square root of predicted residual sum of squares). The test example is Case 290C (liquid nitrogen flow over hydrofoil).

| Surrogate | Parameter | P_{diff} | | T_{diff} | |
|-----------|----------------------|------------|---------|------------|---------|
| | | Values | Weights | Values | Weights |
| PRS | # of training points | 9 | | 9 | |
| | Minimum of data | 1.657 | | 0.449 | |
| | Mean of data | 2.222 | | 0.483 | |
| | Maximum of data | 3.465 | | 0.549 | |
| | # of coefficients | 3 | 0.239 | 5 | 0.666 |
| | R_{adj}^2 | 0.999 | | 1.000 | |
| | PRESS | 0.032 | | 1.05e-4 | |
| | Maximum error | 0.037 | | 7.00e-5 | |
| KRG | RMS error | 0.021 | | 6.00e-5 | |
| | Process variance | 0.098 | 0.659 | 1.02e-3 | 0.315 |
| | PRESS | 0.010 | | 3.98e-4 | |
| RBNN | PRESS | 0.077 | 0.101 | 9.00e-3 | 0.019 |
| | Maximum error | 0.018 | | 6.28e-3 | |
| PWS | PRESS | 0.025 | | 3.25e-4 | |
| | Maximum error | 0.011 | | 1.49e-4 | |

Table 6-5. Predicted and actual P_{diff} and T_{diff} at best-compromise model parameter for liquid N2 (Case 290C). PRS: Polynomial response surface approximation. KRG: Kriging, RBNN: Radial basis neural network. PWS: PRESS-based weighted average surrogate.

| C_{dest} | Response | Simulation | PRS | KRG | RBNN | PWS |
|------------|------------|------------|-------|-------|-------|-------|
| 0.6392 | P_{diff} | 2.012 | 2.017 | 2.012 | 2.003 | 2.012 |
| | T_{diff} | 0.466 | 0.466 | 0.465 | 0.463 | 0.465 |

Table 6-6. Description of flow cases chosen for the validation of the calibrated cryogenic cavitation model. σ_∞ is the cavitation number, Re_∞ is the freestream Reynolds number, T_∞ is the inlet temperature, ρ_l is the liquid density, ρ_v is the vapor density, \dot{m}^- and \dot{m}^+ are evaporation and condensation terms in transport-based cavitation model, and C_{dest} is the best-compromise model parameter. ($C_{prod} = 54.4$)

| Fluid | Geometry | Case # | $T_\infty (K)$ | Re_∞ | σ_∞ | $\rho_l / \rho_v \Big _{T_\infty}$ | % change in \dot{m}^- | % change in \dot{m}^+ | C_{dest} |
|---------|-----------|--------|----------------|-------------------|-----------------|------------------------------------|-------------------------|-------------------------|------------|
| Liq. N2 | Hydrofoil | 290C | 83.06 | 9.0×10^6 | 1.70 | 94.90 | -7.58 | 7.53 | 0.6392 |
| Liq. N2 | Hydrofoil | 296B | 88.54 | 1.1×10^7 | 1.61 | 56.25 | -1.00 | 12.34 | 0.6392 |
| Liq. N2 | Ogive | 312D | 83.00 | 9.0×10^6 | 0.46 | 95.47 | 9.12 | 19.18 | 0.6392 |
| Liq. H2 | Hydrofoil | 249D | 20.70 | 2.0×10^7 | 1.57 | 46.97 | -14.79 | 26.57 | 0.767 |
| Liq. H2 | Hydrofoil | 255C | 22.20 | 2.5×10^7 | 1.49 | 31.60 | -8.96 | 29.01 | 0.767 |
| Liq. H2 | Ogive | 349B | 21.33 | 2.3×10^7 | 0.38 | 39.91 | 20.96 | 34.28 | 0.767 |

Influence of Turbulence Modeling on Predictions

We compare the influence of turbulence modeling on the predictions with the help of two benchmark cases of flow over a hydrofoil with liquid nitrogen (Case 290C), and liquid hydrogen (Case 249D). We compare the performance of the standard $k - \varepsilon$ two-equation turbulence model (Launder and Spalding, 1974) with the non-equilibrium $k - \varepsilon$ turbulence model (Shyy et al., 1997). While the governing equations for the two models are the same (Equations (6.19)-(6.22)), the model constants are given in Table 6-7. We use $\beta_1 = 0.9$, and $\beta_2 = 1.15$.

Table 6-7. Model parameters in Launder-Spalding and non-equilibrium $k - \varepsilon$ turbulence models.

| Model | C_μ | C_{ε_1} | C_{ε_2} | σ_k | σ_ε |
|---|---------|--|--|------------|----------------------|
| Standard $k - \varepsilon$ (Launder and Spalding, 1974) | 0.09 | 1.44 | 1.92 | 1.0 | 1.3 |
| Non-equilibrium $k - \varepsilon$ model (Shyy et al., 1997) | 0.09 | $\beta_1 + (1.4 - \beta_1)(P_t / \varepsilon)$ | $\beta_2 + (1.9 - \beta_2)(P_t / \varepsilon)$ | 0.8927 | 1.15 |

The predicted surface pressure and temperature for the two test cases, shown in Figure 6-13, clearly demonstrate only moderate influence of turbulence models on the predictions.

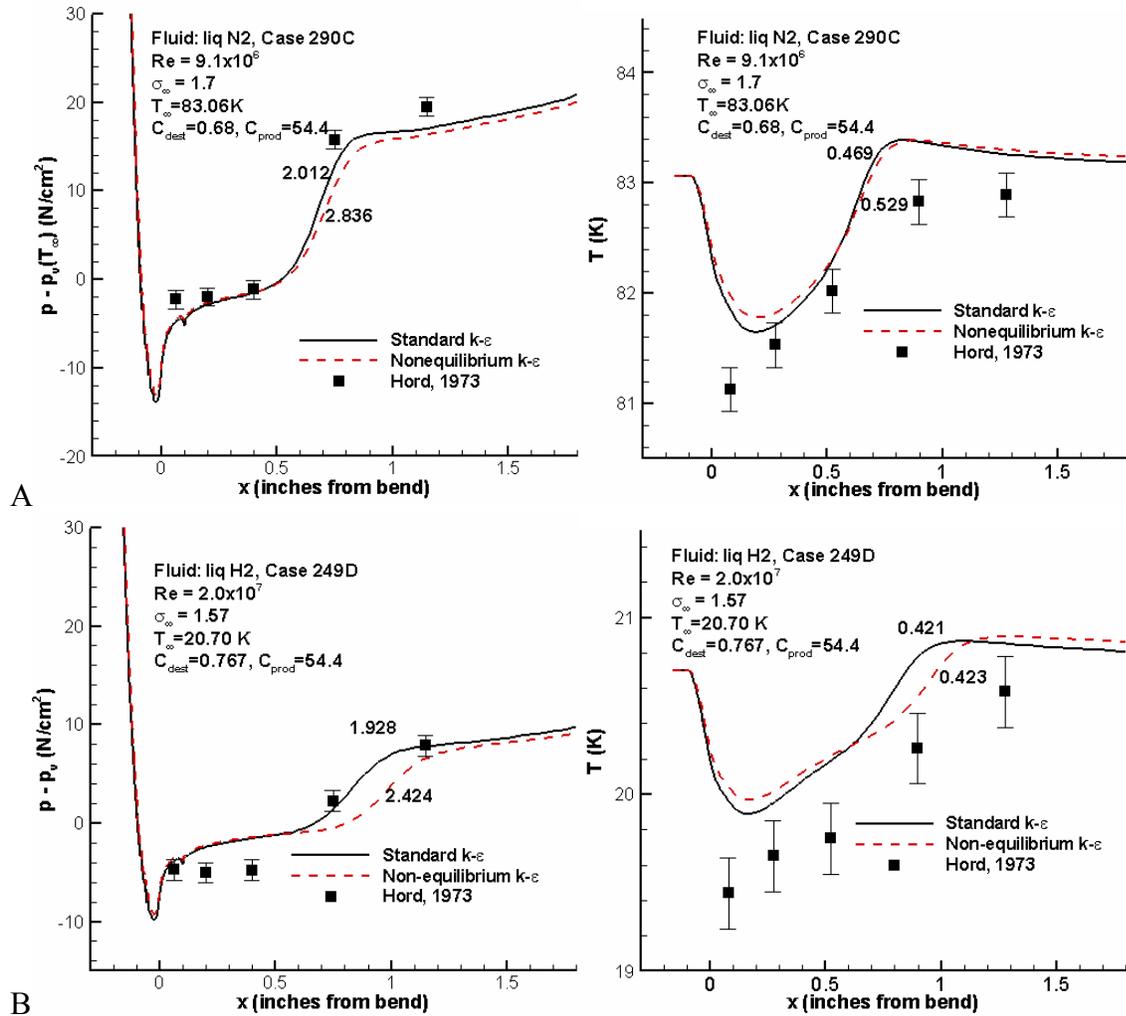


Figure 6-13. Influence of turbulence modeling on surface pressure and temperature predictions in cryogenic cavitating conditions. A) 290C, liquid nitrogen, hydrofoil. B) 249D, liquid hydrogen, hydrofoil.

CHAPTER 7 IMPROVING HYDRODYNAMIC PERFORMANCE OF DIFFUSER VIA SHAPE OPTIMIZATION

Introduction

The space shuttle main engine (SSME) is required to operate over a wide range of flow conditions. This requirement imposes numerous challenges on the design of turbomachinery components. One concept that is being explored is the use of an expander cycle for an upper stage engine. A schematic of a representative expander cycle for a conceptual upper stage engine is shown in Figure 7-1. Oxidizer and fuel pumps are used to feed LOX (liquid oxygen) and LH₂ (liquid hydrogen) to the combustion chamber of the main engine. The combustion products are discharged from the nozzle. The pumps are driven by turbines that use the gasified fuel as the working fluid.

There is a continuing effort to develop subsystems, like turbopumps and turbines, used in a typical expander cycle based upper stage engine. The requirements on the design of subsystems are influenced by size, weight, efficiency, and manufacturability of the system. In addition, often the requirements for subsystems are coupled, for example, the above-mentioned design constraints require turbopumps to operate at high speeds to have high efficiency with low weight and compact design. This may require seeking alternate designs for the turbopumps, as the current designs may not prove adequate over a wide range of operating conditions. Mack et al. (2005b, 2006) optimized the design of a radial turbine that allows high turbopump speeds, performs comparable to an axial turbine at design conditions, and yields good performance at off-design conditions.

Dorney et al. (2006a) have been exploring different concepts for turbopump design. A simplified schematic of a pump is shown in Figure 7-2. Oxidizer or fuel enters from the left. The pressure increases as the fluid passes through the impeller. The fluid emerging from the impeller

periphery typically has high tangential velocity, which is partially converted into pressure by passing it over a diffuser. While conducting the tests with water, Dorney et al. (2006a) have found that a diffuser with vanes is more efficient than a vaneless diffuser at off-design conditions. Over a range of operating conditions, the performance of a pump is driven by the flow in the diffuser. Generally, the diffuser will stall before the impeller, and the performance of the diffuser will drop off more rapidly at off-design operating conditions. The main objectives of the current effort are, (1) to improve the hydrodynamic performance of a diffuser using advanced optimization techniques, and (2) to study the features of the diffuser vanes that influence its performance.

Improvement in performance *via* shape optimization has been successfully achieved in many areas. For example, there are numerous instances of improvement in lift to drag ratio *via* airfoil or wing shape design (Obayashi et al., 2000; Sasaki et al., 2000, 2001; Papila et al., 2002; Emmerich et al., 2002; Huyse et al., 2002; Giannakoglou, 2002), shape of blades is optimized to increase the efficiency of turbines (Papila et al. 2002, Mengistu et al. 2006) and pumps (Samad et al., 2006) in past. Shape design typically involves optimization that requires a significant number of function evaluations to explore different concepts. If the cost of evaluating a single design is high, as is usually the case, surrogate models of objectives and constraints are frequently used to reduce the computational burden.

Surrogate based optimization approach is widely used in the design of space propulsion systems, such as diffuser (Madsen et al. 2000), supersonic turbines (Papila et al. 2002), rocket injectors (Vaidyanathan et al. 2000, Goel et al. 2007a), combustion chambers (Mack et al., 2005a), and radial turbines (Mack et al. 2005b, 2006). Detailed reviews on surrogate based optimization are provided by Li and Padula (2005), and Queipo et al. (2005). There are many

surrogate models but it is not clear which surrogate model performs the best for any particular problem. In such a scenario, one possible approach to account for uncertainties in predictions is to use an ensemble of surrogates (Goel et al., 2006b). This multiple surrogate approach has been demonstrated to work well for several problems including system identification (Goel et al., 2006d), and hardware design (Samad et al., 2006).

Specifically, the objectives of the present study are: (1) to improve the hydrodynamic performance of the diffuser *via* shape optimization of vanes, (2) to identify important regions in diffuser vanes that help optimize pressure ratio, and (3) to demonstrate the application of multiple surrogates based strategy for space propulsion systems.

The chapter is organized as follows. We define the geometry of the vanes and numerical tools to evaluate the diffuser vane shapes in next section. Then we describe the relevant details in surrogate model based optimization and the results obtained for the current optimization problem. The analysis of the physics involved in flow over optimal vane and empirical considerations are discussed afterwards. Finally, we summarize the major findings of this work.

Problem Description

Representative radial locations in the ‘meanline’ pump flow path (not-to-scale) are shown in Figure 7-3 (Dorney et al., 2006a). The fluid enters from the left, and is guided to the unshrouded impeller *via* inlet guide vanes (IGV) assembly. The flow from the impeller passes through the diffuser before being collected in the discharge collector. In this study, we focus our efforts on a configuration of 15 inlet guide vanes, seven main and seven splitter blades, and 17 diffuser vanes. The length of the diffuser vanes is also fixed according to the location of the collector. Our goal is to maximize the performance of the diffuser, characterized by the ratio of pressure at the inlet and the outlet, hereafter called as ‘pressure ratio’. The performance of the diffuser is governed by the shape of the diffuser vane.

Vane Shape Definition

The description of geometry is the most important step in the shape optimization. The current shape of the diffuser vanes, referred as baseline design, (shown in Figure 7-4(A)) has been created using ‘meanline’ and geometry generation codes at NASA (Dorney, 2006b); and yields a pressure-ratio of 1.074. It is obvious from Figure 7-4(B) that the existing shape allows the flow to remain attached while passing on the vanes, but causes significant flow separation while leaving the vane. The separation induces significant loss of pressure recovery that is the primary goal of using a diffuser. The baseline design has been created subject to several constraints, including: (1) the number of vanes is set at 17 based on the number of bolts used to attach the two sides of the experimental rig, (2) the vanes accommodate a 3/8-inch diameter bolt, with enough excess material to allow manufacturing, and (3) the length of the vanes is set based on the location of the collector. These constraints resulted in an initial optimized design that looked quite similar to the baseline design, and yielded similar performance in both isolated component and full stage simulations. In an effort to more thoroughly explore the design space for diffusers (and allowing for future improvements in materials and manufacturing), the constraints resulting from the bolts (the number and thickness of the vanes) were relaxed. Other factors that must be considered in the following optimization include: (1) the designs resulting from the optimization techniques shown in this paper are based on diffuser-alone simulations (different results may be obtained if the optimizations were based on full stage simulations), and (2) no effort has been made to determine if the proposed designs would meet stress and/or manufacturing requirements.

To reduce the separation, we represent the geometry of a vane by five sections using a circular arc, and Bezier curves as shown in Figure 7-5. Sections 1, 3, and 4 are Bezier curves, and Section 2 is a circular arc. The shape of the inlet nose (between points B_1 and B_2) is obtained

from the existing baseline vane shape. The circular arc in Section 2 is described by fixing the radius r ($r = 0.08$), the location of the center C_1 (-3.9, 6.1), and the start and end angles (based on the baseline vane shape provided by Dr Daniel Dorney from NASA). The arc begins at angle $\pi + \theta_r + \theta/2$ (point A_1) and ends at $\pi + \theta_r - \theta/2$ (point A_2). Here, we use $\theta_r = 22.5^\circ$ and θ is a design variable.

A typical parametric Bezier curve $f(x)$, shown in Figure 7-6, is defined with the help of two end points P_0 and P_1 , and two control points P_2 and P_3 as follows.

$$f(x) = P_0(1-x)^3 + 3P_1x(1-x)^2 + 3P_2x^2(1-x) + P_3x^3, x \in [0,1] \quad (7.1)$$

The coordinate of any point on the Bezier curve is obtained by substituting the value of x accordingly. Here, the coordinates of control points are obtained by supplying the slope and the length of the tangents at end points (Papila, 2001) such that, different Bezier curves are generated by varying the length of tangents.

The points A_1 and A_2 , shown by dots in Figure 7-5, and the tangents to the arc, serve as the end points location and tangents used to define the Bezier curves in Sections 1 and 3. Coordinates and the slopes at points B_1 and B_2 (obtained from the data of inlet nose of the baseline design, courtesy Dr Daniel Dorney) are used to define Bezier curves in Sections 1 and 4. The Bezier curve in Section 1 (Figure 7-5) is parameterized using the coordinates and slope of tangents at B_2 and A_2 . The lengths of the tangents t_1 and t_2 control the shape of the curve. While one end point coordinates and slope for Bezier curves in Sections 3 and 4 are known at points A_1 and B_1 , the second end point and slope are obtained by defining the point P (P_y, P_z). The slope of tangent at point P is taken as *five degrees* more than the slope of the line joining points P and B_1 . The additional slope is specified to avoid all the points in Section 4 fall in a straight line. The values of the fixed parameters are decided based on the inputs from designer (Dr Daniel Dorney,

NASA). The lengths of tangents t_1-t_6 serve as variables to generate Bezier curves in Sections 1, 3 and 4.

The ranges of the design variables, summarized in Table 7-1, are selected such that we obtain practically feasible vane geometries and the corresponding grids (discussed in next section). We note that present Bezier curve and circular arc based definition of diffuser vane shapes provides significantly different vane geometries than the baseline design, particularly near the outlet region. The main rationale behind this choice of vane shape is the relaxation of manufacturing and stress constraints, ease of parameterization, and better control of the curvature of the vane that allows the flow to remain attached.

Mesh Generation, Boundary Conditions, and Numerical Simulation

Performance of diffuser vane geometries is analyzed using the NASA PHANTOM code (Dorney and Sondak, 2006b) developed at Marshall Space Flight Center to analyze turbomachinery flows. This 3D, unsteady, Navier-Stokes code utilizes structured, overset O- and H-grids to discretize and to analyze the unsteady flow resulting from the relative motion of rotating components. The code is based on the Generalized Equations Set methodology (Sondak and Dorney, 2003), and implements a modified Baldwin-Lomax turbulence model (Baldwin and Lomax, 1978). The inviscid and viscous fluxes are discretized using a third-order spatially accurate Roe's scheme, and second-order central differencing (Tannehill et al., 1997), respectively. The unsteady terms are modeled using a second order accurate scheme.

For this problem, we solve incompressible, unsteady, non-rotating, turbulent, single phase, constant-material-property flow over diffuser vane. The working fluid is water. By taking the advantage of periodicity, only a single vane is analyzed here. A combination of H- and O-grids with 13065 grid points has been used to analyze diffuser vane shapes. A typical grid is shown in Figure 7-7. The boundary conditions imposed on the flow domain are as follows. Mass flux, total

temperature, and flow angles (circumferential and radial) are specified at the inlet. Mass flux is fixed at the outlet. All solid boundaries are modeled as no slip, adiabatic walls, with zero normal derivative of pressure. Periodic boundary condition is enforced at outer boundaries. With this setup, it takes approximately 15 minutes on a single Intel Xeon processor (2.0 GHz, 1.0 GB RAM) to simulate each design.

Surrogate-Based Design and Optimization

As discussed earlier, the surrogate model based approach is suitable to reduce the computational cost of optimization. The stepwise procedure of surrogate based analysis and optimization is explained with the help of Figure 7-8. Firstly, we identify objectives, constraints, and design variables. Next, we develop a procedure to evaluate different designs. Subsequently, we construct multiple surrogate models of the objectives and constraints. We use these surrogate models to optimize the performance of the system and to characterize the influence of different design variables on the objectives and constraints using global sensitivity analysis (GSA). The optimal design(s) is (are) validated by performing numerical simulation. If we are satisfied with the performance of the subsystem, we terminate the search procedure. Otherwise, we refine the design space in the region of interest. We can also fix the least important variables at the optimal values (as realized by optimization) or mean values to reduce the complexity of the surrogate modeling, and repeat this procedure till convergence. Details of different steps in surrogate model based analysis and optimization method, in context of the current problem, are given as follows.

Surrogate Modeling

The major steps in surrogate modeling are shown in Figure 7-9.

1. **Design of Experiments (DOEs).** The design of experiment is the sampling plan in design variable space and is effective in reducing the computational expense of

generating high-fidelity surrogate models. We use Latin hypercube sampling (LHS), D-optimality criterion, and face-centered cubic central composite designs to select the design sites (points) for conducting simulations.

2. **Numerical Simulations at Selected Locations.** The computationally expensive model is executed at all design points selected using the DOE.
3. **Construction of Surrogate Model.** Surrogate models are relatively computationally inexpensive models to evaluate designs. We use an ensemble of surrogates as proposed by Goel et al. (2006b). The details of different surrogate models are given in Chapter 2.
4. **Model Validation.** The purpose of this step is to establish the predictive capabilities of the surrogate model away from the available data (generalization error). This step can also be used to refine DOE.

Since we have nine design variables, we selected 110 design points (allows adequate data to evaluate 55 coefficients of a quadratic polynomial response surface) using Latin hypercube sampling (LHS) to construct surrogate models. We generated LHS designs using MATLAB routine 'lhsdesign' with 100 iterations for maximization of the minimum distance between points. We evaluated each diffuser vane shape using PHANTOM. This dataset is referred as 'Set A'. The range of the data, given in Table 7-2, indicated potential of improvement in the performance of the diffuser by shape optimization.

We constructed four surrogate models, polynomial response surface approximation (PRS), kriging (KRG), radial basis neural network (RBNN), and a PRESS-based weighted average surrogate (PWS) of the objective 'pressure ratio'. We used quadratic polynomial for PRS, and linear trend model with the Gaussian correlation function for kriging. For RBNN, the 'spread'

coefficient was taken as 0.5 and the error ‘goal’ was the square of five-percent of the mean value of the response at data points. The parameters, α and β , for PWS model (refer to Chapter 4) were 0.05 and -1, respectively. The summary of quality indicators for different surrogate models is given in Table 7-2. All error indicators are desired to be low compared to the response data, except R_{adj}^2 , which is desired to be close to one. The PRESS (Chapter 2) and RMS error ($\sim 1.0e-2$) were very high compared to the range of data. This indicated that all surrogate models poorly approximated the actual response, and were likely to yield inaccurate results if used for global sensitivity analysis and optimization. To identify the cause of poor surrogate modeling, we conducted a lack-of-fit test (refer to Appendix D) for PRS. A low p-value (~ 0.017) indicated that the chosen order of the polynomial was inadequate in the selected design space. Since, the data available at 110 points is insufficient to estimate 220 coefficients in a cubic polynomial, this issue of model inadequacy also reflected the lack of data.

We addressed the issue of model accuracy or data inadequacy using two parallel approaches. Firstly, we added more data in the design space to improve the quality of fit. We sampled 330 additional points using Latin hypercube sampling, such that we had 440 design points to fit a cubic polynomial (220 coefficients). We call this dataset as ‘Set B’. Secondly, the low mean value of the response data (1.041) at 110 points compared to the baseline design (1.074) indicated that large portion of the current design space was undesirable due to inferior performance. Hence, it might be appropriate to identify the region where we expect improvements in the performance of the designs, and construct surrogate models by sampling additional design points in that region (reasonable design space approach, Balabanov et al., 1999). To identify the region of interest, we used the surrogate models, constructed with Set A data (110 design points), to evaluate response at a large number of points in design space.

Specifically, we evaluated responses at a grid of four Gaussian points in each direction (total $4^9=262,144$ points). We chose the Gaussian points, instead of usual uniform grids, because the Gaussian points lie inside the design domain, and are less susceptible to extrapolation errors than the corners of uniform grids that might fall outside the convex hull of LHS design points used to construct surrogate models. Any point with a predicted performance (due to any surrogate model) of 1.080 units or better (0.5% improvement over the baseline design) was considered to belong to the potential region of interest. This process identified 29,681 unique points (~11%) in the potential good region. We selected 110 points from this data set using D-optimality criterion in this smaller region. D-optimal designs were generated using MATLAB routine ‘candexch’ with a maximum of 100 iterations to maximize D-efficiency (Myers and Montgomery, 1995, pp. 93). This 110 points dataset is called ‘Set C’.

As before, we conducted simulations at data points in the Sets B and C using PHANTOM. One point in each set failed to provide an appropriate mesh. The mean, minimum, and maximum values of the pressure ratio for the two datasets are summarized in Table 7-3. We observed only minor differences in the mean pressure-ratio of the Set B compared to the Set A (Table 7-2), but the responses in the Set C data indicated high potential of improvement. This clearly demonstrates the effectiveness of the reasonable design space approach used to identify the region of interest.

We approximated the data in the Set B, and the Set C, using four surrogates. We employed a reduced cubic, and a reduced quadratic polynomial for PRS approximation of the Set B and the Set C data, respectively. As can be seen from different error measures in Table 7-3, the quality of surrogate models fitted to the Set B, and the Set C data was significantly better than the surrogate models fitted to the Set A data. This improvement in surrogate approximation was attributed to

the increase in the sampling density (Set B) allowing a cubic model, and the reduction of the design space (Set C). Both PRS models did not fail the lack-of-fit test (p-value $\sim 0.90+$) indicating the adequacy of fitting surrogate models in the respective design spaces. The PRESS metric and weights associated with different surrogates suggested that the PRS was the best surrogate model for the Set B and the Set C data, unlike kriging for the Set A data.

We used the surrogate models fit to the Set B data for global sensitivity analysis and the surrogate models fitted to the Set C data for optimization. The optimization of the performance using the surrogate models fitted to the Set B data was inferior to the optimal design obtained using the Set C data based surrogates. Some of the optimal designs from the Set B data based surrogates could not be analyzed. This anomaly arose because large design space was sampled with limited data, such that large regions remained unsampled; and hence, susceptible to significant errors, particularly near the corners where optima were found. The same issue restricted the use of surrogate models constructed using Set C data for global sensitivity analysis as there were large extrapolation errors outside the region of interest where no point was sampled. Hence, surrogate models constructed using Set B data were more suited for conducting global sensitivity analysis.

Global Sensitivity Assessment

Global sensitivity analysis (GSA) is useful to characterize the importance of different design variables. This information can be used to identify the most, and the least important design variables. We fix the least important design variables to reduce the complexity of the problem. Here, we use a variance based, non-parametric approach to perform global sensitivity analysis. In this approach, the response function is decomposed into unique additive functions of variables, and their interactions, such that the mean of each additive function is zero. This decomposition allows the variance (V) to be computed as a sum of individual partial variance of

each variable (V_i), and partial variance of interactions (V_{ij}) of different variables. The sensitivity of the response function with respect to each variable is assessed by comparing the sensitivity indices (S_i, S_{ij}) that is the relative magnitude of partial and total variance of each variable. Using Sobol's (1993) approach of decomposing the variables into two groups; first group with a single variable ' i ', and second group Z with all variables, except i^{th} variable, we compute the main (S_i) and total effect (S_i^{total}) of sensitivity indices as follows.

$$S_i = V_i / V, \quad S_i^{total} = (V_i + V_{iZ}) / V \quad (7.2)$$

In the above equation, V_{iZ} denotes the partial variance of all interactions of i^{th} variable. A detailed description of the global sensitivity analysis approach is given in Appendix C.

We used Gauss quadrature numerical integration scheme with four Gauss points along each direction (total $4^9 = 262,144$ points) to evaluate different integrals in GSA. The response at each point was evaluated using surrogate models fit to the Set B data. Corresponding sensitivity indices of main effects of different variables are shown in Figure 7-10. Although there were differences in the exact magnitude of sensitivity indices from various surrogates, all surrogates indicated that the pressure ratio was most influenced by three variables, P_z , t_2 , and P_y . A comparison of sensitivity indices of total and main effect of design variables (using PWS) in Figure 7-11 suggested that the interactions between variables are small but non-trivial.

To validate the findings of the global sensitivity analysis, we evaluated the variation in the response function (pressure ratio) by varying one variable at a time, while keeping the remaining variables at the mean values. We specified five equi-spaced levels for each design variable, and used trapezoidal rule to compute the actual variance. The responses at design points were evaluated by performing actual numerical simulations. The results of actual variance computations are shown in Figure 7-12.

The one-dimensional variance computation results also indicated that variables, P_z , t_2 , and P_y , were more important than all other variables. This validated the findings of the global sensitivity analysis. The differences in the results of one-dimensional variance computation and global sensitivity analysis can be explained as follows: (1) the number of points used to compute one-dimensional variance is small, (2) one-dimensional variance computation does not account for interactions between variables, and (3) there are approximation errors in using surrogate models for global sensitivity analysis. Nevertheless, the main implications of the finding clearly show that the performance of the diffuser vane was pre-dominantly affected by the location of point P and the length of tangent t_2 in Figure 7-5.

Optimization of Diffuser Vane Performance

Next, we used surrogate models, fit to the Set C data, to maximize the pressure ratio by exploring different diffuser vane shapes. To avoid the danger of large extrapolation errors in the unsampled region, we employed the surrogate model (PRS surrogate) from the Set A as a constraint (all points in the feasible region should have predicted response greater than a threshold value of 1.080). Our use of PRS from the Set A as the constraint was motivated by the simplicity of the constraint function, and the fact that PRS contributed to the most number of points in the potential region of interest.

We used sequential quadratic programming optimizer to find the optimal shapes. The optimal configuration of blade shapes, obtained using different surrogate models as function evaluators, is shown in Figure 7-13, and the corresponding optimal design variables are given in Table 7-4. The optimal designs obtained from all surrogates were close to each other in both function and design space. A few minor differences were observed in relatively insignificant design variables (refer to the results of global sensitivity analysis). Notably, all design variables touched the bounds for PRS, and were close to the corner for other surrogate models. The small

value of θ indicated sharper nose, as was used in the baseline design. Also, most tangents were at their lower bounds that resulted in low curvature sections. Near the point P, the tangents were on their upper limits to facilitate gradual transition in the slope. The optimal vane was thinner in the middle section, and was longer compared to the baseline design (Figure 7-13). The central region of the optimal design was non-convex compared to the convex section for the baseline design.

We simulated all four candidate optimal designs from different surrogate models to evaluate the improvements. The actual and predicted performances from different surrogates are compared in Table 7-5. We observed that the error in approximation for different surrogates was comparable to their respective PRESS errors. Nevertheless, PRS was the most accurate surrogate model and furnished the best performance shape. RBNN was the worst surrogate model. PRESS-based weighted average surrogate model performed significantly better than the worst surrogate. The best predicted diffuser vane yielded significant improvements in the performance (1.117) compared to the baseline design (1.074). We refer to this design as ‘intermediate optimal’ design.

It is obvious, that the design space refinement in the region of interest based on multiple surrogates has pay-offs in the improved performance of surrogates, and the identification of optimal design. The high confidence in the optimal predictions was also derived from the similar performance of all surrogate models. The results also showed the incentives (protection against the worst design, proper identification of the reasonable design space) of investing a small amount of computational resources in constructing multiple surrogate models (less than the cost of a single simulation) for computationally expensive problems, and then the extra cost of evaluating multiple optima.

We compared instantaneous (at the end of simulation) and time-averaged flow fields from the intermediate optimal design (from PRS) with the baseline design in Figure 7-14. The intermediate optimal design allowed smoother turning of the flow compared to the baseline design, and reduced the losses due to separation of the flow, which were significantly high in the baseline design. Consequently, the pressure at the outlet was higher for this intermediate optimal design. We also noted an increase in pressure loading on the vane for the intermediate optimal design.

Design Space Refinement – Dimensionality Reduction

We used first design space refinement by identifying the region of interest. This helped in identifying the intermediate optimal design. Since most design variables in the intermediate optimal design were at the boundary of the design space (Table 7-4), further improvements in the performance of the diffuser might be obtained by relaxing the limits of the design variables. To reduce the computational expense, we reduce the design space by utilizing the findings of the global sensitivity analysis. We fixed six relatively insignificant design variables at the optimal design (predicted using PRS), and expanded the range of the three most important design variables. The modified ranges of the design variables and the fixed parameters are given in Table 7-6. We selected 20 design points using a combination of face-centered central composite design (15 points FCCD), and LHS designs (five points) of experiments. The range of pressure ratio at 20 design points is given in Table 7-7. Note that, all tested designs in the refined design space performed equal to or better than the intermediate optimal design.

As before, we constructed four surrogate models in the refined design space. The performance metrics, specified in Table 7-7, indicated that all surrogate models approximated the response function very well. The weights associated with different surrogate models suggested that a quadratic PRS approximation represents the data the best. This result is not unexpected

since any smooth function can be represented by a second order polynomial, if the domain of application is small enough. As before, PWS model was comparable to the best surrogate model.

Final Optimization

The design variables for the four optimal designs of the diffuser vane obtained using different surrogate models and corresponding surrogate-predicted, and actual (CFD simulation) pressure ratio are listed in Table 7-8. The error in predictions of surrogate models compared well with the quality indicators, and all surrogate models had only minor differences in the performance. Also, the optimal vane shapes from different surrogates were similar. In this case also, polynomial response surface approximation conceded the smallest errors in prediction. While the performance of the optimal diffuser vane has improved compared to the intermediate optimal design (compare to Table 7-5), the contribution of the optimization process was insignificant. One of the data points resulted in a better performance than the predicted optimal. This result was not surprising because the optimal design existed at a corner that was already sampled leaving little scope for further improvement. Nevertheless, the optimizers correctly concentrated on the best region. As expected, the optimized design was thinner and streamlined to further reduce the losses, and to improve the pressure recovery.

Analysis of Optimal Diffuser Vane Shape

We analyzed the optimal diffuser vane shape ($t_2 = 0.60$, $P_y = -2.00$, $P_z = 6.00$; P -ratio = 1.151) according to the flow structure and the other considerations as follows.

Flow Structure

The instantaneous and time-averaged pressure contours for the optimal vane shape (best data point) are shown in Figure 7-15. We observed further reduction in separation losses and smoother turning of the flow compared to the intermediate optimal design obtained before design space refinement (Table 7-5). Consequently, the pressure rise in the diffuser was higher. The

optimal vane shape had a notable curvature in the middle section on the lower side of the vane (near the point P). This curvature decelerated the flow, and led to faster increase in pressure (notice the shift of higher pressure region towards the inlet in Figure 7-14 and Figure 7-15).

Vane Loadings

The shapes and pressure loads on the baseline, intermediate optimal, and final optimal vanes are shown in Figure 7-16. We noted the increase in the mean pressure on the diffuser vane for the optimal design. The pressure loads near the inlet tip, and the pressure loading on the diffuser vane, given by the area bound by the pressure profile on the two sides, had reduced by optimization. However, the optimized diffuser vane might be susceptible to high stresses as the optimal design was thinner compared to the baseline vane. The intermediate optimal design served as a compromise design with relatively higher pressure ratio (~4%) compared to the baseline design, and lower pressure loading on the vanes compared to the optimal design.

In future, this problem would be studied by accounting for manufacturing and structural considerations, like stresses in the diffuser vanes. One can either specify a constraint to limit stress to be less than the feasible value or alternatively, one can solve a multi-objective optimization problem with two competing objectives, minimization of stress or pressure loading in the vane, and maximization of pressure ratio.

Empirical Considerations

Typically, the vane shape design is carried out using empirical considerations on the gaps between adjacent diffuser vanes as shown in Figure 7-17. The empirical suggestions on the ratio of different gaps (Stepanoff, 1993), and actual values obtained for different vanes are given in Table 7-9. Contrary to the empirical relations, the ratio of length to width gap (L/W_1), and ratio of width gaps (W_2/W_1) decreased as the pressure ratio increases, though the actual magnitude of length and width of the gaps increase. The discrepancies between the optimal design and the

empirical optimal ratios (Stepanoff, 1993) are explained by the multiple design considerations used for the empirical optimal. We note that the optimization was carried out for a 2D diffuser vane not for the combination of vanes and flow for which the empirical ratios are provided. This allowed a variable height of the vane for optimization. However, the empirical ratios are obtained by assuming a constant channel height, so that the area ratio of the channel reduces to the ratio of width gaps (W_2/W_1). Nevertheless, this requires further investigation to understand the cause of discrepancies between the empirical ratios and that obtained for the optimal design.

Summary and Conclusions

We used surrogate model based optimization strategy to maximize the hydrodynamic performance of a diffuser, characterized by the increase in pressure ratio, by modifying the shape of the vanes. The shape of the diffuser vanes was defined by a combination of Bezier curves and a circular arc. We defined the shape of the vane using nine design variables and used surrogate models to represent the pressure ratio. We used lack-of-fit test to identify the issues of model inadequacy, and insufficiency of the data to represent the pressure ratio. We addressed these issues by, (1) adding more data points, and (2) identifying the region of interest using the less-accurate surrogate models. More samples were added in the region of interest using the information from multiple surrogate models.

The surrogate models, constructed with increased data and/or in smaller design space, were significantly more accurate than the initial surrogate models. Also, during the course of design space refinement, the best surrogate model changed from kriging (initial data) to polynomial response surface approximation (all subsequent results). Had we followed the conventional approach of identifying the best surrogate model with the first design of experiments, and then used that surrogate model for optimization, we might have not captured the best design. Thus, we can say that the results reflect the improvements in the performance using the design space

refinement approach, and using multiple surrogates constructed by incurring a low computational cost.

We conducted a surrogate model based sensitivity analysis to identify the most important design variables in the entire design space. Three design variables controlling the shape of the upper and lower side of the vane were found to be most influential. We used surrogate model in the reduced design space to identify the optimal design in nine variable design space. This intermediate optimal design improved the pressure ratio by more than four percent compared to the baseline design.

Since all the design variables for intermediate optimal design hit the bounds, we further refined the design space by fixing the least important variables on optimal values to reduce the design space, and relaxing the bounds on the most important design variables. The optimal design obtained using the surrogate models in the refined design space further improved the performance of the diffuser by more than seven percent compared to the baseline design. The pressure losses in the flow were reduced, and a more uniform pressure increase on the vane was obtained. However, the optimal vane shape might be susceptible to failure due to high stresses. This behavior was attributed to the absence of stress constraint that allowed using thinner vanes to maximize the performance. In the future, the optimization would be carried out by considering the multi-disciplinary analysis accounting for stress constraint, manufacturability, and pressure increase.

In terms of the vane shapes, as expected, thin vanes helped improve the hydrodynamic performance significantly. The interesting aspect was the change in the sign of curvature of the vane on the suction side that allows an initial speeding of the flow followed by a continuous pressure recovery without flow separation.

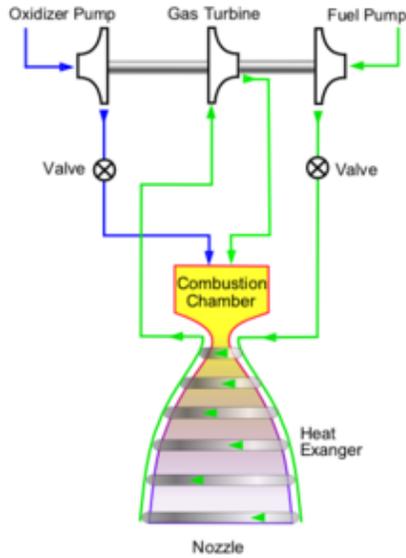


Figure 7-1. A representative expander cycle used in the upper stage engine (courtesy: wikipedia entry on expander cycles).

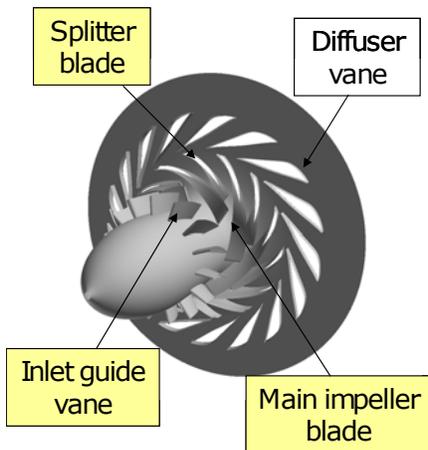


Figure 7-2. Schematic of a pump.

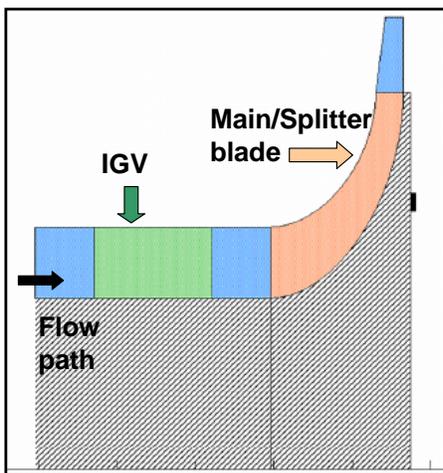


Figure 7-3. Meanline pump flow path. IGV is inlet guide vane.

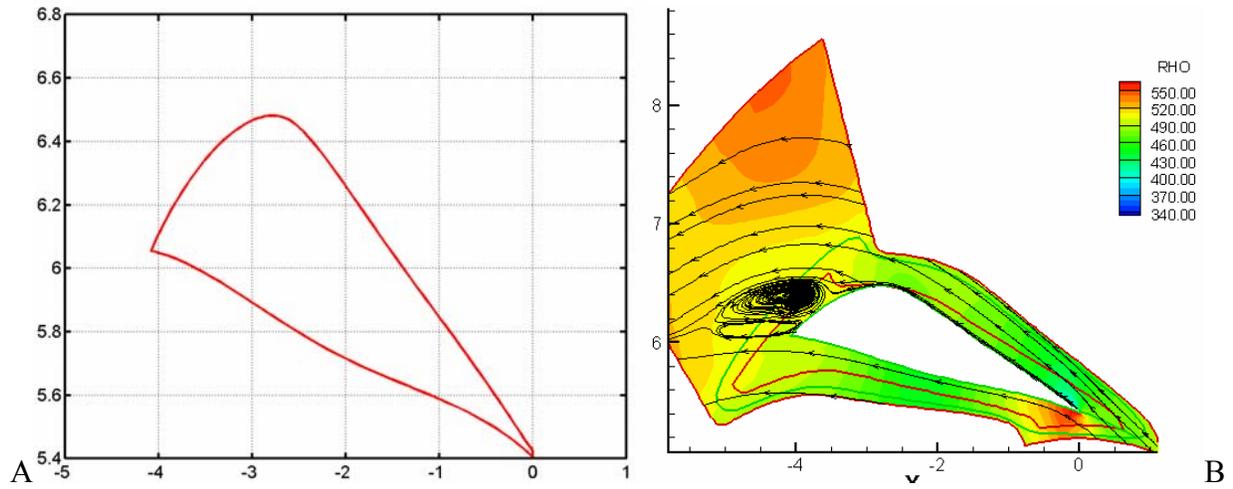


Figure 7-4. Baseline diffuser vane shape and time-averaged flow. A) Diffuser vane shape. B) Streamlines and time-averaged pressure.

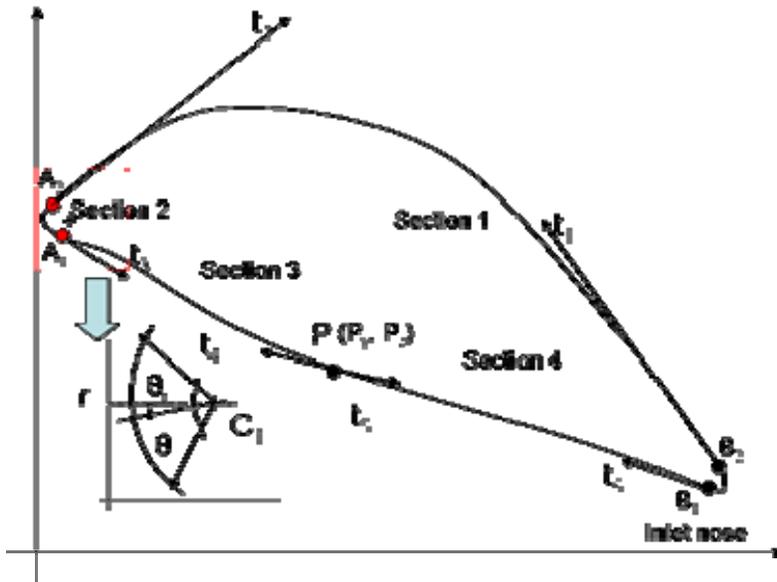


Figure 7-5. Definition of the geometry of the diffuser vane (refer to Table 7-1 for variable description).

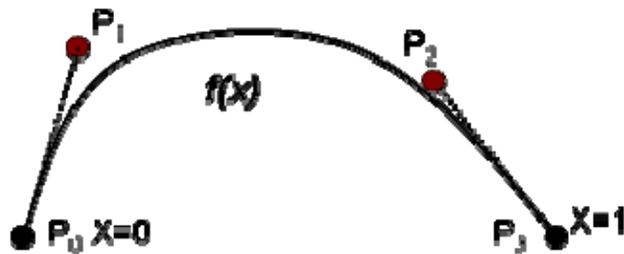


Figure 7-6. Parametric Bezier curve.

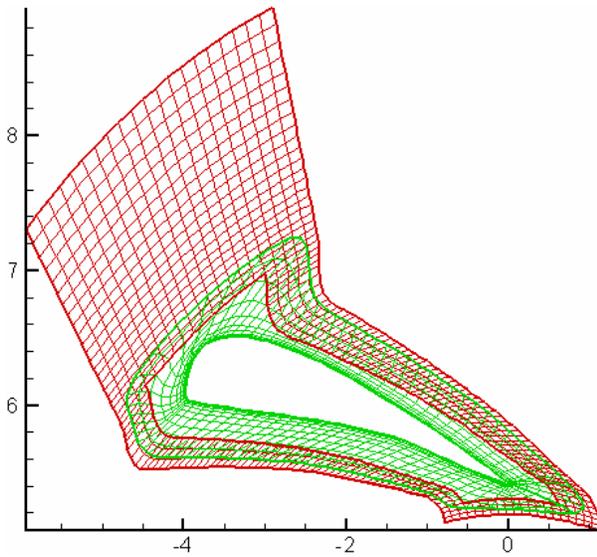


Figure 7-7. A combination of H- and O-grids to analyze diffuser vane. Body-fitted O-grids are shown in green and algebraic H-grid is shown in red.

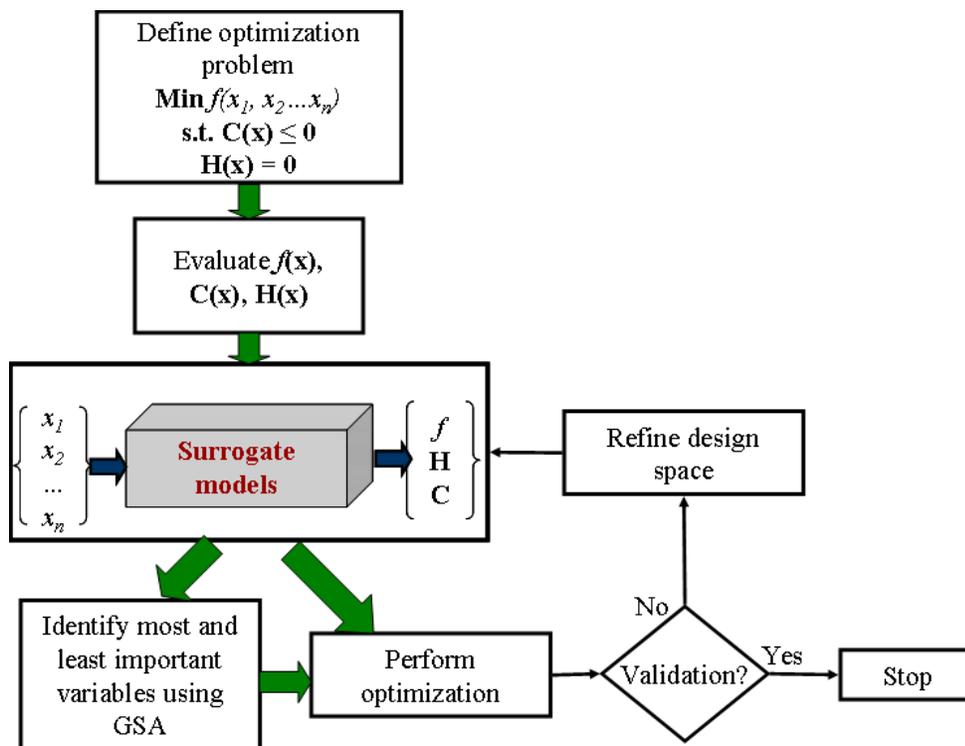


Figure 7-8. Surrogate based design and optimization procedure.

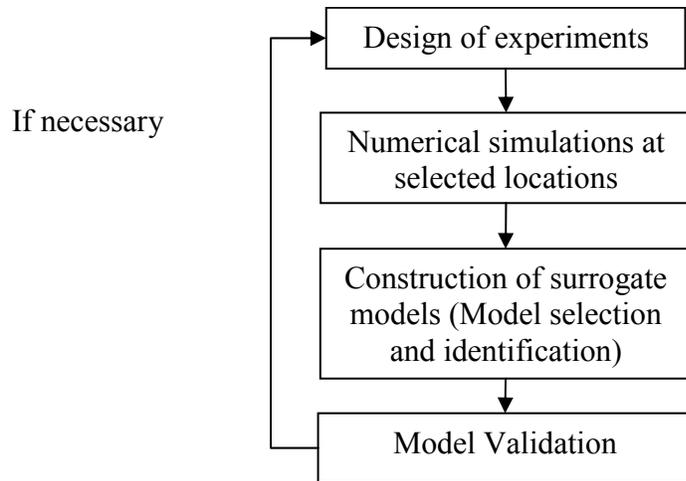


Figure 7-9. Surrogate modeling.

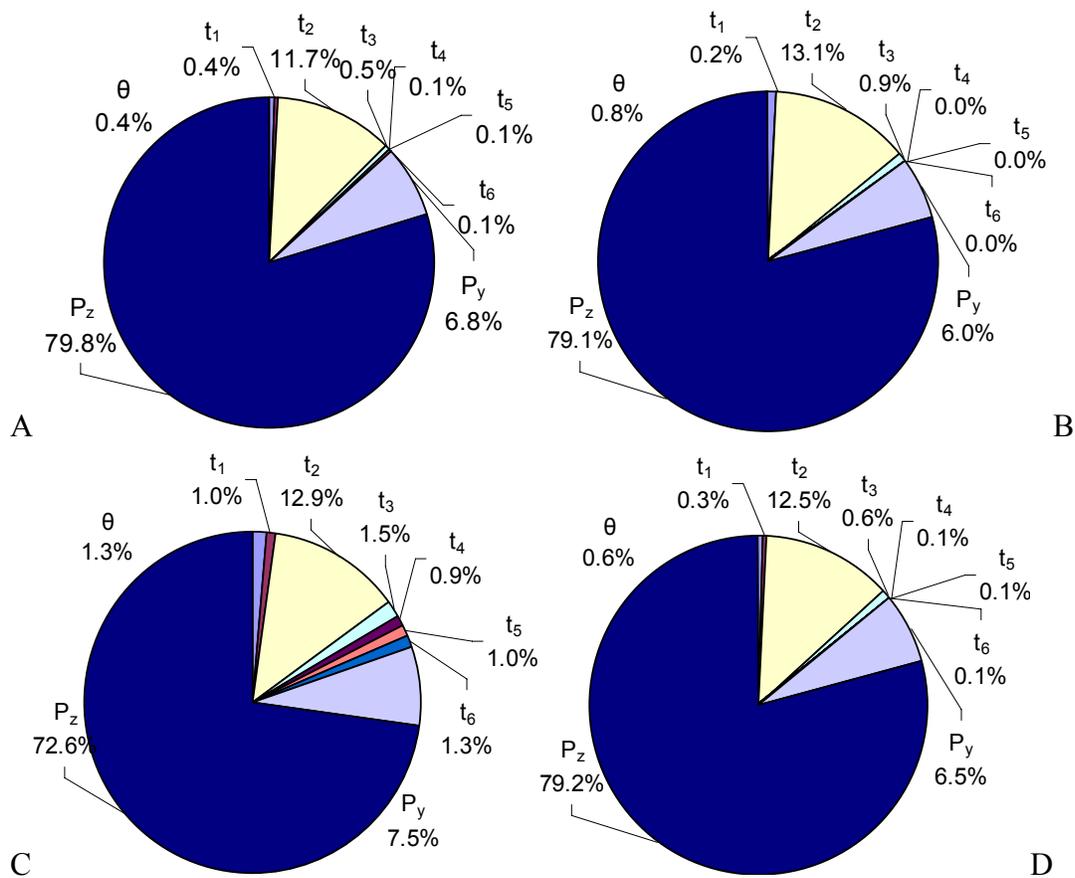


Figure 7-10. Sensitivity indices of main effect using various surrogate models (Set B). A) Polynomial response surface approximation. B) Kriging. C) Radial basis neural network. D) PRESS-based weighted average surrogate model.

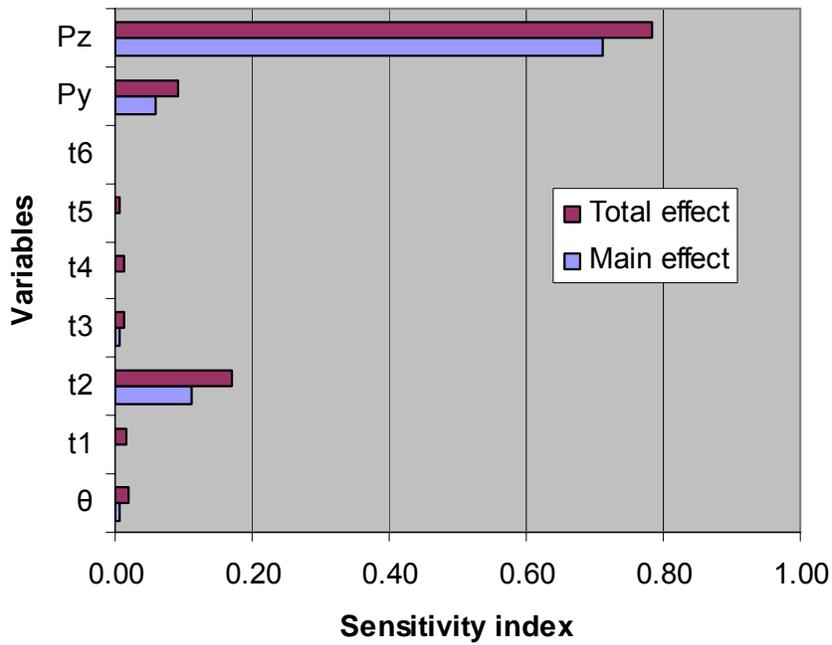


Figure 7-11. Sensitivity indices of main and total effects of different variables using PWS (Set B).

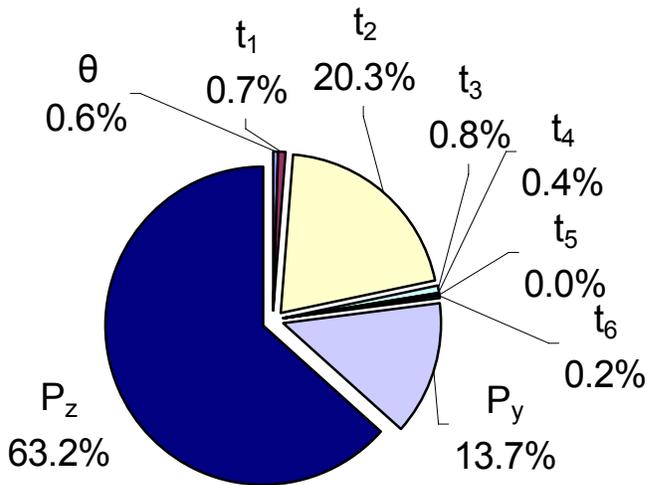


Figure 7-12. Actual partial variance of different design variables (no interactions are considered).

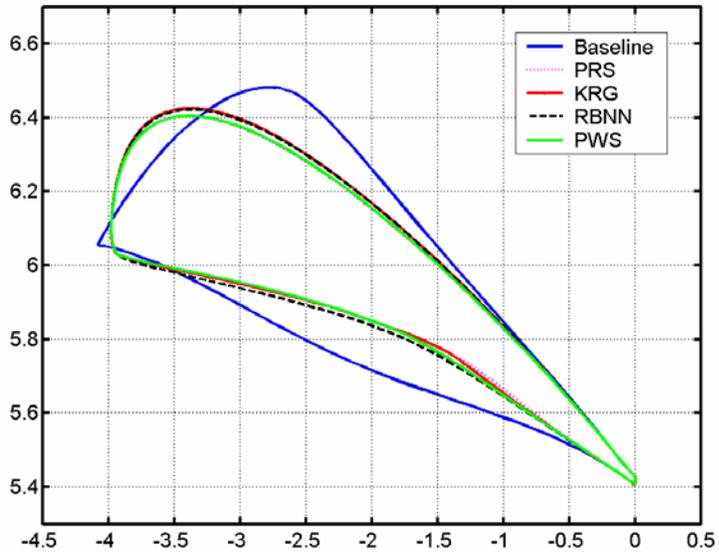


Figure 7-13. Baseline and optimal diffuser vane shape obtained using different surrogate models. PRS indicates function evaluator is polynomial response surface, KRG stands for kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate model as function evaluator.

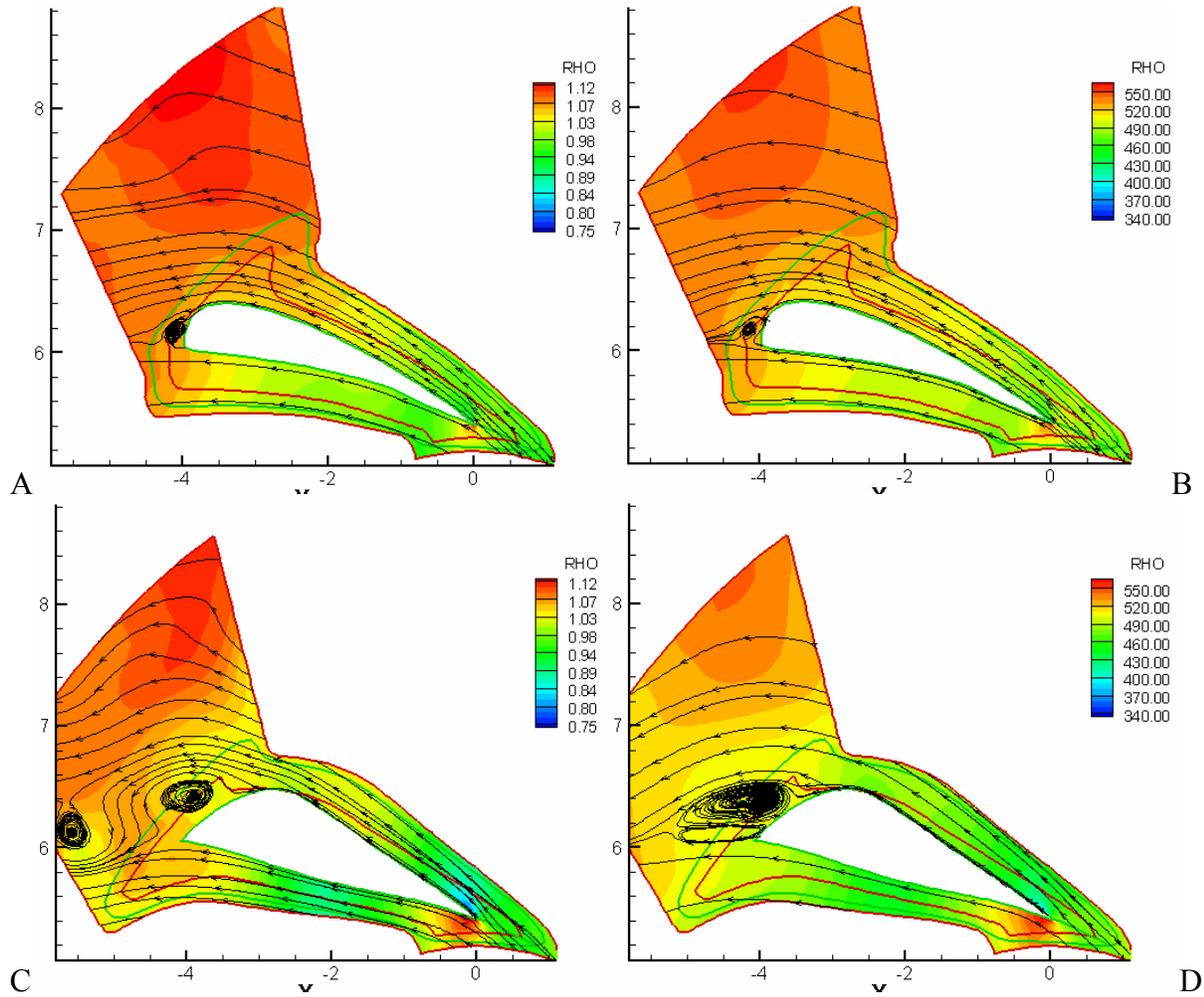


Figure 7-14. Comparison of instantaneous and time-averaged flow fields of intermediate optimal (PRS) and baseline designs. A) Instantaneous – Intermediate optimal design. B) Time-averaged – Intermediate optimal design. C) Instantaneous – Baseline design. D) Time-averaged – Baseline design.

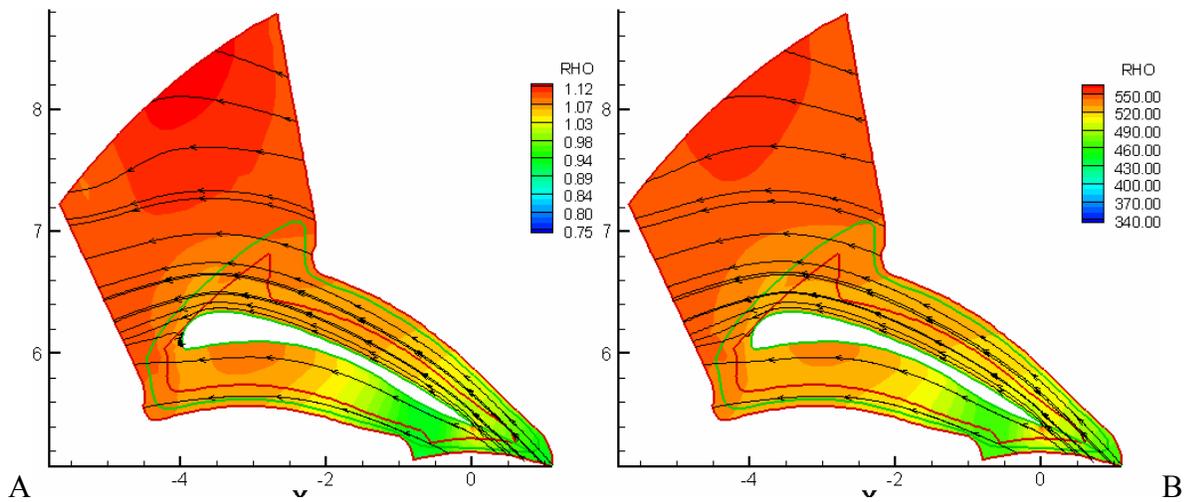


Figure 7-15. Instantaneous and time-averaged pressure for the final optimal diffuser vane shape.
 A) Instantaneous pressure and streamlines. B) Time-averaged pressure and streamlines.

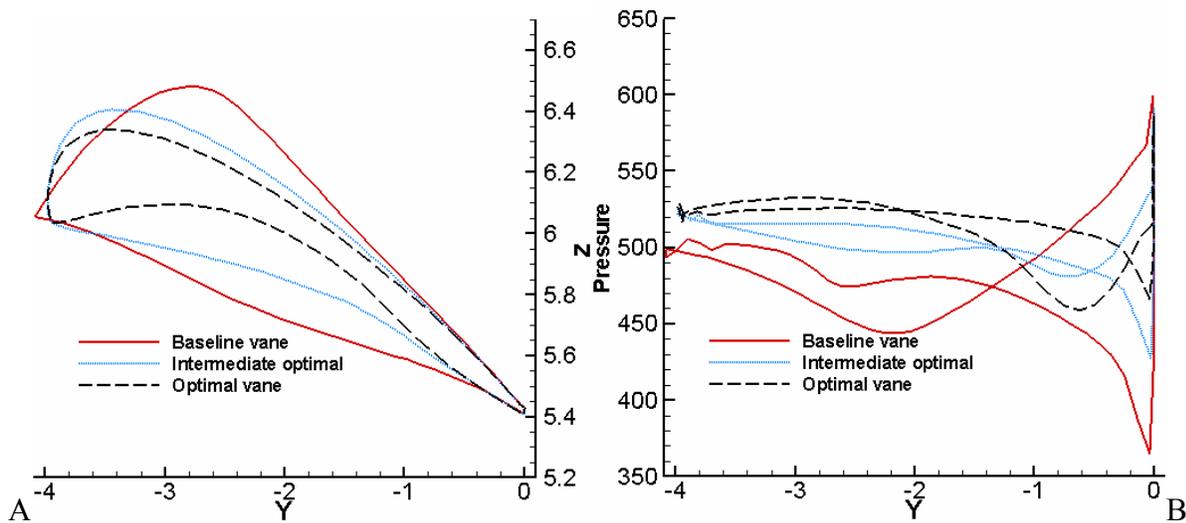


Figure 7-16. Pressure loadings on different vanes. A) Different vane shapes. B) Corresponding pressure loadings.

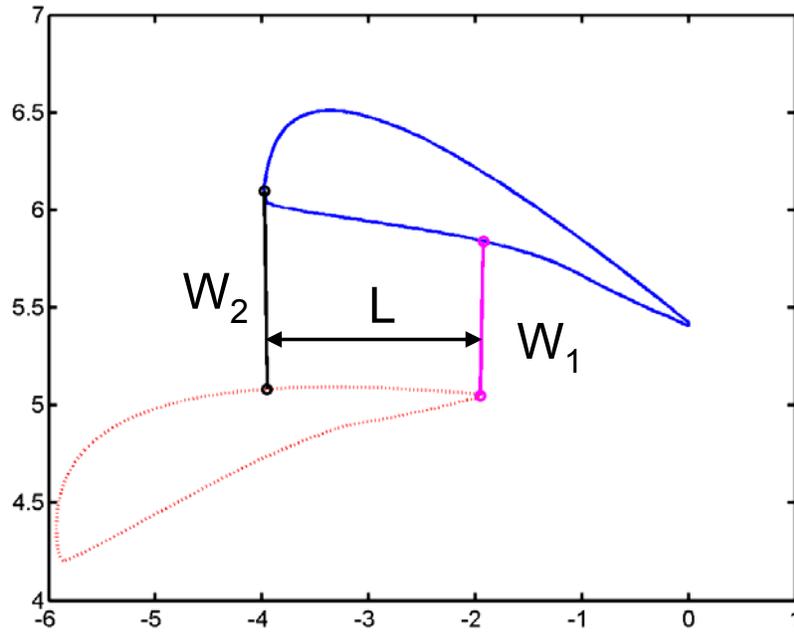


Figure 7-17. Gaps between adjacent vanes.

Table 7-1. Design variables and corresponding ranges. Angle θ is given in degrees and all other dimensions are scaled according to the baseline design.

| | Minimum | Maximum |
|----------|---------|---------|
| θ | 60° | 110° |
| t_1 | 1.25 | 2.50 |
| t_2 | 0.75 | 1.50 |
| t_3 | 0.05 | 0.30 |
| t_4 | 0.50 | 1.00 |
| t_5 | 0.50 | 1.00 |
| t_6 | 1.00 | 2.00 |
| P_y | -2.50 | -2.00 |
| P_z | 5.70 | 5.85 |

Table 7-2. Summary of pressure ratio on data points and performance metrics for different surrogate models fitted to Set A. We tabulate the weights associated with the surrogates used to construct PRESS-based weighted surrogate (PWS). PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural networks, RMSE: Root mean square error, PRESS: Predicted residual sum of squares. Here we give the square root of PRESS so as to facilitate easy comparison with RMSE.

| Surrogate | Parameter | Value | Weight |
|----------------|---------------------|---------|--------|
| Pressure ratio | # of points | 110 | |
| | Minimum of data | 1.001 | |
| | Mean of data | 1.041 | |
| | Maximum of data | 1.093 | |
| PRS | R_{adj}^2 | 0.863 | 0.32 |
| | RMSE | 8.00e-3 | |
| | PRESS | 1.33e-2 | |
| | Max absolute error | 2.62e-2 | |
| | Mean absolute error | 4.40e-3 | |
| KRG | Process variance | 1.09e-4 | 0.41 |
| | PRESS | 1.02e-2 | |
| RBNN | PRESS | 1.58e-2 | 0.27 |
| | Max absolute error | 2.03e-2 | |
| | Mean absolute error | 3.33e-3 | |
| PWS | Max absolute error | 1.39e-2 | |
| | Mean absolute error | 1.95e-3 | |

Table 7-3. Range of data, quality indicators for different surrogate models, and weights associated with the components of PWS. PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural networks, PWS: PRESS-based weighted surrogate, RMSE: Root mean square error, PRESS: Predicted residual sum of squares (in PRS terminology), Here, we give the square root of PRESS so as to facilitate easy comparison with RMSE. We used a reduced cubic and a reduced quadratic polynomial to approximate the Set B and Set C data, respectively.

| Surrogate | Parameter | Set B | | Set C | |
|----------------|---------------------|---------|--------|---------|--------|
| | | Value | Weight | Value | Weight |
| Pressure ratio | # of points | 439 | | 109 | |
| | Minimum of data | 1.000 | | 1.052 | |
| | Mean of data | 1.040 | | 1.075 | |
| | Maximum of data | 1.097 | | 1.105 | |
| PRS | R_{adj}^2 | 0.959 | 0.44 | 0.978 | 0.36 |
| | RMSE | 4.07e-3 | | 1.65e-3 | |
| | PRESS | 4.84e-3 | | 8.74e-3 | |
| | Max absolute error | 1.02e-2 | | 3.84e-3 | |
| | Mean absolute error | 2.67e-3 | | 1.05e-3 | |
| KRG | Process variance | 1.11e-4 | 0.37 | 1.05e-4 | 0.34 |
| | PRESS | 5.89e-3 | | 9.47e-3 | |
| RBNN | PRESS | 1.17e-2 | 0.19 | 1.07e-2 | 0.30 |
| | Max absolute error | 1.53e-2 | | 2.57e-2 | |
| | Mean absolute error | 1.47e-3 | | 2.89e-3 | |
| PWS | Max absolute error | 6.85e-3 | | 7.91e-3 | |
| | Mean absolute error | 1.30e-3 | | 1.04e-3 | |

Table 7-4. Optimal design variables and pressure ratio (P-ratio) obtained using different surrogates constructed using Set C data. PRS is polynomial response surface, KRG is kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate.

| Surrogate | θ | T_1 | t_2 | t_3 | t_4 | t_5 | t_6 | P_y | P_z | Predicted P-ratio |
|-----------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------------------|
| PRS | 60.00 | 1.25 | 0.75 | 0.05 | 1.00 | 1.00 | 1.00 | -2.00 | 5.85 | 1.120 |
| KRG | 63.58 | 1.25 | 0.80 | 0.05 | 0.81 | 0.88 | 1.43 | -2.00 | 5.85 | 1.111 |
| RBNN | 64.49 | 1.32 | 0.78 | 0.07 | 0.94 | 0.54 | 1.08 | -2.04 | 5.84 | 1.105 |
| PWS | 63.65 | 1.25 | 0.75 | 0.05 | 1.00 | 0.53 | 1.09 | -2.00 | 5.85 | 1.109 |

Table 7-5. Comparison of actual and predicted pressure ratio of optimal designs obtained from multiple surrogate models (Set C). Each row shows the results of optimal design using a particular surrogate, and different columns show the prediction of different surrogate models at each optimal design. PRS is polynomial response surface, KRG is kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate.

| Design by | Actual P-ratio | Surrogate prediction using | | | |
|------------------------|----------------|----------------------------|--------|--------|--------|
| | | PRS | KRG | RBNN | PWS |
| PRS | 1.117 | 1.120 | 1.103 | 1.084 | 1.103 |
| KRG | 1.113 | 1.112 | 1.111 | 1.085 | 1.104 |
| RBNN | 1.106 | 1.106 | 1.104 | 1.105 | 1.105 |
| PWS | 1.114 | 1.116 | 1.108 | 1.102 | 1.109 |
| Average absolute error | | 0.0014 | 0.0060 | 0.0186 | 0.0072 |

Table 7-6. Modified ranges of design variables and fixed parameters in refined design space.

| Variables | Min | Max | Fixed parameters | Value | Fixed parameters | Value |
|-----------|-------|-------|------------------|------------|------------------|-------|
| T_2 | 0.60 | 0.75 | θ | 60° | t_4 | 1.00 |
| P_y | -2.00 | -1.50 | t_1 | 1.25 | t_5 | 1.00 |
| P_z | 5.85 | 6.00 | t_3 | 0.05 | t_6 | 1.00 |

Table 7-7. Range of data, summary of performance indicators, and weighted associated with different surrogate models in the refined design space. PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural networks, PWS: PRESS-based weighted surrogate, RMSE: Root mean square error, PRESS: Predicted residual sum of squares (in PRS terminology). Here we give the square root of PRESS so as to facilitate easy comparison with RMSE.

| Surrogate | Parameter | Value | Weight |
|----------------|---------------------|------------------|---------|
| Pressure ratio | # of points | 20 | |
| | Minimum of data | 1.117 | |
| | Mean of data | 1.136 | |
| | Maximum of data | 1.151 | |
| PRS | R_{adj}^2 | 0.956 | 0.59 |
| | RMSE | 1.75e-3 | |
| | PRESS | 2.74e-3 | |
| | Max absolute error | 2.53e-3 | |
| | Mean absolute error | 1.11e-3 | |
| | KRG | Process variance | 1.21e-4 |
| RBNN | PRESS | 7.14e-3 | |
| | PRESS | 1.11e-2 | 0.16 |
| | Max absolute error | <1.0e-6 | |
| | Mean absolute error | <1.0e-6 | |
| PWS | PRESS | 4.21e-3 | |
| | Max absolute error | 1.51e-3 | |
| | Mean absolute error | 6.62e-4 | |

Table 7-8. Design variables and pressure ratio at the optimal designs predicted by different surrogates. PRS is polynomial response surface, KRG is kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate.

| | T_2 | P_y | P_z | Actual P-ratio | Surrogate predictions using | | | |
|------------------------|-------|-------|-------|----------------|-----------------------------|----------|----------|----------|
| | | | | | PRS | KRG | RBNN | PWS |
| PRS | 0.60 | -1.99 | 6.00 | 1.150 | 1.151 | 1.151 | 1.155 | 1.152 |
| KRG | 0.60 | -1.87 | 6.00 | 1.149 | 1.150 | 1.152 | 1.152 | 1.151 |
| RBNN | 0.61 | -1.97 | 5.99 | 1.150 | 1.150 | 1.150 | 1.155 | 1.151 |
| PWS | 0.60 | -1.96 | 6.00 | 1.150 | 1.151 | 1.152 | 1.154 | 1.152 |
| Average absolute error | | | | | 7.5E-04 | 1.50E-03 | 4.25E-03 | 1.75E-03 |

Table 7-9. Actual and empirical ratios of gaps between adjacent diffuser vanes.

| | W_1 | W_2 | L | L/W_1 | W_2/W_1 | $P\text{-ratio}$ |
|----------------------|-------|-------|------|---------|-----------|------------------|
| Empirical relations | | | | 4.00 | 1.60 | |
| Baseline vane | 0.66 | 0.94 | 2.09 | 3.15 | 1.42 | 1.074 |
| Intermediate optimal | 0.79 | 1.07 | 2.03 | 2.55 | 1.35 | 1.117 |
| Optimal vane | 0.94 | 1.11 | 2.02 | 2.15 | 1.19 | 1.150 |

CHAPTER 8 SUMMARY AND FUTURE WORK

In this chapter, we summarize the main contributions in the form of major conclusions derived from this work and discuss the prospects of future work.

As we stated in the beginning, the main goal of this work was to develop methodologies for optimal design of space propulsion systems. Firstly, we revisited the challenges in the optimal design of space propulsion systems and illustrated the need of using surrogate models to alleviate high computational expense. Then we highlighted the issues that influence the effective use of surrogate models for design and optimization. To this end, we illustrated the risks in design of experiments based on a single criterion, and difficulties in the choice of surrogate model, and proposed some remedies. We also offered insight into a somewhat neglected topic of appraising the accuracy of error estimation models that can aid in optimization by correctly identifying regions of high uncertainties. Finally, we showed the application of above strategies to two applications of relevance to space propulsion systems: (1) we used surrogate model based strategy for model validation and sensitivity evaluation of a cryogenic cavitation model, and (2) we optimized the hydrodynamic performance of the diffuser. We briefly recapitulate important lessons from this work and the scope of future work as follows.

Pitfalls of Using a Single Criterion for Experimental Designs

Summary and Learnings

- We illustrated the non-dominated nature of different types of experimental designs while considering multiple criteria, and demonstrated the risks in using a single criterion to construct experimental designs. Particularly, min-max RMS bias design, which minimizes the maximum RMS bias error in the entire design space, may yield a design that is very sensitive to noise.
- Popular experimental designs like LHS and D-optimal designs can leave large holes in design space that may potentially cause large errors in approximation without any indication of poor quality of surrogate model.

- Face-centered central composite design, which is an intuitive design, performs very well on multiple criteria but this design is not practical for high-dimensional spaces.
- D-optimal designs are better than LHS designs in reducing maximum errors but the latter is better for space-averaged bias errors.
- To alleviate the risk of using a single criterion based experimental design, we explored multiple strategies. We demonstrated possible advantages of simultaneously considering complimentary criteria. In particular, we showed improvements in space-averaged errors by combining two such criteria, model based D-optimality criterion, which caters to variance, and geometry based LHS design, which improves the distribution of points in design space.
- We further evidenced the elimination of poor experimental designs by selecting one out of three experimental designs according to an appropriate criterion.

Future Work

We have posed the problem of need to simultaneously consider multiple criteria to reduce risk of running into poor experimental designs. However, a significant amount of research is required to answer the following two questions, (1) which criteria should be considered?, and (2) how to simultaneously accommodate multiple criteria?

Ensemble of Surrogates

Summary and Learnings

- We demonstrated that a single surrogate model may not be appropriate for all problems. Instead the best surrogate depends on the design of experiment, nature of the problem, and the amount of data used to develop surrogate model.
- We showed that simultaneously using multiple surrogate models may prove more effective than using a single surrogate model.
- We proposed a method to develop a weighted-average of surrogates using a cross-validation estimate of surrogate performance. The proposed weighted surrogate performed at par with the best individual surrogate and protected us from the worst surrogate with lower sensitivity to the choice of design of experiment, sampling density, and the dimensionality of the problem.
- Using multiple surrogates, we could identify regions where uncertainty in predictions is significant, and measures like, adaptive sampling can be taken to improve predictions.

Future Work

The simultaneous use of multiple surrogates is a low cost method to account for uncertainties due to choice of approximation model. Some of the possible areas that can be explored in future are as follows: (1) to identify the suitability of many different surrogates for model averaging, (2) to develop different schemes for model averaging, (3) to develop different methods for selection of weights, and (4) to develop error prediction models for weighted average surrogates.

Accuracy of Error Estimates for Noise-free Functions

Summary and Learnings

Though practically very useful, error models have not received much attention in engineering optimization, possibly because of lack of confidence in their accuracy. We compared different error estimation models with the help of various example problems.

- The main finding was that model independent error measures perform equal to or better than the model dependent error estimators.
- Generalized cross-validation error yields a reasonable estimate of actual root mean square error in the entire design space, though it usually overestimates the errors. While estimated root mean square error for polynomial response surface (PRS) approximation consistently underpredicts actual errors, it performs much better than process variance for kriging.
- Among local error estimation models, no error model characterizes the entire error field well for all problems.
- Estimated standard error for PRS underestimates actual RMS errors and identifies the high error regions quite accurately. This error estimation model is least influenced by the choice of design of experiment and nature of the problem. On the other hand, root mean square bias error model performs the best when the discrepancy between assumed true model and actual function is small. RMS bias error mostly overestimates actual errors in PRS.
- Mean square error measure for kriging underestimates actual errors, and shows little variation with the nature of the problem. This error model very accurately identifies potential high error regions.
- The standard deviation of responses, which is a model-independent pointwise error measure, also characterizes the actual root mean square errors reasonably well for different

designs of experiments, problems, and surrogates. However, this model performs very poorly when any constituent surrogate predictions are very bad.

- We also explored possibilities of simultaneously using multiple error models to improve the error prediction characteristics. A geometric averaging of error measures, like estimated RMS error and PRESS for PRS, estimated standard error and RMS bias error for PRS, mean square error and standard deviation of responses for kriging, improves robustness with respect to designs of experiments and nature of the problem, in prediction of actual root mean square error in the entire design space.
- We showed that simultaneous use of multiple error models for locating the high error regions reduces the risk of failure, though it causes false alarms quite often.
- We observed encouraging results in identifying the appropriate error model using the error prediction capabilities at data points using a generalized cross validation based approach.

Future Work

We showed that while global error estimation models are reasonably well developed, there is a need to improve the capability of local error modeling, particularly in the area of model-independent error measures that can be used for any problem. Besides, we can significantly benefit by developing better ways to use an ensemble of error models. Some areas worth-exploring are weighted averaging of error models, reduction in the false alarm while using multiple error models, and so on.

System Identification of Cryogenic Cavitation Model

Summary and Learnings

We studied one cryogenic cavitation model (Merkle et al., 1998) in detail to assess the influence of thermo-sensitive material properties, model parameters, and thermal effects on the prediction of cavitating flow variables in cryogenic environments; and used surrogate based approach to calibrate the model parameters. Firstly, we studied the influence of variation in model parameters C_{dest} and C_{prod} , and uncertainties in material properties, latent heat of vaporization L and vapor density ρ_v , on the prediction of pressure and temperature in a liquid nitrogen flow over a hydrofoil in a suitably designed tunnel. This benchmark case was

extensively studied by Hord (1973a) for different fluids and flow conditions. The performance of cavitation model was characterized by L_2 -norm of the deviation of predicted surface pressure (P_{diff}) and temperature (T_{diff}) data from the experimental data. We approximated the two prediction indicators using surrogate models to limit the computational cost. The main conclusions are enumerated as follows.

- Using a global sensitivity analysis approach proposed by Sobol (1993), we found that the model parameter C_{dest} influenced the performance of the cavitation model the most, and C_{prod} was the least influential parameter. Relatively, uncertainties in material properties were less significant but not negligible. Uncertainty in vapor density was more important of the two selected material properties.
- Further, applying the information from the sensitivity analysis, we calibrated the cryogenic cavitation model for different fluids using the model parameter C_{dest} . The objective of calibration was to simultaneously minimize the competing responses P_{diff} and T_{diff} . Again, using a multiple surrogate based optimization strategy, and noting the importance of pressure predictions that influence cavitation more than the temperature predictions, we found $C_{dest, LN2} = 0.6392$, $C_{dest, LH2} = 0.767$, and $C_{prod} = 54.4$ as the best compromise model parameters.
- The choice of calibrated model parameters improved robustness with respect to different geometries and operating conditions. From physical point of view, the reduction in the parameter C_{dest} , instigated an earlier onset of condensation, and hence the cavity closure, which was difficult to predict using original parameters.
- The application of multiple surrogate models was found very effective in this model validation and calibration exercise.
- The role of thermal environment on predictions was assessed using analysis and simulations. We presented an analytical framework to assess the influence of variation in material properties on the cavitation model performance. We found that the wall heat transfer due to thermal boundary condition does not affect the flow variables significantly, but the thermal effects *via* phase change were very influential in determining the cavity morphology.

Future Work

While the current effort clearly showed improvements in the prediction capabilities of the present cryogenic cavitation model, we noted difficulties in simultaneous prediction of pressure and temperature fields. To this end, one needs to critically probe the current cryogenic cavitation

model and physics of cavitation in cryogenic environment to develop better models. Besides, the surrogate-based approach of model validation can be extended to other flow problems.

Shape Optimization of Diffuser Vanes

Summary and Learnings

We used surrogate based methodology to improve the hydrodynamic performance of diffuser that is an important propulsion component.

- The optimized design has thinner, non-convex shaped diffuser vanes.
- While these vanes may be susceptible to the manufacturing and structural difficulties, the improvements in the hydrodynamic performance are significant.
- We observed that initial speeding of the flow can help avoid separation at a later stage, which helps reduce pressure loss.
- The design approach was aided by different aspects of surrogate modeling. Specifically, the reasonable design space approach helped us identify the high performance region, the dimensionality of the problem was reduced following a global sensitivity analysis, and the use of multiple surrogate models protected us from obtaining a sub-optimal design as the best surrogate modeled changed during the course of optimization.

Future Work

This problem served as a proof of concept for the future development of space propulsion systems that may involve cavitation or other complex flows. With respect to the diffuser design with the current operating conditions, one can improve the practical utility of the current effort by considering manufacturing and structural constraints. The current analysis was based on a vane-only analysis. In future, we can combine flow path with the vane to improve the performance.

Though the methodologies developed in this work are applied to the design of space-propulsion systems, they are generic in nature and can be utilized for any specialty.

APPENDIX A
THEORETICAL MODELS FOR ESTIMATING POINTWISE BIAS ERRORS

Let the true response $\eta(\mathbf{x})$ at a design point \mathbf{x} be represented by a polynomial $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$, where $\mathbf{f}(\mathbf{x})$ is the vector of basis functions, and $\boldsymbol{\beta}$ is the vector of coefficients. The vector $\mathbf{f}(\mathbf{x})$ has two components: $\mathbf{f}^{(1)}(\mathbf{x})$ is the vector of basis functions used in the polynomial response surface model, and $\mathbf{f}^{(2)}(\mathbf{x})$ is the vector of additional basis functions that are missing in the linear regression model. Similarly, the coefficient vector $\boldsymbol{\beta}$ can be written as a combination of vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ that represent the true coefficients associated with the basis function vectors $\mathbf{f}^{(1)}(\mathbf{x})$ and $\mathbf{f}^{(2)}(\mathbf{x})$, respectively. Precisely,

$$\eta(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} = \begin{bmatrix} \mathbf{f}^{(1)T} & \mathbf{f}^{(2)T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}^{(2)}. \quad (\text{A1})$$

Assuming normally distributed noise ε with zero mean and variance σ^2 ($N(0, \sigma^2)$), the observed response $y(\mathbf{x})$ at a design point \mathbf{x} is given as

$$y(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon. \quad (\text{A2})$$

If there is no noise, the true response $\eta(\mathbf{x})$ is the same as the observed response $y(\mathbf{x})$.

Then, the true response for N_s design points $(\mathbf{x}^{(i)}, i = 1, \dots, N_s)$ in matrix notation is

$$\mathbf{y} = X\boldsymbol{\beta} = \begin{bmatrix} X^{(1)} & X^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} = X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)}, \quad (\text{A3})$$

where \mathbf{y} is the vector of observed responses at the data points, $X^{(1)}$ is the Gramian matrix constructed using the basis functions corresponding to $\mathbf{f}^{(1)}(\mathbf{x})$, and $X^{(2)}$ is constructed using the missing basis functions corresponding to $\mathbf{f}^{(2)}(\mathbf{x})$. As an example, a Gramian matrix in two

variables when the PRS model is quadratic and the true response is cubic (with monomial basis functions)

$$X = \left[\begin{array}{cccccc|cccc} 1 & x_1^{(1)} & x_2^{(1)} & x_1^{(1)2} & x_1^{(1)}x_2^{(1)} & x_2^{(1)2} & x_1^{(1)3} & x_1^{(1)2}x_2^{(1)} & x_1^{(1)}x_2^{(1)2} & x_2^{(1)3} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_1^{(2)2} & x_1^{(2)}x_2^{(2)} & x_2^{(2)2} & x_1^{(2)3} & x_1^{(2)2}x_2^{(2)} & x_1^{(2)}x_2^{(2)2} & x_2^{(2)3} \\ \vdots & \vdots \\ 1 & x_1^{(i)} & x_2^{(i)} & x_1^{(i)2} & x_1^{(i)}x_2^{(i)} & x_2^{(i)2} & x_1^{(i)3} & x_1^{(i)2}x_2^{(i)} & x_1^{(i)}x_2^{(i)2} & x_2^{(i)3} \\ \vdots & \vdots \\ 1 & x_1^{(N_s)} & x_2^{(N_s)} & x_1^{(N_s)2} & x_1^{(N_s)}x_2^{(N_s)} & x_2^{(N_s)2} & x_1^{(N_s)3} & x_1^{(N_s)2}x_2^{(N_s)} & x_1^{(N_s)}x_2^{(N_s)2} & x_2^{(N_s)3} \end{array} \right] \quad (\text{A4})$$

$\underbrace{\hspace{10em}}_{X^{(1)}} \quad \underbrace{\hspace{10em}}_{X^{(2)}}$

The predicted response $\hat{y}(\mathbf{x})$ at a design point \mathbf{x} is given as a linear combination of approximating basis function vector $\mathbf{f}^{(1)}(\mathbf{x})$ with corresponding estimated coefficient vector \mathbf{b} :

$$\hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}. \quad (\text{A5})$$

The estimated coefficient vector \mathbf{b} is evaluated using the data for N_s design points as (Myers and Montgomery, 1995, Chapter 2)

$$\mathbf{b} = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} \mathbf{y}. \quad (\text{A6})$$

Substituting for \mathbf{y} from Equation (A3) in Equation (A6) gives

$$\mathbf{b} = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} [X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)}], \quad (\text{A7})$$

that can be rearranged as

$$\mathbf{b} = \boldsymbol{\beta}^{(1)} + A\boldsymbol{\beta}^{(2)}, \text{ where } A = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} X^{(2)}, \quad (\text{A8})$$

is called the alias matrix. Equation (A8) can be rearranged as

$$\boldsymbol{\beta}^{(1)} = \mathbf{b} - A\boldsymbol{\beta}^{(2)} \quad (\text{A9})$$

Note that this relation is valid only if Equation (A3) is satisfied (i.e., no noise). The bias error at N_s design points is defined as

$$\mathbf{e}_b = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - X^{(1)}\mathbf{b}. \quad (\text{A10})$$

Substituting for \mathbf{y} from Equation (A3) and for \mathbf{b} from Equation (A8) gives

$$\mathbf{e}_b = X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)} - X^{(1)}[\boldsymbol{\beta}^{(1)} + A\boldsymbol{\beta}^{(2)}] = [X^{(2)} - X^{(1)}A]\boldsymbol{\beta}^{(2)}. \quad (\text{A11})$$

Thus, the bias error at N_s design points is a function of the coefficient vector $\boldsymbol{\beta}^{(2)}$ only. The error at a general design point \mathbf{x} is the difference between the true response and the predicted response, $e(\mathbf{x}) = \eta(\mathbf{x}) - \hat{y}(\mathbf{x})$. When bias error is dominant, $e(\mathbf{x}) \approx e_b(\mathbf{x})$, where $e_b(\mathbf{x})$ is the bias error at design point \mathbf{x} . Substituting values from Equations (A1) and (A5) gives

$$e_b(\mathbf{x}) \approx \eta(\mathbf{x}) - \hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}^{(2)} - (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}. \quad (\text{A12})$$

This expression can be used to estimate pointwise estimates of RMS bias error and/or bounds on bias error.

Data-Independent Error Measures

Firstly, we develop measures to estimate error prior to generation of data that is no experiment/simulation is conducted. Such error measures are useful for determining the experimental designs. We assume that there is no noise in data, and the true function ($\mathbf{f}(\mathbf{x})$) and the polynomial response surface model ($\mathbf{f}^{(1)}(\mathbf{x})$) are known such that Equation (A3) is satisfied. Substituting Equation (A9) in Equation (A12) and rearranging terms, we get,

$$e_b(\mathbf{x}) \approx [(\mathbf{f}^{(2)}(\mathbf{x})) - A^T(\mathbf{f}^{(1)}(\mathbf{x}))]^T \boldsymbol{\beta}^{(2)}. \quad (\text{A13})$$

This shows that for a given experimental design (alias matrix A is fixed), the bias error at a general design point \mathbf{x} depends only on the true coefficient vector $\boldsymbol{\beta}^{(2)}$. So, pointwise bounds or root mean square estimates of bias error can be obtained by supplying the distribution of the true coefficient vector $\boldsymbol{\beta}^{(2)}$.

Data-Independent Bias Error Bounds

The bound on bias errors characterize the maximum error considering all possible true functions. Mathematically, pointwise data-independent bias error bounds are given as,

$$e_b^I(\mathbf{x}) = \max_{\boldsymbol{\beta}} (e_b(\mathbf{x})) = \max_{\boldsymbol{\beta}^{(2)}} \left(\left[(\mathbf{f}^{(2)}(\mathbf{x})) - A^T (\mathbf{f}^{(1)}(\mathbf{x})) \right]^T \boldsymbol{\beta}^{(2)} \right). \quad (\text{A14})$$

For a given experimental design, the term $\left[(\mathbf{f}^{(2)}(\mathbf{x})) - A^T (\mathbf{f}^{(1)}(\mathbf{x})) \right]$ is a function of \mathbf{x} only.

Define,

$$\mathbf{m}(\mathbf{x}) = \left[(\mathbf{f}^{(2)}(\mathbf{x})) - A^T (\mathbf{f}^{(1)}(\mathbf{x})) \right]. \quad (\text{A15})$$

Then, Equation (A14) can be written as,

$$e_b^I(\mathbf{x}) = \max_{\boldsymbol{\beta}^{(2)}} \left(\mathbf{m}^T(\mathbf{x}) \boldsymbol{\beta}^{(2)} \right). \quad (\text{A16})$$

Obviously Equation (A16) is maximized, when each term of the coefficient vector $\boldsymbol{\beta}^{(2)}$ takes an extreme value.

With no data on the distribution, the principle of maximum entropy was followed by assuming that all components of the coefficient vector $\boldsymbol{\beta}^{(2)}$ have a uniform distribution between $-\gamma$ and γ (γ is a constant). Then,

$$e_b^I(\mathbf{x}) = \gamma \sum_{i=1}^{N_{\beta_2}} |m_i(\mathbf{x})|. \quad (\text{A17})$$

Data-Independent RMS Bias Error

The data-independent root mean square of bias error $e_b^{rmsI}(\mathbf{x})$ at design point \mathbf{x} , that characterizes the average error, is obtained by computing its L_2 norm $\sqrt{E_{\boldsymbol{\beta}}(e_b^2(\mathbf{x}))}$ ($E_{\boldsymbol{\beta}}(\mathbf{x})$ denotes the expected value with respect to $\boldsymbol{\beta}$), where

$$E_{\boldsymbol{\beta}}(e_b^2(\mathbf{x})) = E_{\boldsymbol{\beta}}(e_b(\mathbf{x})e_b^T(\mathbf{x})) \approx E_{\boldsymbol{\beta}}\left(\left[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A\right]\boldsymbol{\beta}^{(2)}\left[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A\right]\boldsymbol{\beta}^{(2)T}\right). \quad (\text{A18})$$

Since the term $[\mathbf{f}^{(2)}(\mathbf{x}) - A^T \mathbf{f}^{(1)}(\mathbf{x})]$ depends only on the design point \mathbf{x} and the experimental design, the above equation can be rewritten as

$$\begin{aligned} E_{\beta} (e_b^2(\mathbf{x})) &\approx E_{\beta} \left([\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] \boldsymbol{\beta}^{(2)} \boldsymbol{\beta}^{(2)T} [\mathbf{f}^{(2)}(\mathbf{x}) - A^T \mathbf{f}^{(1)}(\mathbf{x})] \right) \\ &= [\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] E_{\beta} (\boldsymbol{\beta}^{(2)} \boldsymbol{\beta}^{(2)T}) [\mathbf{f}^{(2)}(\mathbf{x}) - A^T \mathbf{f}^{(1)}(\mathbf{x})]. \end{aligned} \quad (\text{A19})$$

Then,

$$e_b^{rmsI}(\mathbf{x}) = \sqrt{E_{\beta} (e_b^2(\mathbf{x}))} \approx \sqrt{[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] E_{\beta} (\boldsymbol{\beta}^{(2)} \boldsymbol{\beta}^{(2)T}) [\mathbf{f}^{(2)}(\mathbf{x}) - A^T \mathbf{f}^{(1)}(\mathbf{x})]}. \quad (\text{A20})$$

It is obvious from Equation (A20) that the RMS bias error at any point depends on the distribution of the coefficient vector $\boldsymbol{\beta}^{(2)}$. For a uniform distribution between $-\gamma$ and γ for all components of the coefficient vector $\boldsymbol{\beta}^{(2)}$, with simple algebra it can be shown that

$E_{\beta} (\boldsymbol{\beta}^{(2)} \boldsymbol{\beta}^{(2)T}) = \gamma^2 / 3 I$, where I is an $N_{\beta_2} \times N_{\beta_2}$ identity matrix. Substituting this and Equation (A15) in Equation (A20), the pointwise RMS bias error is

$$\begin{aligned} e_b^{rmsI}(\mathbf{x}) &\approx \sqrt{[\mathbf{f}^{(2)T}(\mathbf{x}) - \mathbf{f}^{(1)T}(\mathbf{x})A] \left(\gamma^2 / 3 I \right) [\mathbf{f}^{(2)}(\mathbf{x}) - A^T \mathbf{f}^{(1)}(\mathbf{x})]} \\ &= \gamma / \sqrt{3} \|\mathbf{m}(\mathbf{x})\|. \end{aligned} \quad (\text{A21})$$

where $\|\cdot\|$ represents the norm of the quantity.

Data-Dependent Error Measures

Next, we develop the error measures posterior to generation of data such that the error measures yield an estimate of actual error in approximation. Unlike data-independent error measures, the choice of the true coefficient vector is restricted by the constraint that the data at sampled locations must be satisfied.

Bias Error Bound Formulation

Papila et al. (2005) presented a method to compute pointwise bias error bounds, when data was exactly satisfied by the assumed true model, and there was no noise in data. This formulation was generalized by Goel et al. (2006a) to estimate bias error bounds while accounting for noise in the data, and rank deficiencies in the matrix of equations. Their method to estimate pointwise bias error bound $e_b^{beb}(\mathbf{x})$ is given as follows:

$$e_b^{beb}(\mathbf{x}) = \text{Maximize}_{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}} |e_b(\mathbf{x})|$$

Subject to :

$$\begin{aligned} \mathbf{y} - \mathbf{1}\varepsilon &\leq X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)} \leq \mathbf{y} + \mathbf{1}\varepsilon \\ \mathbf{c}_l^{(1)} &\leq \boldsymbol{\beta}^{(1)} \leq \mathbf{c}_u^{(1)} \\ \mathbf{c}_l^{(2)} &\leq \boldsymbol{\beta}^{(2)} \leq \mathbf{c}_u^{(2)}, \end{aligned} \tag{A22}$$

where ε (called as relaxation) is the acceptable deviation from the data \mathbf{y} , and $\mathbf{c}_l, \mathbf{c}_u$ are the lower and upper bounds on the coefficient vector $\boldsymbol{\beta}$, respectively. The minimum relaxation required to get a feasible solution of this problem can be obtained by solving,

$$\text{Minimize}_{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \varepsilon_{\min}} \varepsilon_{\min}$$

Subject to:

$$\begin{aligned} \mathbf{y} - \mathbf{1}\varepsilon_{\min} &\leq X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)} \leq \mathbf{y} + \mathbf{1}\varepsilon_{\min} \\ \varepsilon_{\min} &\geq 0 \\ \mathbf{c}_l^{(1)} &\leq \boldsymbol{\beta}^{(1)} \leq \mathbf{c}_u^{(1)} \\ \mathbf{c}_l^{(2)} &\leq \boldsymbol{\beta}^{(2)} \leq \mathbf{c}_u^{(2)}. \end{aligned} \tag{A23}$$

The value of relaxation $\varepsilon \geq \varepsilon_{\min}$ is then chosen as suggested by Goel et al. (2006a).

Equation (A22) can be cast as: $\max(\max(e_b(\mathbf{x})), -\min(e_b(\mathbf{x})))$, where $e_b(\mathbf{x})$ is computed using (A12). So, bias error bounds at point \mathbf{x} can be obtained by solving two linear programming problems, one to maximize bias error, and second to minimize bias error, subject to data and appropriate bounds on the coefficient vector $\boldsymbol{\beta}$.

Root Mean Square Bias Error Formulation

The root mean square of bias error $e_b^{rms}(\mathbf{x})$ at design point \mathbf{x} is obtained by computing its L_2 norm (Goel et al. 2006c, 2007b) as follows:

$$e_b^{rms}(\mathbf{x}) = \sqrt{E_\beta(e_b^2(\mathbf{x}))} = \sqrt{E_\beta(e_b(\mathbf{x})e_b^T(\mathbf{x}))}, \quad (\text{A24})$$

where $E_\beta(g(x))$ is the expected value of $g(x)$ with respect to β . Substituting for $e_b(\mathbf{x})$ from Equation (A12),

$$E_\beta(e_b^2(\mathbf{x})) = E_\beta(e_b(\mathbf{x})e_b^T(\mathbf{x})) = E_\beta\left(\left((\mathbf{f}^{(1)}(\mathbf{x}))^T(\boldsymbol{\beta}^{(1)} - \mathbf{b}) + (\mathbf{f}^{(2)}(\mathbf{x}))^T\boldsymbol{\beta}^{(2)}\right)\left((\mathbf{f}^{(1)}(\mathbf{x}))^T(\boldsymbol{\beta}^{(1)} - \mathbf{b}) + (\mathbf{f}^{(2)}(\mathbf{x}))^T\boldsymbol{\beta}^{(2)}\right)^T\right). \quad (\text{A25})$$

Since, $\mathbf{f}^{(1)}(\mathbf{x})$ and $\mathbf{f}^{(2)}(\mathbf{x})$ depend only on \mathbf{x} , Equation (A25) can be rearranged as,

$$E_\beta(e_b^2(\mathbf{x})) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \underbrace{E_\beta\left(\left(\boldsymbol{\beta}^{(1)} - \mathbf{b}\right)\left(\boldsymbol{\beta}^{(1)} - \mathbf{b}\right)^T\right)}_1 (\mathbf{f}^{(1)}(\mathbf{x})) + (\mathbf{f}^{(2)}(\mathbf{x}))^T \underbrace{E_\beta\left(\boldsymbol{\beta}^{(2)}\left(\boldsymbol{\beta}^{(1)} - \mathbf{b}\right)^T\right)}_2 (\mathbf{f}^{(1)}(\mathbf{x})) + (\mathbf{f}^{(1)}(\mathbf{x}))^T \underbrace{E_\beta\left(\left(\boldsymbol{\beta}^{(1)} - \mathbf{b}\right)\boldsymbol{\beta}^{(2)T}\right)}_3 (\mathbf{f}^{(2)}(\mathbf{x})) + (\mathbf{f}^{(2)}(\mathbf{x}))^T \underbrace{E_\beta\left(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)T}\right)}_4 (\mathbf{f}^{(2)}(\mathbf{x})). \quad (\text{A26})$$

This expression can be estimated if the distribution of the coefficient vector $\boldsymbol{\beta}$ is known.

Determining the Distribution of Coefficient Vector $\boldsymbol{\beta}$

We know that RMS bias errors (Equation (A24)), and bias error bounds (Equation (A22)) depend on the information about the coefficient vector $\boldsymbol{\beta}$. The distribution of coefficient vector $\boldsymbol{\beta}$ can be obtained using the data at sampling points as follows. Since the rank of matrix X is lesser than the number of coefficients, there are $N_e = N_\beta - \text{rank}(X)$, zero eigenvalues, and correspondingly, N_e null eigenvectors. Using the properties of null eigenvectors (\mathbf{V}) $X\mathbf{V} = 0$,

we obtain the distribution of coefficient vector $\boldsymbol{\beta}$ as a linear combination of N_e null eigenvectors, given as,

$$\boldsymbol{\beta} = \boldsymbol{\beta}_\varepsilon + \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i \mathbf{V}_i \quad (\text{A27})$$

where coefficient vector $\boldsymbol{\beta}_\varepsilon$ satisfies the data within the limit of minimum relaxation ε_{\min} ,

$\|\boldsymbol{\beta}_\varepsilon\|$ is the norm of vector $\boldsymbol{\beta}_\varepsilon$, and γ_i is the random coefficient associated with i^{th} null

eigenvector \mathbf{V}_i . For a given experimental design, $\boldsymbol{\beta}_\varepsilon$ and \mathbf{V}_i are fixed, so the distribution of $\boldsymbol{\beta}$ is

related to the distribution of γ_i . Using the maximum entropy principle, we assume that all γ_i

follow a uniform distribution $[-\gamma, \gamma]$, where γ is a constant. For a given experimental design,

the coefficient vector $\boldsymbol{\beta}_\varepsilon$ is obtained in two steps:

- We identify the minimum-norm solution ($\boldsymbol{\beta}_m$) of under-determined system of equations $X\boldsymbol{\beta} = \mathbf{y}$. If matrix X is full rank, solution $\boldsymbol{\beta}_m$ exactly satisfies the system of equations, and obviates next step. Otherwise $X\boldsymbol{\beta}_m = \mathbf{y}_m$; and the difference between \mathbf{y} and \mathbf{y}_m denotes imperfections in modeling of the data.
- We specify bounds $\mathbf{c}_l, \mathbf{c}_u$ in Equation (A23) using $\boldsymbol{\beta}_m$ as,

$$\mathbf{c}_l = \boldsymbol{\beta}_m - \gamma |\boldsymbol{\beta}_m| \quad \text{and} \quad \mathbf{c}_u = \boldsymbol{\beta}_m + \gamma |\boldsymbol{\beta}_m|, \quad (\text{A28})$$

and using $\boldsymbol{\beta}_m$ as initial guess in Equation (A23), we find $\boldsymbol{\beta}_\varepsilon$.

Substituting Equation (A27) in Equation (A12), we can rewrite the expression to evaluate bias error at a design point \mathbf{x} as,

$$e_b(\mathbf{x}) = f^T(\mathbf{x})\boldsymbol{\beta} - (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b} = f^T(\mathbf{x}) \left(\boldsymbol{\beta}_\varepsilon + \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i \mathbf{V}_i \right) - (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}. \quad (\text{A29})$$

This can be rearranged as,

$$e_b(\mathbf{x}) = \underbrace{f^T(\mathbf{x})\boldsymbol{\beta}_\varepsilon - (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}}_{\text{Constant}(e_0)} + \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i. \quad (\text{A30})$$

The first term in above expression is a constant e_0 for given data (fixed \mathbf{b}), and experimental design (fixed $\boldsymbol{\beta}_\varepsilon$). Then, the bias error depends only on the distribution of γ_i ,

$$e_b(\mathbf{x}) = e_0 + \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i. \quad (\text{A31})$$

The advantages of the above formulation of bias error is that, now we can develop analytical expressions for bias error bounds and root mean square bias errors, if we know γ_i .

Analytical Expression for Pointwise Bias Error Bound

The pointwise bounds on bias error are obtained by maximizing Equation (A31) over all possible values of coefficients γ_i . That is,

$$e_b^{beb}(\mathbf{x}) = \max_{-\gamma \leq \gamma_i \leq \gamma} \left| e_0 + \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i \right|. \quad (\text{A32})$$

This can be rearranged as,

$$e_b^{beb}(\mathbf{x}) = \begin{cases} e_0 + \max_{-\gamma \leq \gamma_i \leq \gamma} \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i; & e_0 \geq 0 \\ -e_0 - \min_{-\gamma \leq \gamma_i \leq \gamma} \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i; & e_0 < 0 \end{cases}. \quad (\text{A33})$$

In contrast to Equation (A22), this optimization problem (Equation (A33)) has only side constraints on γ_i , because all coefficient vectors satisfy the data within limit of ε_{\min} . Since all γ_i follow a uniform distribution $[-\gamma, \gamma]$, we rewrite the second term in the above expression,

$$\max \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i = \gamma \|\boldsymbol{\beta}_\varepsilon\| \sum_{i=1}^{N_e} |f^T(\mathbf{x}) \mathbf{V}_i|, \quad (\text{A34})$$

and,

$$\min \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i = -\max \sum_{i=1}^{N_e} \gamma_i f^T(\mathbf{x}) \mathbf{V}_i. \quad (\text{A35})$$

The expression in Equation (A33) is easy to evaluate, even when, the bounds on γ_i are nonuniform. Thus, with the new formulation, we have an analytical expression for pointwise bias error bounds that significantly reduces the computational expense.

Analytical Estimate of Root Mean Square Bias Error

Next, we develop analytical estimates of the expected values of different terms in Equation (A26) by using the assumption on the distribution of γ_i , as follows. Lets consider (i, j)

component $E_\beta(\beta_i^{(2)}(\beta_j^{(1)} - b_j))$ of matrix $E_\beta(\boldsymbol{\beta}^{(2)}(\boldsymbol{\beta}^{(1)} - \mathbf{b})^T)$. Substituting Equation (A27),

$$E_\beta(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) = E_\gamma \left(\left(\beta_{\varepsilon i}^{(2)} + \|\boldsymbol{\beta}_\varepsilon\| \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \right) \left(\beta_{\varepsilon j}^{(1)} + \|\boldsymbol{\beta}_\varepsilon\| \sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} - b_j \right) \right). \quad (\text{A36})$$

Denoting $(\beta_{\varepsilon j}^{(1)} - b_j) = \beta_{\varepsilon j}^{(1)*}$ and expanding terms,

$$E_\beta(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) = E_\gamma \left(\begin{aligned} &\beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \beta_{\varepsilon i}^{(2)} \|\boldsymbol{\beta}_\varepsilon\| \sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} + \beta_{\varepsilon j}^{(1)*} \|\boldsymbol{\beta}_\varepsilon\| \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \\ &+ \|\boldsymbol{\beta}_\varepsilon\| \sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} \|\boldsymbol{\beta}_\varepsilon\| \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \end{aligned} \right). \quad (\text{A37})$$

Rearranging terms,

$$\begin{aligned} E_\beta(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) &= E_\gamma \left(\beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} \right) + E_\gamma \left(\beta_{\varepsilon i}^{(2)} \|\boldsymbol{\beta}_\varepsilon\| \sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} \right) \\ &+ E_\gamma \left(\beta_{\varepsilon j}^{(1)*} \|\boldsymbol{\beta}_\varepsilon\| \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \right) + E_\gamma \left(\|\boldsymbol{\beta}_\varepsilon\|^2 \sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \right). \end{aligned} \quad (\text{A38})$$

Since $\beta_{\varepsilon i}^{(2)}, \beta_{\varepsilon j}^{(1)*}, \|\boldsymbol{\beta}_\varepsilon\|$ are constant,

$$\begin{aligned}
E_{\beta}(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) &= \beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \beta_{\varepsilon i}^{(2)} \|\mathbf{\beta}_{\varepsilon}\| E_{\gamma} \left(\sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} \right) \\
&+ \beta_{\varepsilon j}^{(1)*} \|\mathbf{\beta}_{\varepsilon}\| E_{\gamma} \left(\sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \right) + \|\mathbf{\beta}_{\varepsilon}\|^2 E_{\gamma} \left(\sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \right).
\end{aligned} \tag{A39}$$

Using the property $E(\sum x) = \sum E(x)$, and noting that $V_{qj}^{(1)}, V_{pi}^{(2)}$ are constant and $E_{\gamma}(\gamma_i) = 0$,

Equation (A39) can be further simplified as,

$$\begin{aligned}
E_{\beta}(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) &= \beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \beta_{\varepsilon i}^{(2)} \|\mathbf{\beta}_{\varepsilon}\| \left(\sum_{q=1}^{N_e} E_{\gamma}(\cancel{\gamma_q}) V_{qj}^{(1)} \right) \\
&+ \beta_{\varepsilon j}^{(1)*} \|\mathbf{\beta}_{\varepsilon}\| \left(\sum_{p=1}^{N_e} E_{\gamma}(\cancel{\gamma_p}) V_{pi}^{(2)} \right) + \|\mathbf{\beta}_{\varepsilon}\|^2 E_{\gamma} \left(\sum_{q=1}^{N_e} \gamma_q V_{qj}^{(1)} \sum_{p=1}^{N_e} \gamma_p V_{pi}^{(2)} \right).
\end{aligned} \tag{A40}$$

Using the property $\sum x \sum y = \sum \sum xy$,

$$E_{\beta}(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) = \beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \|\mathbf{\beta}_{\varepsilon}\|^2 E_{\gamma} \left(\sum_{p=1}^{N_e} \sum_{q=1}^{N_e} \gamma_q \gamma_p V_{qj}^{(1)} V_{pi}^{(2)} \right), \tag{A41}$$

$$E_{\beta}(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) = \beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \|\mathbf{\beta}_{\varepsilon}\|^2 \sum_{p=1}^{N_e} \sum_{q=1}^{N_e} E_{\gamma}(\gamma_q \gamma_p) V_{qj}^{(1)} V_{pi}^{(2)}. \tag{A42}$$

$$\text{Since } E_{\gamma}(\gamma_q \gamma_p) = \begin{cases} \gamma^2/3 & p = q \\ 0 & p \neq q \end{cases},$$

$$E_{\beta}(\beta_i^{(2)}(\beta_j^{(1)} - b_j)) = \beta_{\varepsilon i}^{(2)} \beta_{\varepsilon j}^{(1)*} + \|\mathbf{\beta}_{\varepsilon}\|^2 \left(\gamma^2/3 \right) \sum_{p=1}^{N_e} V_{pj}^{(1)} V_{pi}^{(2)}. \tag{A43}$$

Similarly, we can write analytical expressions for other terms and estimate RMS bias errors using Equation (A26).

APPENDIX B
APPLICATIONS OF DATA-INDEPENDENT RMS BIAS ERROR MEASURES

Recap

True function: $\eta(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} = \begin{bmatrix} \mathbf{f}^{(1)T} & \mathbf{f}^{(2)T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}^{(2)}. \quad (\text{B1})$

Approximation: $\hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}, \quad (\text{B2})$

where $\mathbf{b} = (X^{(1)T} X^{(1)})^{-1} X^{(1)T} \mathbf{y}. \quad (\text{B3})$

Data-independent root mean square bias error: $e_b^{rmsl}(\mathbf{x}) = \gamma / \sqrt{3} \|\mathbf{m}(\mathbf{x})\|, \quad (\text{B4})$

where $\mathbf{m}(\mathbf{x}) = [(\mathbf{f}^{(2)}(\mathbf{x})) - A^T (\mathbf{f}^{(1)}(\mathbf{x}))]. \quad (\text{B5})$

Construction of Experimental Designs

We construct a central composite design (CCD), a popular minimum variance design, in an N_v -dimensional cube $V = [-1, 1]^{N_v}$ for N_v ranging from two to five. For N_v dimensions, the CCD has $2^{N_v} + 2N_v + 1$ points. Minimum-bias experimental designs are constructed using two parameters $0 \leq \alpha_1, \alpha_2 \leq 1$ that define the coordinates of sampled data points. Points corresponding to vertices are placed at $x_i = \pm\alpha_1$ for $i = 1, \dots, N_v$, and axial points are located at $x_i = \pm\alpha_2, x_j = 0$ for all $j \neq i$, for each $i = 1, \dots, N_v$. In addition, one point is placed at the origin. Figure B-1 shows an example of an ED as a function of parameters α_1, α_2 in two dimensions.

The min-max RMS bias design can be obtained by identifying parameters α_1, α_2 such that the maximum value of (data independent) RMS bias error in design space $(e_b^{rmsl})_{\max}$ is minimized. Mathematically, this is a two-level optimization problem:

$$e_{\min}^b = \min_{\alpha_1, \dots, \alpha_m} (e_b^{rmsI})_{\max} = \min_{0 \leq \alpha_1, \dots, \alpha_m \leq 1} (\max_V |e_b^{rmsI}(\mathbf{x})|). \quad (\text{B6})$$

The inner problem is the maximization of RMS bias error over the design space (this is a function of α_i and pointwise errors are estimated by Equation (B4)), and the outer level problem is minimization over the parameters α_i . Note that knowledge of $\mathbf{b}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}$ is not required to generate an ED.

We used the conventional assumption of a quadratic PRS and a true model being a cubic polynomial to compute bias errors. The maximum RMS bias error for a given combination of parameters α_1 and α_2 was obtained by evaluating pointwise RMS bias errors on a uniform 11^{Nv} grid. The min-max RMS bias designs for different dimensional spaces are given in Table 3-5. In low dimensions, the optimal min-max RMS design was obtained by placing vertex points inwards while keeping the axial point on the face. However, for higher dimensions the optimal design required the axial point to be close to the center while forcing the vertex point on the corner. This design has a very low maximum and average RMS bias error but this configuration leads to very standard errors, that is, this design would be very sensitive to noise errors.

Why Min-max RMS Bias Designs Place Points near Center for Four-dimensional Space?

The explanation of this unexpected result obtained for higher dimensional cases is as follows. There are three types of cubic terms: $x_i^3, x_i x_j^2, x_i x_j x_k$. The experimental design is a compromise to minimize errors due to the three. The term x_i^3 is modeled as x_i , no matter where the vertex and axial points are located. Similarly, the $x_i x_j x_k$ term will always be modeled by zero. The term $x_i x_j^2$ will be modeled by $c x_i$, where the constant c is determined by the position of axial and vertex points. Vertex points favor $c = \alpha_1^2$, and axial points favor $c = 0$. Because the

$x_i x_j x_k$ term is modeled by zero, it generates large bias errors. As the dimension of the problem increases, there are relatively more of these terms. For $N_v = 2$, their number is zero out of four terms; for $N_v = 3$, it is one out of 10 terms; for $N_v = 4$, it is four out of 20 and for $N_v = 5$, it is 10 out of 35. Furthermore, unlike the other terms, all the $x_i x_j x_k$ terms have a common point where they peak: the all-positive or all-negative corners. These terms then dominate the choice of experimental design for high dimensional spaces. If axial points are placed inside the design space, the term $x_i x_j^2$ contributes to the error at all-positive and all-negative corners ($c \leq \alpha_1^2$). However, when the axial points (hence α_2) go to zero, we can have $c \rightarrow \alpha_1^2$ which minimizes additional errors at the all-positive and all-negative corners.

Verification of Experimental Designs

To verify the results for min-max RMS bias ED, we compared predicted RMS bias errors with actual RMS bias errors in four-dimensional design space. A cubic true function was approximated by a quadratic polynomial. Response and errors were predicted on a uniform grid of 11^4 points. Pointwise RMS bias (predicted) error was estimated using Equation (B4) with constant $\gamma = 1$. To compute actual RMS errors, a large number of true polynomials were generated by randomly selecting the coefficient vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ from a uniform distribution over $[-1, 1]$. For each true polynomial, the true response was obtained using Equation (B1) and the vector \mathbf{b} was evaluated using Equation (A6). The actual error at a point due to each true polynomial was computed by taking the difference between actual and predicted response as

$$e_c(\mathbf{x}) = y(\mathbf{x}) - \hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}_c^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}_c^{(2)} - (\mathbf{f}^{(1)}(\mathbf{x}))^T \mathbf{b}_c, \quad (\text{B7})$$

where the subscript c represents an instance of the true polynomial.

Pointwise actual RMS bias errors $e_{act}^{rms}(\mathbf{x})$ were estimated by averaging actual errors over a large number of polynomials in the root mean squares sense as

$$e_{act}^{rms}(\mathbf{x}) = \sqrt{\sum_{c=1}^{N_P} e_c^2(\mathbf{x}) / N_P}. \quad (\text{B8})$$

N_P ($= 100,000$) true polynomials were used to compute the actual errors. Two experimental designs with extreme values of axial location parameter α_2 were compared. The actual errors and the predicted errors along with the correlations between actual and predicted error estimates are given in Table B-3. Maximum and space-averaged values of actual RMS bias errors and predicted RMS bias errors compared very well. The correlations between pointwise actual RMS bias errors and predicted RMS bias errors were also high. The small change in maximal actual RMS bias error with the location of axial design point confirmed the outcome that maximal RMS bias errors had low variation with respect to the parameter α_2 and placing the axial points close to the center minimized the maximal actual RMS error.

Comparison of Experimental Designs

Different standard designs, for two-dimensional design spaces, available in the literature—minimum variance design ED 1 (minimizes maximum variance, Myers and Montgomery, 1995), minimum space-averaged bias design ED 2 (minimizes space-averaged bias error, Qu et al., 2004), min-max bias error bound design ED 3 (minimizes maximum bias error bound, Papila et al., 2005), and min-max RMS bias design ED 4 (minimizes maximum RMS bias error, Goel et al., 2006c)—were compared using metrics defined in Chapter 3. Errors were computed using a uniform 41x41 grid and results are summarized in Table B-2.

Interestingly, the design based on the most commonly used bias error metric, the minimum space-averaged bias design (ED 2), performed most poorly on all the metrics except the space-

averaged bias errors (RMS and bound). Since for ED 2 the sampled points were located in the interior, there was a significant extrapolation region that explained the very high maximum errors. This suggested that averaging of bias error over the entire space may not be the best criterion to create experimental designs. On the other hand, all other designs gave comparable performance on all metrics. As expected, the experimental designs based on bias errors (min-max bias error bound design, ED 3; and min-max RMS bias design, ED 4) performed better on bias errors and the ED based on min-variance (ED 1) reduced estimated standard error more than the other designs. The differences between the min-max bias error bound design (ED 3) and the min-max RMS bias error design (ED 4) were small. As expected, the results show that a design based on a single criterion does not perform the best on all metrics. Instead we obtain different non-dominated tradeoff solutions.

RMS Bias Error Estimates for Trigonometric Example

Bias error estimates are sometimes criticized because of the assumption that a higher order polynomial than the fitting model is the true function. We demonstrate that this assumption on the true model is practically useful if it captures the main characteristics of the true unknown function. Suppose the true function is a trigonometric polynomial given by

$$\eta(\mathbf{x}) = \sum_{\substack{j+k \leq 6 \\ j, k \geq 0}} \text{Real}\left(-ia_{jk} e^{ijx_1} e^{ikx_2}\right), \quad x_1, x_2 \in [-1, 1]. \quad (\text{B9})$$

The coefficients a_{06}, a_{60} were assumed to be uniformly distributed over $[-10, 10]$ and the remaining a_{jk} were assumed to be uniformly distributed over $[-1, 1]$. This function represents the cumulative response of *sine* waves of different wavelengths. For the given range of parameters j, k, x_1 , and x_2 , the highest frequency (shortest wavelength corresponding to $j = 6$ and

$k = 6$) components in the true function have more than one cycle in design space. To estimate actual errors, 10000 ($N_p = 10000$) combinations of a_{jk} were used.

This function was approximated using a cubic polynomial and bias errors were estimated assuming the true model to be quintic. A uniform 4x4 grid was used for sampling (16 points) and a uniform 21x21 grid was used to estimate errors. The distribution of actual RMS bias error and predicted RMS bias error in design space is shown in Figure B-2. The predicted RMS bias error was scaled by a factor of 55 to compare with the actual errors in design space. Note that prior to generation of data, the actual magnitude of the error was not important as the estimated errors were scaled by an ‘unknown’ factor γ (refer to Equation (B4)) that was arbitrarily set to one. The predicted RMS bias errors correctly identified the presence of high error zones along the diagonals. However, predictions were inaccurate near the center and close to the boundary, where the effect of high frequency terms in the true function was significant. The correlation between the actual RMS bias errors and the predicted RMS bias errors was 0.68. The error contours and correlation coefficient indicated a reasonable agreement between actual RMS bias error and predicted RMS bias error estimates considering the fact that the true function was a high frequency trigonometric polynomial.

For this example problem, the actual distribution of the coefficient vector $\beta^{(2)}$ can be obtained from the distribution of constants a_{jk} in the true function by expanding each *sine* term in Maclaurin series as

$$\sin(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)!} x^{2n-1}, \quad (\text{B10})$$

and observing the coefficients of the quintic terms ($\eta(\mathbf{x})$ is odd, so there are no quartic terms). It was found that the coefficients $\beta_6^{(2)}$ and $\beta_{11}^{(2)}$ (corresponding to x_1^5 and x_2^5 , respectively)

approximately followed a uniform distribution with range $[-700, 700]$ and the coefficients $\beta_7^{(2)}$, $\beta_8^{(2)}$, $\beta_9^{(2)}$, and $\beta_{10}^{(2)}$ (corresponding to $x_1^4x_2$, $x_1^3x_2^2$, $x_1^2x_2^3$, and $x_1x_2^4$, respectively) followed approximately a uniform distribution with range $[-70, 70]$. Using the modified distribution of coefficients and hence $E(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)T})$, RMS bias errors were estimated. Corresponding actual RMS bias error and scaled predicted RMS bias error (scaled by 0.177) contours are shown in Figure B-3. For this case, the agreement between the actual RMS bias error and the predicted RMS bias error improved significantly (compare Figure B-2(A) and Figure B-3(A)) and the correlation between the actual RMS bias errors and the predicted RMS bias errors increased to 0.94. This indicates that supplying the correct distribution of the coefficient vector for the basis functions missing from the response surface model $f^{(2)}(\mathbf{x})$ is helpful to assess the errors.

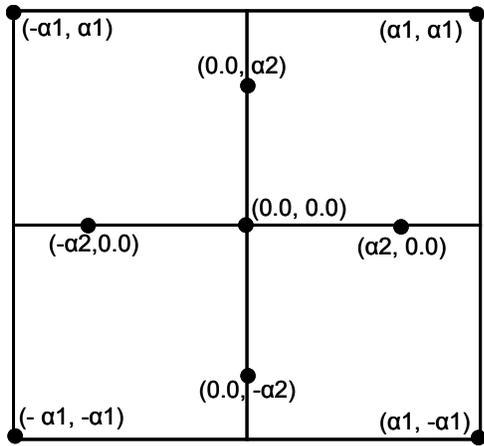


Figure B-1. Two-dimensional illustration of central composite experimental design constructed using two parameters α_1 and α_2 .

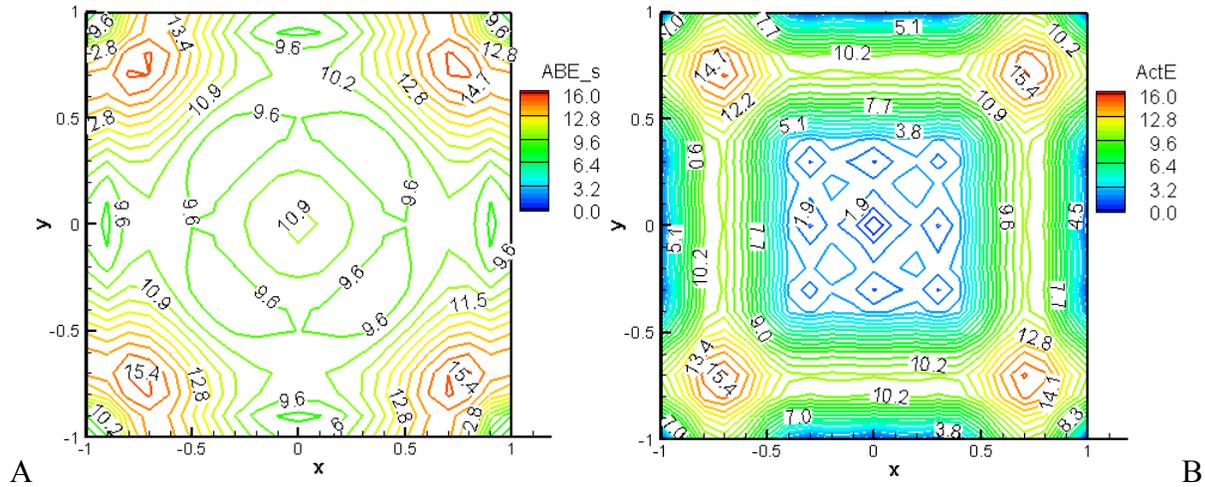


Figure B-2. Contours of scaled predicted RMS bias error and actual RMS error when assumed true model to compute bias error was quintic while the true model was trigonometric (Equation (B9)) (scaled bias error = $55 \times$ predicted RMS bias error). A) Scaled predicted RMS bias error. B) Actual RMS error.

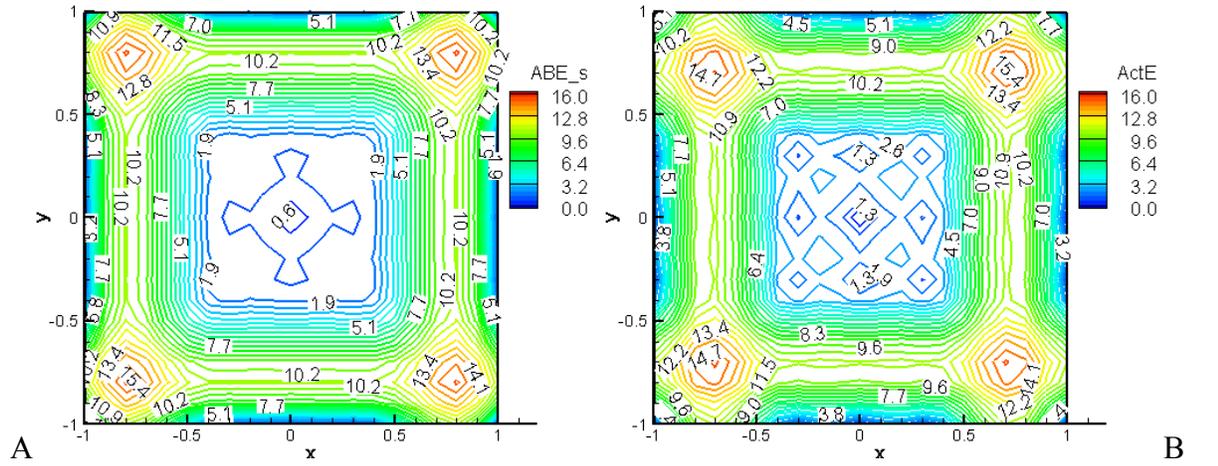


Figure B-3. Contours of scaled predicted RMS bias error and actual RMS error when different distributions of $\beta^{(2)}$ were specified (assumed true model was quintic while the true model was trigonometric (Equation (B9)), (scaled bias error = 0.177*RMS bias error). A) Scaled predicted RMS bias error. B) Actual RMS error.

Table B-1. Design variables and maximum RMS bias errors for min-max RMS bias central composite designs in $N_v=2-5$ dimensional spaces. Errors were computed on a uniform 41^2 grid ($N_v = 2$), 21^3 grid ($N_v = 3$) and, 11^{N_v} grid ($N_v > 3$), in space $V = [-1, 1]^{N_v}$.

| N_v | α_1 | α_2 | $(e_b^{rms})_{\max}$ | $(e_b^{rms})_{\text{avg}}$ |
|-------|------------|------------|----------------------|----------------------------|
| 2 | 0.954 | 1.000 | 0.341 | 0.269 |
| 3 | 0.987 | 1.000 | 0.659 | 0.518 |
| 4 | 1.000 | 0.100 | 1.155 | 0.927 |
| 5 | 1.000 | 0.100 | 1.826 | 1.200 |

Table B-2. Comparison of different experimental designs for two dimensions. (1) Minimum maximum variance design. (2) Minimum space-averaged bias design. (3) Min-max bias error bound design. (4) Min-max RMS bias design. Metrics of measurement were maximum estimated standard error, space-averaged estimated standard error, maximum bias error bound (BEB), maximum RMS bias error, space-averaged RMS bias error (BE). Errors are computed on a uniform 41×41 grid in space $V = [-1, 1]^2$. α_1 and α_2 define the location of sampled points in Figure B-1.

| ED | α_1 | α_2 | $(e_{es})_{\max}$ | $(e_{es})_{\text{avg}}$ | $ e_b^I _{\max}$ | $(e_b^{rms})_{\max}$ | $(e_b^{rms})_{\text{avg}}$ |
|----------------------|------------|------------|-------------------|-------------------------|------------------|----------------------|----------------------------|
| (1) Min max variance | 1.000 | 1.000 | 0.898 | 0.670 | 1.170 | 0.385 | 0.302 |
| (2) Min avg BE | 0.700 | 0.707 | 1.931 | 0.869 | 2.364 | 0.690 | 0.168 |
| (3) Min max BEB | 0.949 | 0.949 | 0.993 | 0.681 | 1.001 | 0.351 | 0.261 |
| (4) Min max RMS BE | 0.954 | 1.000 | 0.973 | 0.688 | 1.029 | 0.341 | 0.269 |

Table B-3. Comparison of actual RMS bias errors and predicted RMS bias errors for min-max RMS bias central composite experimental designs in four-dimensional space. Errors are computed on a uniform 11^4 grid.

| α_1 | α_2 | $(e_b^{rms})_{\max}$ | $(e_{act}^{rms})_{\max}$ | $(e_b^{rms})_{\text{avg}}$ | $(e_{act}^{rms})_{\text{avg}}$ | $r(e_b^{rms}, e_{act}^{rms})$ |
|------------|------------|----------------------|--------------------------|----------------------------|--------------------------------|-------------------------------|
| 1.000 | 0.100 | 1.155 | 1.158 | 0.927 | 0.927 | 1.000 |
| 1.000 | 1.000 | 1.176 | 1.180 | 0.827 | 0.827 | 1.000 |

APPENDIX C GLOBAL SENSITIVITY ANALYSIS

Global sensitivity analysis was first presented by Sobol in 1993. This method is used to estimate the effect of different variables on the total variability of the function. Some of the advantages of conducting a global sensitivity analysis include, (1) assessing importance of the variables, (2) fixing non-essential variables (which do not affect the variability of the function) thus, reducing the problem dimensionality. Homma and Saltelli (1996) (analytical functions and study of a chemical kinetics model), Saltelli et al. (1999) (analytical functions), Vaidyanathan et al. (2004b) (liquid rocket injector shape design), Jin et al. (2004) (piston shape design), Jacques et al. (2004) (flow parameters in a nuclear reactor), and Mack et al. (2005a) (bluff body shape optimization) presented some applications of the global sensitivity analysis. The theoretical formulation of the global sensitivity analysis is given as follows:

A function $f(\mathbf{x})$ of a square integrable objective as a function of a vector of independent uniformly distributed random input variables \mathbf{x} in domain $[0, 1]$ is assumed. The function can be decomposed as the sum of functions of increasing dimensionality as

$$f(\mathbf{x}) = f_0 + \sum_i f_i(x_i) + \sum_{i < j} f_{ij}(x_i, x_j) + \dots + f_{12\dots N_v}(x_1, x_2, \dots, x_{N_v}), \quad (C1)$$

where $f_0 = \int_{\mathbf{x}=0}^1 f d\mathbf{x}$. If the following condition

$$\int_0^1 f_{i_1\dots i_s} dx_k = 0, \quad (C2)$$

is imposed for $k = i_1, \dots, i_s$, then the decomposition described in Equation (C1) is unique. In context of global sensitivity analysis, the total variance denoted as $V(f)$ can be shown equal to

$$V(f) = \sum_{i=1}^{N_v} V_i + \sum_{1 \leq i, j \leq N_v} V_{ij} + \dots + V_{1\dots N_v}, \quad (C3)$$

where $V(f) = E((f - f_0)^2)$, and each of the terms in Equation (C3) represents the partial contribution or partial variance of the independent variables (V_i) or set of variables to the total variance, and provides an indication of their relative importance. The partial variances can be calculated using the following expressions:

$$\begin{aligned} V_i &= V(E[f | x_i]), \\ V_{ij} &= V(E[f | x_i, x_j]) - V_i - V_j, \\ V_{ijk} &= V(E[f | x_i, x_j, x_k]) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k, \end{aligned} \quad (C4)$$

and so on, where V and E denote variance and expected value respectively. Note that,

$$E[f | x_i] = \int_0^1 f_i dx_i \quad \text{and} \quad V(E[f | x_i]) = \int_0^1 f_i^2 dx_i.$$

This formulation facilitates the computation of the sensitivity indices corresponding to the independent variables and set of variables. For example, the first and second order sensitivity indices can be computed as

$$S_i = \frac{V_i}{V(f)}, \quad S_{ij} = \frac{V_{ij}}{V(f)}. \quad (C5)$$

Under the independent model inputs assumption, the sum of all the sensitivity indices is equal to one. The first order sensitivity index for a given variable represents the main effect of the variable, but it does not take into account the effect of the interaction of the variables. The total contribution of a variable to the total variance is given as the sum of all the interactions and the main effect of the variable. The total sensitivity index of a variable is then defined as

$$S_i^{total} = \frac{V_i + \sum_{j, j \neq i} V_{ij} + \sum_{j, j \neq i} \sum_{k, k \neq i} V_{ijk} + \dots}{V(f)}. \quad (C6)$$

Note that the above referenced expressions can be easily evaluated using surrogate models of the objective functions. Sobol (1993) has proposed a variance-based non-parametric approach to estimate the global sensitivity for any combination of design variables using Monte Carlo

methods. To calculate the total sensitivity of any design variable x_i , the design variable set is divided into two complementary subsets of x_i and Z ($Z = x_j, \forall j = 1, N_v; j \neq i$). The purpose of using these subsets is to isolate the influence of x_i from the influence of the remaining design variables included in Z . The total sensitivity index for x_i is then defined as

$$S_i^{total} = V_i^{total} / V(f), \quad (C7)$$

$$V_i^{total} = V_i + V_{i,Z}, \quad (C8)$$

where V_i is the partial variance of the objective with respect to x_i , and $V_{i,Z}$ is the measure of the objective variance that is dependent on interactions between x_i and Z . Similarly, the partial variance for Z can be defined as V_Z . Therefore the total objective variability can be written as

$$V = V_i + V_Z + V_{i,Z}. \quad (C9)$$

While Sobol had used Monte Carlo simulations to conduct the global sensitivity analysis, the expressions given above can be easily computed analytically if $f(\mathbf{x})$ can be represented in a closed form (e.g., polynomial response surface approximation).

APPENDIX D
LACK-OF-FIT TEST WITH NON-REPLICATE DATA FOR POLYNOMIAL RESPONSE
SURFACE APPROXIMATION

A standard lack-of-fit test is a statistical tool to determine the influence of bias error (order of polynomial) on the predictions (Myers and Montgomery, 1995). The test compares the estimated magnitudes of the error variance and the residuals unaccounted for by the fitted model. Lets say, we have M unique locations of the data and at j^{th} location, we repeat the experiment n_j times, such that total number of points used to construct surrogate model is $N_s = \sum_{j=1}^M n_j$. The sum of squares due to pure error is given by,

$$SS_{pe} = \sum_{j=1}^M \sum_{k=1}^{n_j} (y_{jk} - \bar{y}_j)^2, \quad (D1)$$

where \bar{y}_j is the mean response at the j^{th} sample location, given as, $\bar{y}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} y_k$.

The sum of square of residuals due to lack-of-fit of the polynomial response surface model is,

$$SS_{lof} = \sum_{j=1}^M n_j (\hat{y}(\mathbf{x}_j) - \bar{y}_j)^2, \quad (D2)$$

where $\hat{y}(\mathbf{x}_j)$ is the predicted response at the sampled location \mathbf{x}_j . In matrix form, the above expressions are given as,

$$SS_{pe} = \sum_{j=1}^M \mathbf{y}_{n_j}^T \left(I_{n_j} - \frac{\mathbf{1}_{n_j} \mathbf{1}_{n_j}^T}{n_j} \right) \mathbf{y}_{n_j}, \quad (D3)$$

$$SS_{lof} = \mathbf{y}^T \left(I_{N_s} - X(X^T X)^{-1} X^T \right) \mathbf{y} - \sum_{j=1}^M \mathbf{y}_{n_j}^T \left(I_{n_j} - \frac{\mathbf{1}_{n_j} \mathbf{1}_{n_j}^T}{n_j} \right) \mathbf{y}_{n_j}, \quad (D4)$$

where $\mathbf{1}_{n_j}$ is the $(n_j \times 1)$ vector of ones, I_{n_j} is $(n_j \times n_j)$ identity matrix, I_{N_s} is $(N_s \times N_s)$ identity matrix.

We formulate F-ratio using the two residual sum of squares as,

$$F = \frac{(SS_{lof}/d_{lof})}{(SS_{pe}/d_{pe})}, \quad (D5)$$

where $d_{lof} = N_s - N_\beta$, and $d_{pe} = N_s - M$, are the degrees of freedom associated with SS_{lof} and SS_{pe} , respectively. The lack-of-fit in the surrogate model is detected with α -level of significance, if the value of F in Equation (D5), exceeds the tabulated $F_{\alpha, d_{lof}, d_{pe}}$ value, where the latter quantity is the upper 100α percentile of the central F -distribution.

When the data is obtained from the numerical simulations, the replication of simulations does not provide an estimate of noise (SS_{pe}), since all replications return exactly the same value. In such scenario, the variance of noise can be estimated by treating the observations at neighboring designs as ‘near’ replicates (Hart, 1997, pp. 123). We adopt the method proposed by Neill and Johnson (1985), and Papila (2002) to estimate the lack-of-fit for non-replicate simulation. In this method, we denote a near-replicate design point \mathbf{x}_{jk} (as the k^{th} replicate of the j^{th} point $\bar{\mathbf{x}}_j$) such that

$$\mathbf{x}_{jk} = \bar{\mathbf{x}}_j + \boldsymbol{\delta}_{jk}, \quad (D6)$$

where $\boldsymbol{\delta}_{jk}$ represents the disturbance vector. Then, the Gramian matrix is written as

$$X = \bar{X} + \Delta, \quad (D7)$$

where \bar{X} matrix is constructed using $\bar{\mathbf{x}}_j$ for near replicate points and matrix $\Delta = X - \bar{X}$. Now the estimated response at the design points (including near-replicates) is given as,

$$\hat{\mathbf{y}} = \mathbf{y} - \Delta \mathbf{b}. \tag{D8}$$

where \mathbf{b} is the estimated coefficient vector. Now, we compute SS_{pe} and SS_{lof} by replacing \mathbf{y} , \mathbf{y}_{n_j} , and X in Equations (D1) and (D2) with $\hat{\mathbf{y}}$, $\hat{\mathbf{y}}_{n_j}$, and \bar{X} .

LIST OF REFERENCES

- Abramowitz M, Stegun IA, (Eds.), 1972, Integration, §25.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, pp 885-887.
- Ahuja V, Hosangadi A, Arunajatesan S, 2001, Simulations of cavitating flows using hybrid unstructured meshes. *Journal of Fluids Engineering*, **123(2)**:331-340.
- AIAA, 1994, Editorial policy statement on numerical accuracy and experimental uncertainty. *AIAA Journal*, **32**:3.
- AIAA, 1998, Guide for the verification and validation of computational fluid dynamics simulations. *AIAA G-077-1998*.
- Ariew R, 1976, Ockham's razor: A historical and philosophical analysis of Ockham's razor principle of parsimony. *Ph.D. Thesis*, Philosophy, The University of Illinois, Urbana-Champaign.
- ASME Editorial Board, 1994, Journal of heat transfer editorial policy statement on numerical accuracy. *ASME Journal of Heat Transfer*, **116**:797-798.
- Athavale MM, Singhal AK, July 2001, Numerical analysis of cavitating flows in rocket turbopump elements. *Proceedings of the 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Salt Lake City UT, AIAA-2001-3400.
- Balabanov VO, Kaufman M, Knill DL, Haim D, Golovidov O, Giunta AA, Grossman B, Mason WH, Watson LT, Haftka RT, September 1996, Dependence of optimal structural weight on aerodynamic shape for a high-speed civil transport. *Proceedings of the 6th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Bellevue WA, AIAA-96-4046:599-612.
- Balabanov VO, Giunta AA, Golovidov O, Grossman B, Mason WH, Watson LT, Haftka RT, 1999, Reasonable design space approach to response surface approximation. *Journal of Aircraft*, **36(1)**:308-315.
- Baker ML, Munson MJ, Duchow E, Hoppus GW, Alston KY, January 2004a, System level optimization in the integrated hypersonic aeromechanics tool (IHAT). *Proceedings of the 42nd AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2004-0618.
- Baker ML, Munson MJ, Hoppus GW, Alston KY, September 2004b, Weapon system optimization in the integrated hypersonic aeromechanics tool (IHAT). *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany NY, AIAA-2004-4316.
- Baldwin BS, Lomax H, January 1978, Thin-layer approximation and algebraic model for separated turbulent flow. *Proceedings of the 16th Aerospace Science Meeting*, Huntsville AL, AIAA-1978-257.

- Barthelemy J-FM, Haftka RT, 1993, Approximation concepts for optimum structural design – A review. *Structural Optimization*, **5**:129-144.
- Batchelor GK, 1967, *An Introduction to Fluid Dynamics*. Cambridge University Press, New York.
- Bishop C, 1995, *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.
- Box GEP, Draper N, 1959, A basis for the selection of a response surface design. *Journal of the American Statistical Association*, **54**:622-654.
- Box GEP, Draper NR, 1963, The choice of a second order rotatable design. *Biometrika*, **50(3)**:335–352.
- Bramantia A, Barbaa PD, Farinaa MB, Savinia A, 2001, Combining response surfaces and evolutionary strategies for multi-objective Pareto-optimization in electromagnetics. *International Journal of Applied Electromagnetics and Mechanics*, **15**:231-236.
- Brennen CE, 1994, *Hydrodynamics of Pumps*. Oxford University Press.
- Brennen CE, 1995, *Cavitation and Bubble Dynamics*. Oxford University Press.
- Burman J, Papila N, Shyy W, Gebart BR, September 2002, Assessment of response surface-based optimization techniques for unsteady flow around bluff bodies. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta GA, AIAA-2002-5596.
- Charania AC, Bradford JE, Olds JR, Graham M, April 2002, System level uncertainty assessment for collaborative RLV design. *Proceedings of the 38th Combustion, Airbreathing Propulsion, Propulsion Systems Hazards, and Modeling and Simulation Subcommittees Meeting*, (sponsored by JANNAF), Destin FL.
- Chankong V, Haimes YY, 1983, *Multiobjective Decision Making Theory and Methodology*. Elsevier Science, New York.
- Chung HS, Alonso JJ, September 2000, Comparison of approximation models with merit functions for design optimization. *Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach CA, AIAA-2000-4754.
- Cooper P, 1967, Analysis of single and two-phase flows in turbopump inducers, *ASME Journal of Engineering Power*, **89**:577-588.
- Craig KJ, Stander N, Dooge D, Varadappa S, September 2002, MDO of automotive vehicle for crashworthiness and NVH using response surface methods. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Atlanta GA, AIAA-2002-5607.

- Currin C, Mitchell TJ, Morris MD, Ylvisaker D, 1998, A Bayesian approach to the design and analysis of computer experiments. *Technical Report, Oak Ridge National Laboratory*, Oak Ridge TN, ORNL-6498.
- Deb K, 2001, *Multi-objective Optimization using Evolutionary Algorithms*. Wiley: Chichester UK.
- Delannoy Y, Reboud JL, 1993, Heat and mass transfer on a vapor cavity, *Proceedings of the ASME Fluids Engineering Conference*, Washington DC, **165**:209-214.
- Deshpande M, Feng J, Merkle CL, 1997, Numerical modeling of the thermodynamic effects of cavitation. *Journal of Fluids Engineering*, **119(2)**:420-427.
- Dixon LCW, Szegő GP, 1978, *Towards Global Optimization 2*, North-Holland, Amsterdam.
- Dornberger R, Büche D, Stoll P, September 2000, Multidisciplinary optimization in turbomachinery design. *Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering*, Barcelona.
- Dorney DJ, Rothermel J, Griffin LW, Thornton RJ, Forbes JC, Skelley SE, Huber FW, July 2006a, Design and analysis of a turbopump for a conceptual expander cycle upper-stage engine. *Proceedings of the Symposium of Advances in Numerical Modeling of Aerodynamics and Hydrodynamics in Turbomachinery*, Miami FL, FEDSM 2006-98101.
- Dorney DJ, Sondak DL, 2006b, PHANTOM: Program user's manual, Version 20.
- Dorney DJ, 2006c, NASA Marshall Space Flight Center, *Personal Communications*.
- Draper NR, Lawrence WE, 1965, Designs which minimize model inadequacies: Cuboidal regions of interest. *Biometrika*, **52(1-2)**:111-118.
- Draper NR, Smith H, 1998, *Applied Regression Analysis*. Third Edition, John Wiley & Sons Inc., New York.
- Efron B, 1983, Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, **78**:316-331.
- Efron B, Tibshirani R, 1993, *Introduction to the Bootstrap*. Chapman & Hall, New York.
- Emmerich MJ, Giotis A, Ozdemir M, Back T, Giannakoglou K, September 2002, Metamodel-assisted evolutionary strategies, *Proceedings of the Parallel Problem Solving from Nature VII Conference*, Granada Spain, pp 361-370.
- Fang H, Rais-Rohani M, Liu Z, Horstemeyer MF, 2005, A comparative study of meta-modeling methods for multi-objective crashworthiness optimization. *Computers and Structures*, **83**:2121-2136.

- Farina M, Sykulski JK, 2001, Comparative study of evolution strategies combined with approximation techniques for practical electromagnetic optimization problems. *IEEE Transactions on Magnetics*, **37(5)**:3216-3220.
- Fedorov VV, Montepiedra G, Nachtsheim CJ, 1999, Design of experiments for locally weighted regression. *Journal of Statistical Planning and Inference*, **81**:363-383.
- Franc JP, Rebattet C, Coulon A, 2003, An experimental investigation of thermal effects in a cavitating inducer, *Proceedings of the Fifth International Symposium on Cavitation*, Osaka, Japan.
- Friedman J, Stuetzle W, 1981, Projection pursuit regression. *JASA: Theory and Methods*, **76**:817-823.
- Garcia R, 2001, NASA Marshall Space Flight Center, *Personal Communications*.
- Gelder TF, Ruggeri RS, Moore RD, 1966, Cavitation similarity considerations based on measured pressure and temperature depressions in cavitated regions of Freon-114. *NASA Technical Note D-3509*.
- Giannakoglou KC, 2002, Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence. *Progress in Aerospace Sciences*, **38**:43-76.
- Gibbs M, 1997, Bayesian Gaussian Processes for Regression and Classification. *Ph.D. Thesis*, Cambridge University.
- Girosi F, 1998, An equivalence between sparse approximation and support vector machines. *Neural Computation*, **10(6)**:1455-1480.
- Giunta AA, Watson LT, September 1998, A comparison of approximation modeling techniques: Polynomial versus interpolating models. *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis MO, **1**:392-404, AIAA-98-4758.
- Goel T, Vaidyanathan RV, Haftka RT, Queipo NV, Shyy W, Tucker PK, September 2004, Response surface approximation of Pareto optimal front in multi-objective optimization. *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany NY, AIAA-2004-4501.
- Goel T, Mack Y, Haftka RT, Shyy W, Queipo NV, January 2005, Interaction between grid and design space refinement for bluff-body facilitated mixing. *Proceedings of the 43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2005-0125.
- Goel T, Haftka RT, Papila M, Shyy W, 2006a, Generalized bias error bounds for response surface approximation. *International Journal of Numerical Methods in Engineering*, **65(12)**:2035-2059.

- Goel T, Haftka RT, Shyy W, Queipo NV, 2006b, Ensemble of surrogates. *Structural and Multidisciplinary Optimization Journal*, doi: 10.1007/s00158-006-0051-9 (a different version was presented at 11th AIAA/ISSMO Multi-disciplinary Analysis and Optimization Conference, Portsmouth VA, 6-8 September 2006, AIAA-2006-7047).
- Goel T, Haftka RT, Shyy W, Watson LT, January 2006c, Pointwise RMS bias error estimates for design of experiments. *Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2006-0724.
- Goel T, Zhao J, Thakur SS, Haftka RT, Shyy W, July 2006d, Surrogate model-based strategy for cryogenic cavitation model validation and sensitivity evaluation. *Proceedings of the 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Sacramento CA, AIAA-2006-5047.
- Goel T, Haftka RT, Shyy W, Watson LT, July 2006e, Combining bias and variance based criteria for selecting experimental designs. *Proceedings of the NSF Design, Service, and Manufacturing Grantees and Research Conference*, St. Louis MO.
- Goel T, Vaidyanathan R, Haftka RT, Shyy W, Queipo NV, Tucker PK, 2007, Response surface approximation of Pareto optimal front in multi-objective optimization. *Computer Methods in Applied Mechanics and Engineering*, **196**:879-893.
- Goel T, Haftka RT, Shyy W, Watson LT, 2007, Pitfalls of using a single criterion for selecting experimental designs. *Submitted to International Journal of Numerical Methods in Engineering*.
- Golub G, Heath M, Wahba G, 1979, Generalized cross-validation as a method for choosing a Goodridge parameter. *Technometrics*, **21**:215-223.
- Griffin MD, French JR, 1991, *Space Vehicle Design*, AIAA, Washington DC.
- Gupta A, Ruffin SM, 2000, Aerothermodynamic performance enhancement of sphere-cones using the artificially blunted leading-edge concept. *Journal of Spacecraft and Rockets*, **37(2)**:235-241.
- Hall P, 1986, On the bootstrap and confidence intervals. *Annals of Statistics*, **14**:1431-1452.
- Hansen PC, 1992, Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, **34**:561-580.
- Hart JD, 1997, *Non-parametric Smoothing and Lack of Fit Tests*, Springer-Verlag, New York.
- Hedayat A, Sloane N, Stufken J, 1999, *Orthogonal Arrays: Theory and Applications*. Springer Series in Statistics, Springer Verlag, New York.
- Hesterberg T, Moore DS, Monaghan S, Clipson A, Epstein R, 2005, *Bootstrap Methods and Permutation Tests*, W H Freeman, New York, Chapter 14.

- Hill GA, Olson ED, September 2004, Application of response surface-based methods to noise analysis in the conceptual design of revolutionary aircraft. *Proceedings of the 10th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Albany NY, AIAA-2004-4437.
- Holl JW, Billet ML, Weir DS, 1975, Thermodynamic effects on developed cavitation, *Journal of Fluids Engineering*, **97(4)**:507-516.
- Homma T, Saltelli A, 1996, Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, **52(1)**:1-17.
- Hord J, 1973a, Cavitation in liquid cryogenics, II-Hydrofoil. *NASA CR-2156*.
- Hord J, 1973b, Cavitation in liquid cryogenics, III-Ogives. *NASA CR-2156*.
- Hosangadi A, Ahuja V, Ungewitter RJ, 2002, Simulations of cavitating inducer flow fields. *Proceedings of the 38th Combustion, Airbreathing Propulsion, Propulsion Systems Hazards, and Modeling and Simulation Subcommittees Meeting* (sponsored by JANNAF), Destin FL.
- Hosangadi A, Ahuja V, June 2003, A generalized multi-phase framework for modeling cavitation in cryogenic fluids. *Proceedings of the 33rd AIAA Fluid Dynamics Conference and Exhibit*, Orlando FL, AIAA-2003-4000.
- Hosangadi A, Ahuja V, Ungewitter RJ, July 2003, Generalized numerical framework for cavitation in inducers. *Proceedings of the 4th ASME/JSME Joint Fluids Engineering Conference*, Honolulu HI, FEDSM-2003-45408.
- Hosangadi A, Ahuja V, 2005, Numerical study of cavitation in cryogenic fluids. *Journal of Fluids Engineering*, **127(2)**:267-281.
- Hosder S, Watson LT, Grossman B, Mason WH, Kim H, Haftka RT, Cox SE, 2001, Polynomial response surface approximations for the multidisciplinary design optimization of a high speed civil transport. *Optimization and Engineering*, **2(4)**:431-452.
- Huber F, 2001, Turbine aerodynamic design tool development. In: *Space Transportation Fluids Workshop*, Marshall Space Flight Center AL.
- Humble RW, Henry GN, Larson WJ, 1995, *Space Propulsion Analysis and Design*. McGraw Hill Inc., New York, Chapter 5.
- Huque Z, Jahingir N, July 2002, Application of collaborative optimization on a RBCC inlet/ejector system. *Proceedings of the 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Indianapolis IN, AIAA-2002-3604.
- Hutchinson M, de Hoog F, 1985, Smoothing noisy data with spline functions. *Numerical Mathematics*, **47**:99-106.

- Huysse L, Padula SL, Lewis RM, Li W, 2002, Probabilistic approach to free-form airfoil shape optimization under uncertainty. *AIAA Journal*, **40(9)**:1764-1772.
- Iman R, Conover W, 1982, A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics, Part B-Simulation and Computation*, **11**:311-334.
- Jacques J, Lavergne C, Devictor N, March 2004, Sensitivity analysis in presence of model uncertainty and correlated inputs. *Proceedings of 4th International Conference on Sensitivity of Model Output (SAMO 2004)*, Santa Fe NM, pp 317-323.
- Jin R, Chen W, Simpson TW, 2001, Comparative studies of meta-modeling techniques under multiple modeling criteria. *Structural and Multi-disciplinary Optimization*, **23(1)**:1-13.
- Jin R, Chen W, Sudijanto A, March 2004, Analytical meta-model based global sensitivity analysis and uncertainty propagation for robust design. *Proceedings of the SAE World Congress and Exhibition*, Detroit MI, Paper 2004-01-0429.
- JMP®, *The statistical discovery software*TM, Version 5, Copyright © 1989-2002, SAS Institute Inc., Cary,NC, USA.
- Johnson M, Moore L, Ylvisaker D, 1990, Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, **26**:131-148.
- Jones D, Schonlau M, Welch W, 1998, Expensive global optimization of expensive black-box functions. *Journal of Global Optimization*, **13(4)**:455-492.
- Jun S, Jeon Y, Rho J, Lee D, September 2004, Application of collaborative optimization using response surface methodology to an aircraft wing design. *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany NY, AIAA-2004-4442.
- Kaufman M, Balabanov V, Burgee SL, Giunta AA, Grossman B, Haftka RT, Mason WH, Watson LT, 1996, Variable-complexity response surface approximations for wing structural weight in HSCT design. *Computational Mechanics*, **18(2)**:112-126.
- Keane AJ, 2003, Wing optimization using design of experiments, response surface and data fusion methods. *Journal of Aircraft*, **40(4)**:741-750.
- Khuri AI, Cornell JA, 1996, *Response Surfaces: Designs and Analyses*, 2nd edition, Marcel Dekker Inc., New York.
- Kim Y, Lee D, Kim Y, Yee K, September 2002, Multidisciplinary design optimization of supersonic fighter wing using response surface methodology. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Atlanta GA, AIAA-2002-5408.

- Kleijnen J, Deflandre D, 2004, Validation of regression meta-models in simulation: Bootstrap approach. *European Journal of Operational Research*, **170(1)**:120-131.
- Knill DL, Giunta AA, Baker CA, Grossman B, Mason WH, Haftka RT, Watson LT, 1999, Response surface models combining linear and Euler aerodynamics for supersonic transport design. *Journal of Aircraft*, **36(1)**:75-86.
- Kubota A, Kato H, Yamaguchi H, 1992, A new modeling of cavitating flows: A numerical study of unsteady cavitation on a hydrofoil section, *Journal of Fluid Mechanics*, **240**:59-96.
- Kunz RF, Boger DA, Stinebring DR, Chyczewski TS, Lindau JW, Gibeling HJ, 2000, A preconditioned Navier-Stokes method for two-phase flows with application to cavitation. *Computers and Fluids*, **29(8)**:849-875.
- Kupper LL, Meydrecht EF, 1973, A new approach to mean squared error estimation of response surfaces. *Biometrika*, **60(3)**:573-579.
- Laslett G, 1994, Kriging and splines: An empirical comparison of their predictive performance in some applications. *JASA: Applications and Case Studies*, **89**:391-400.
- Lauder BE, Spalding DB, 1974, The numerical computation of turbulent flows. *Computer Methods in Applied Mechanics and Engineering*, **3(2)**:269-289.
- Leary S, Bhaskar A, Keane A, 2003, Optimal orthogonal array-based Latin hypercubes. *Journal of Applied Statistics*, **30**:585-598.
- Lemmon EW, McLinden MO, Huber ML, 2002, *REFPROP: Reference Fluid Thermodynamic and Transport Properties*, NIST Standard Database 23, version 7.0.
- Lepsch RA Jr, Stanley DO, Unal R, 1995, Dual-fuel propulsion in single-stage advanced manned launch system vehicle. *Journal of Spacecraft and Rockets*, **32(3)**:417-425.
- Lertnuwat B, Sugiyama K, Matsumoto Y, 2001, Modeling of thermal behavior inside a bubble. *Proceedings of the Fourth International Symposium on Cavitation*, Pasadena, CA.
- Levy Y, Fan H-Y, Sherbaum V, 2005, A numerical investigation of mixing processes in a novel combustor application. *Journal of Heat Transfer*, **127(12)**:1334-1343.
- Li W, Padula S, May 2005, Approximation methods for conceptual design of complex systems. *Approximation Theory XI: Gatlinburg*, (eds. Chui C, Neaumtu M, Schumaker L), Nashboro Press, Brentwood, TN, pp 241-278 (also appeared in *Proceedings of the 11th International Conference on Approximation Theory*).
- Lophaven SN, Nielsen HB, Sondergaard J, 2002, DACE: A Matlab kriging toolbox, version 2.0, *Information and Mathematical Modeling*, Technical University of Denmark.

- Mack Y, Goel T, Shyy W, Haftka RT, Queipo NV, January 2005a, Multiple surrogates for shape optimization of bluff body-facilitated mixing. *Proceedings of the 43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2005-333.
- Mack Y, Goel T, Shyy W, Haftka RT, 2005b, Surrogate model based optimization framework: A case study in aerospace design. *Evolutionary Computation in Dynamic and Uncertain Environments* (eds. Yang S, Ong YS, Jin Y), Springer Kluwer Academic Press (in press).
- Mack Y, Shyy W, Haftka RT, Griffin LW, Snellgrove L, Huber F, July 2006, Radial turbine preliminary aerodynamic design optimization for expander cycle liquid rocket engine. *Proceedings of the 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Sacramento CA, AIAA-2006-5046.
- Madavan NK, Rai MM, Huber FW, 2001, Redesigning gas-generator turbines for improved unsteady aerodynamic performance using neural networks. *Journal of Propulsion and Power*, **17(3)**:669-677, (also presented at the *AIAA/ASME/SAE/ ASEE 35th Joint Propulsion Conference*, 20-24 June, 1999, Los Angeles CA, AIAA-99-2522).
- Madsen JI, Shyy W, Haftka RT, 2000, Response surface techniques for diffuser shape optimization. *AIAA Journal*, **38(9)**:1512-1518.
- Martin JD, 2005, A Methodology for Evaluating System-Level Uncertainty in the Conceptual Design of Complex Multidisciplinary Systems. *PhD Thesis*, The Pennsylvania State University, University Park PA.
- Martin JD, Simpson TW, 2005, Use of kriging models to approximate deterministic computer models. *AIAA Journal*, **43(4)**:853-863.
- Matheron G, 1963, Principles of geostatistics. *Economic Geology*, **58**:1246-1266.
- Matlab®, The language of technical computing, Version 6.5 Release 13. © 1984-2002, The MathWorks Inc.
- McDonald D, Grantham W, Tabor W, Murphy M, September 2000, Response surface model development for global/local optimization using radial basis functions. *Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. Long Beach CA, AIAA-2000-4776.
- McKay M, Conover W, Beckman R, 1979, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**:239-245.
- Mengistu T, Ghaly W, Mansour T, September 2006, Global and local shape aerodynamic optimization of turbine blades. *Proceedings of the 11th Multi-disciplinary Analysis and Optimization Conference*, Portsmouth VA, AIAA-2006-6933.

- Merkle CL, Feng J, Buelow PEO, April 1998, Computational modeling of dynamics of sheet cavitation. *Proceedings of the 3rd International Symposium on Cavitation*, Grenoble France.
- Miettinen KM, 1999, *Nonlinear Multiobjective Optimization*. Kluwer:Boston.
- Mitchell TJ, Morris MD, 1992, Bayesian design and analysis of computer experiments: Two examples. *Statistica Sinica*, **2(2)**: 359-379.
- Montepiedra G, Fedorov VV, 1997, Minimum bias design with constraints. *Journal of Statistical Planning and Inference*, **63(1)**:97-111.
- Morozov VA, 1984, *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag: Berlin.
- Myers RH, Montgomery DC, 1995, *Response Surface Methodology-Process and Product Optimization using Designed Experiments*, John Wiley & Sons Inc: New York.
- NASA online facts, 1991, <http://www-pao.ksc.nasa.gov/kscpao/nasafact/count2.htm> (last accessed March 28, 2007).
- Neill JW, Johnson DE, 1985, Testing linear regression function adequacy without replications. *The Annals of Statistics*, **13(4)**:1482-1489.
- Obayashi S, Sasaki D, Takeguchi Y, Hirose N, 2000, Multi-objective evolutionary computation for supersonic wing-shape optimization. *IEEE Transactions on Evolutionary Computation*, **4(2)**:182-187.
- Oberkampf WL, Trucano TG, Hirsch C, 2004, Verification, validation, and predictive capability in computational engineering and physics. *Applied Mechanics Review*, **57(5)**:345-384.
- Ohtani K, 2000, Bootstrapping R^2 and adjusted R^2 in regression analysis. *Economic Modelling*, **17(4)**:473-483.
- Ong YS, Nair PB, Keane AJ, 2003, Evolutionary optimization of computationally expensive problems via surrogate modeling. *AIAA Journal*, **41(4)**:687-696.
- Orr MJL, 1996, Introduction to radial basis function networks, *Center for Cognitive Science*, Edinburg University, EH 9LW, Scotland, UK. url: <http://www.anc.ed.ac.uk/~mjo/rbf.html> (last accessed March 28, 2007).
- Orr MJL, 1999a, Recent advances in radial basis function networks, *Technical Report*, *Institute for Adaptive and Neural Computation Division for Informatics*, Edinburg University, EH 9LW, Scotland, UK. url: <http://www.anc.ed.ac.uk/~mjo/rbf.html> (last accessed March 28, 2007).

- Orr MJL, 1999b, Matlab functions for radial basis function networks, *Technical Report, Institute for Adaptive and Neural Computation Division for Informatics*, Edinburg University, EH 9LW, Scotland, UK. url: <http://www.anc.ed.ac.uk/~mjo/rbf.html> (last accessed March 28, 2007).
- Owen AB, 1992, Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, **2(2)**:439-452.
- Owen AB, 1994, Controlling correlations in Latin hypercube samples. *Journal of the Statistical Association*, **89**:1517-1522.
- Palmer K, Tsui K. 2001, A minimum bias Latin hypercube design. *Institute of Industrial Engineers Transactions*, **33(9)**:793-808.
- Papila M, Haftka RT, April 1999, Uncertainty and wing structural weight approximations. *Proceedings of the 40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Material Conference*, St. Louis MO, AIAA-99-1312, pp 988-1002.
- Papila M, Haftka RT, 2000, Response surface approximations: Noise, error repair and modeling errors. *AIAA Journal*, **38(12)**:2336-2343.
- Papila M, Haftka RT, April 2001, Uncertainty and response surface approximations. *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Material Conference*, Seattle WA, AIAA-2001-1680.
- Papila M, 2002, Accuracy of Response Surface Approximations for Weight Equations Based on Structural Optimization, *Ph.D. Thesis*, The University of Florida, Gainesville FL.
- Papila M, Haftka RT, Watson LT, 2005, Pointwise bias error bounds and min-max design for response surface approximations. *AIAA Journal*, **43(8)**:1797-1807.
- Papila N, Shyy W, Griffin LW, Huber F, Tran K, July 2000, Preliminary design optimization for a supersonic turbine for rocket propulsion. *Proceedings of the 36th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Huntsville, Alabama, AIAA-2000-3242.
- Papila N, 2001, Neural Networks and Polynomial Response Surface Approximation Techniques for Supersonic Turbine Design Optimization, *Ph.D. Thesis*, The University of Florida, Gainesville FL.
- Papila N, Shyy W, Griffin LW, Dorney DJ, January 2001, Shape optimization of supersonic turbines using response surface and neural network methods. *Proceedings of the 39th Annual Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2001-1065.
- Papila N, Shyy W, Griffin LW, Dorney DJ, 2002, Shape optimization of supersonic turbines using global approximation methods. *Journal of Propulsion and Power*, **18(3)**:509-518.
- Patankar SV, 1980, *Numerical Heat Transfer and Fluid Flow*. Hemisphere: Washington, D.C.

- Perrone M, Cooper L. 1993, When networks disagree: Ensemble methods for hybrid neural networks. In: Mammone RJ. editor. *Artificial Neural Networks for Speech and Vision*. London: Chapman & Hall, pp 126-142.
- Poggio T, Smale S, 2003, The mathematics of learning: Dealing with data. *Notices of the American Mathematical Society*, **50**:537-544.
- Preston A, Colonius T, Brennen CE, 2001, Towards efficient computation of heat and mass transfer effects in the continuum model for bubbly cavitating flows. *Proceedings of the 4th International Symposium on Cavitation*, Pasadena CA.
- Qu X, Haftka RT, Venkataraman S, Johnson TF, 2003, Deterministic and reliability based optimization of composite laminates for cryogenic environments. *AIAA Journal*, **41(10)**:2029-2036.
- Qu X, Venter G, Haftka RT, 2004, New formulation of minimum-bias central composite experimental design and Gauss quadrature. *Structural and Multidisciplinary Optimization*, **28(4)**:231-242.
- Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK, 2005, Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, **41(1)**:1-28.
- Rachid FBF, 2003, A thermodynamically consistent model for cavitating flows of compressible fluids, *International Journal of Non-linear Mechanics*, **38**:1007-1018.
- Rai MM, Madavan NK, 2000, Aerodynamic design using neural networks. *AIAA Journal*, **38(1)**:173-182. (also presented at the 7th *AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, St. Louis MO, AIAA-1998-4928).
- Rai MM, Madavan NK, 2001, Application of artificial neural networks to the design of turbomachinery airfoils. *Journal of Propulsion and Power*, **17(1)**:176-183 (also presented at the 36th *Annual Aerospace Sciences Meeting and Exhibit*, Reno NV, Jan 12-15, 1998, AIAA-98-1003).
- Rais-Rohani M, Singh MN, September 2002, Efficient response surface approach for reliability estimation of composite structures. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Atlanta GA, AIAA-2002-5604.
- Rao CR, 1947, Factorial experiments derivable from combinatorial arrangement of arrays. *Supplement to the Journal of the Royal Statistical Society*, **9(1)**:128-139.
- Rao CR, 2002, *Linear Statistical Inference and Its Applications*. Second Edition, John Wiley and Sons, New York.
- Rapposelli E, Agostino LD, 2003, A barotropic cavitation model with thermodynamic effects, *Proceedings of the Fifth International Symposium on Cavitation*, Osaka, Japan.

- Reboud JL, Sauvage-Boutar E, Desclaux J, 1990, Partial cavitation model for cryogenic fluids. *ASME Cavitation and Multiphase Flow Forum*, Toronto, Canada.
- Redhe M, Forsberg J, Jansson T, Marklund PO, Nilsson L, 2002a, Using the response surface methodology and the D-optimality criterion in crashworthiness related problems: An analysis of the surface approximation error versus the number of function evaluations. *Structural and Multidisciplinary Optimization*, **24(3)**:185-194.
- Redhe M, Eng L, Nilsson L, September 2002b, Using space mapping and surrogate models to optimize vehicle crashworthiness design. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Atlanta GA, AIAA-2002-5536.
- Rikards R, Auzins J, September 2002, Response surface method in design optimization of carbon/epoxy stiffener shells. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Atlanta GA, AIAA-2002-5654.
- Roache PJ, Ghia K, White F, 1986, Editorial policy statement on the control of numerical accuracy. *ASME Journal of Fluids Engineering*, **108(1)**:2.
- Roache PJ, 1998, *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, New Mexico.
- Rogallo RS, Moin P, 1984, Numerical simulation of turbulent flows. *Annual Review of Fluid Mechanics*, **16**:99-137.
- Rowell LF, Braun RD, Olds JR, Unal R, 1999, Multidisciplinary conceptual design optimization of space transport systems. *Journal of Aircraft*, **36(1)**:218-226.
- Ruggeri RS, Moore RD, 1969, Method of prediction of pump cavitation performance for various liquids, liquid temperatures and rotation speeds. *NASA Technical Note, NASA TN D-5292*.
- Sacks J, Ylvisaker D, 1984, Some model robust designs in regression. *The Annals of Statistics*, **12**:1324-1348.
- Sacks J, Schiller S, Welch W, 1989, Designs for computer experiments. *Technometrics*, **31**:41-47.
- Sacks J, Welch WJ, Mitchell TJ, Wynn HP, 1993, Design and analysis of computer experiments. *Statistical Science*, **4**:409-435.
- Saltelli A, Tarantola S, Chan K, 1999, A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, **41(1)**:39-56.
- Samad A, Kim K-Y, Goel T, Haftka RT, Shyy W, July 2006, Shape optimization of turbomachinery blade using multiple surrogate models. *Proceedings of the Symposium of Advances in Numerical Modeling of Aerodynamics and Hydrodynamics in Turbomachinery*, Miami FL, FEDSM 2006-98368.

- Sanchez E, Pintos S, Queipo NV, July 2006, Toward and optimal ensemble of kernel-based approximations with engineering applications. *Proceedings of the IEEE World Conference on Computational Intelligence*, Vancouver BC, Canada, Paper ID 1265.
- Sasaki D, Obayashi S, Sawada K, Himeno R, September 2000, Multi-objective aerodynamic optimization of supersonic wings using Navier-Stokes equations, *Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, Barcelona Spain.
- Sasaki D, Morikawa M, Obayashi S, Nakahashi K, March 2001, Aerodynamic shape optimization of supersonic wings by adaptive range multi-objective genetic algorithms, *Proceedings of the 1st International Conference on Evolutionary Multi-Criterion Optimization*, Zurich, pp 639-652.
- Sasena M, Papalambros P, Goovaerts P, September 2000, Metamodeling sampling criteria in a global optimization framework. *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Optimization Conference*, Long Beach CA, AIAA-2000-4921.
- Sen P, Yang J-B, 1998, *Multiple Criteria Decision Support in Engineering Design*. Springer-Verlag: London.
- Senocak I, Shyy W, 2002, A pressure-based method for turbulent cavitating flow computations. *Journal of Computational Physics*, **176**:363-383.
- Senocak I, Shyy W, 2004a, Interfacial dynamics-based modeling of turbulent cavitating flows, Part-1: Model development and steady-state computations. *International Journal for Numerical Methods in Fluids*, **44(9)**:975-995.
- Senocak I, Shyy W, 2004b, Interfacial dynamics-based modeling of turbulent cavitating flows, Part-2: Time-dependant computations. *International Journal for Numerical Methods in Fluids*, **44(9)**:997-1016.
- Shyy W, 1994, *Computational Modeling for Fluid Flow and Interfacial Transport*. Elsevier: Amsterdam, The Netherlands (Revised print 1997).
- Shyy W, Thakur SS, 1994, A controlled variation scheme in a sequential solver for recirculating flows, Part-I: Theory and Formulation. *Numerical Heat Transfer B-Fundamentals*, **25(3)**:245-272.
- Shyy W, Thakur SS, Ouyang H, Liu J, Blosch E, 1997, *Computational Techniques for Complex Transport Phenomenon*. Cambridge University Press, United Kingdom.
- Shyy W, Tucker PK, Vaidyanathan R, 2001a, Response surface and neural network techniques for rocket engine injector optimization. *Journal of Propulsion and Power*, **17(2)**:391-401.

- Shyy W, Papila N, Vaidyanathan R, Tucker PK, 2001b, Global design optimization for aerodynamics and rocket propulsion components. *Progress in Aerospace Sciences*, **37**:59-118.
- Simpson TW, Peplinski JD, Koch PN, Allen JK, 2001a, Meta-models for computer based engineering design: Survey and recommendations. *Engineering with Computers*, **17(2)**:129-150.
- Simpson TW, Mauery TM, Korte JJ, Mistree F, 2001b, Kriging models for global approximation in simulation-based multidisciplinary design optimization. *AIAA Journal*, **39(12)**:2233-2241.
- Singhal AK, Vaidya N, Leonard AD, June 1997, Multi-dimensional simulation of cavitating flows using a PDF model for phase change. *ASME Fluids Engineering Division Summer Meeting*, Vancouver, Canada FEDSM97-3272.
- Singhal AK, Athavale MM, Li H, Jiang Y, 2002, Mathematical basis and validation of the full cavitation model. *Journal of Fluids Engineering*, **124(3)**:617-624.
- Snieder R, 1998, The role of non-linearity in inverse problems. *Inverse Problems*, **14(3)**:387-404.
- Sobieszczanski-Sobieski J, Haftka RT, 1997, Multidisciplinary aerospace design optimization: Survey of recent developments. *Structural Optimization*, **14**:1-23.
- Sobol IM, 1993, Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, **1(4)**:407-414.
- Sondak DL, Dorney DJ, July 2003, General equation set solver for compressible and incompressible turbomachinery flows, *Proceedings of the 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Huntsville AL, AIAA-2003-4420.
- Stahl HA, Stepanoff AJ, 1956, Thermodynamic aspects of cavitation in centrifugal pumps, *Transactions of ASME*, **78**:1691–1693.
- Stander N, Roux W, Giger M, Redhe M, Fedorova N, Haarhoff J, September 2004, A comparison of meta-modeling techniques for crashworthiness optimization. *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany NY, AIAA-2004-4489.
- Steffen CJ Jr., January 2002a, Response surface modeling of combined-cycle propulsion components using computational fluid dynamics. *Proceedings of the 40th AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2002-0542.
- Steffen CJ Jr., Bond RB, Edwards JR, April 2002b, Three dimensional CFD analysis of the GTX combustor. *Proceedings of the 38th Combustion, Airbreathing Propulsion, Propulsion Systems Hazards, and Modeling and Simulation Subcommittees Meeting* (sponsored by JANNAF), Destin FL.

- Stein M, 1987, Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, **29**:143-151.
- Stepanoff AJ, 1993, *Centrifugal and Axial Flow Pumps*, 2nd Edition, Krieger Publishing Company, Malabar FL.
- Steuer RE, 1986, *Multiple Criteria Optimization: Theory, Computation, and Application*. Wiley: New York.
- Tang B, 1993, Orthogonal array-based Latin hypercubes. *Journal of the American Statistical Association*, **88**:1392-1397.
- Tannehill JC, Anderson DA, Pletcher RH, 1997, *Computational Fluid Mechanics and Heat Transfer*. Taylor and Francis.
- Tenorio L, 2001, Statistical regularization of inverse problems. *SIAM Review*, **43**:347-366.
- Thakur SS, Wright JF, Shyy W, 2002, STREAM: A computational fluid dynamics and heat transfer Navier-Stokes solver. Streamline Numerics Inc. and Computational Thermo-Fluids Laboratory, Department of Mechanical and Aerospace Engineering Technical Report, Gainesville FL.
- Tikhonov AN, Arsenin VY, 1977, *Solutions to Ill-posed Problems*. Wiley, New York.
- Tokumasu T, Kamijo K, Matsumoto Y, 2002, A numerical study of thermodynamic effects of sheet cavitation. *Proceedings of the ASME FEDSM'02*, Montreal, Canada.
- Tokumasu T, Sekino Y, Kamijo K, 2003, A new modeling of sheet cavitation considering the thermodynamic effects. *Proceedings of the 5th International Symposium on Cavitation*, Osaka, Japan.
- Ueberhuber CW, 1997, *Numerical Computation 2: Methods, Software and Analysis*. Springer-Verlag: Berlin, pp 71.
- Umakant J, Sudhakar K, Mujumdar PM, Panneerselvam S, September 2004, Configuration design of air-breathing hypersonic vehicle using surrogate models. *Proceedings of 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany NY, AIAA-2004-4543.
- Utturkar Y, 2005, Computational Modeling of Thermodynamic Effects in Cryogenic Cavitation, *Ph.D. Thesis*, The University of Florida, Gainesville FL.
- Utturkar Y, Wu J, Wang G, Shyy W, 2005a, Recent progress in modeling of cryogenic cavitation for liquid rocket propulsion. *Progress in Aerospace Sciences*, **41(7)**:558-608.
- Utturkar Y, Thakur SS, Shyy W, January 2005b, Computational modeling of thermodynamic effects in cryogenic cavitation. *Proceedings of 43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno NV, AIAA-2005-1286.

- Vaidyanathan R, Papila N, Shyy W, Tucker PK, Griffin LW, Haftka RT, Fitz-Coy N, September 2000, Neural network and response surface methodology for rocket engine component optimization. *Proceedings of 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Long Beach CA, AIAA-2000-4880.
- Vaidyanathan R, Senocak I, Wu J, Shyy W, 2003, Sensitivity evaluation of a transport-based turbulent cavitation model. *Journal of Fluids Engineering*, **125(3)**:447-458.
- Vaidyanathan R, 2004, Investigation of Navier-Stokes Code Verification and Design Optimization, *Ph.D. Thesis*, The University of Florida, Gainesville FL.
- Vaidyanathan R, Tucker PK, Papila N, Shyy W, 2004a, CFD based design optimization for a single element rocket injector. *Journal of Propulsion and Power*, **20(4)**:705-717 (also presented in *41st Annual Aerospace Sciences Meeting and Exhibit*, Reno NV, January 2003, AIAA-2003-296).
- Vaidyanathan R, Goel T, Shyy W, Haftka RT, Queipo NV, Tucker PK, July 2004b, Global sensitivity and trade-off analyses for multi-objective liquid rocket injector design. *Proceedings of 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Ft. Lauderdale, FL, AIAA-2004-4007.
- Vapnik VN, 1998, *Statistical Learning Theory*. Wiley and Sons, New York.
- Venkateswaran S, Lindau JW, Kunz RF, Merkle C, 2002, Computation of multiphase mixture flows with compressibility effects. *Journal of Computational Physics*, **180(1)**:54-77.
- Venter G, Haftka RT, April 1997, Minimum-bias based experimental design for constructing response surface in structural optimization. *Proceedings of the 38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Kissimmee FL, Part 2: 1225-1238, AIAA-1997-1053.
- Ventikos Y, Tzabiras G, 2000, A numerical methods for simulation of steady and unsteady cavitating flows. *Computers and Fluids*, **29(1)**:63-88.
- Versteeg HK, Malalasekera W, 1995, *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*, Pearson Education Limited, England.
- Vittal S, Hajela P, September 2002, Confidence intervals for reliability estimated using response surface methods. *Proceedings of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization Conference*, Atlanta GA, AIAA-2002-5475.
- Wahba G, 1983, Bayesian confidence interval for the cross-validated smoothing spline. *Journal of the Royal Statistical Society B*, **45**:133-150.
- Wang TS, Chen YS, July 1990, A united Navier-Stokes flowfield and performance analysis of liquid rocket engines, *Proceedings of the 26th AIAA/SAE/ASME/ASEE Joint Propulsion Conference*, Orlando FL, AIAA 90-2494.

- Wang G, Senocak I, Shyy W, Ikohagi T, Cao S, 2001, Dynamics of attached turbulent cavitating flows. *Progress in Aerospace Sciences*, **37(6)**:551-581.
- Welch WJ, 1983, A mean squared error criterion for the design of experiments. *Biometrika*, **70(1)**:205-213.
- Wikipedia, 2007, http://en.wikipedia.org/wiki/Expander_cycle last accessed March 8, 2007.
- Williams B, Santner T, Notz W, 2000, Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica*, **10**:1133-1152.
- Williams CKI, Rasmussen CE, 1996, Gaussian processes for regression. *Advances in Neural Information Processing Systems 8*. Eds. Touretzky DS, Mozer MC, Hasselmo ME, MIT Press, pp 514-520.
- Wilson B, Capperelli D, Simpson TW, Frecker M, 2001, Efficient Pareto frontier exploration using surrogate approximations. *Optimization and Engineering*, **2(1)**:31-50.
- Wu J, 2005, Modeling of Turbulent Cavitation Dynamics in Fluid Machinery Flows, *Ph.D. Thesis*, The University of Florida, Gainesville FL.
- Wu J, Senocak I, Wang G, Wu Y, Shyy W, 2003a, Three-dimensional simulation of turbulent cavitating flows in a hollow-jet valve. *Computer Modeling in Engineering and Sciences*, **4(6)**:679-690.
- Wu J, Utturkar Y, Senocak I, Shyy W, Arakere N, June 2003b, Impact of turbulence and compressibility modeling on three-dimensional cavitating flow computations. *Proceedings of the 33rd AIAA Fluid Dynamics Conference and Exhibit*, Orlando FL, AIAA-2003-4264.
- Wu J, Utturkar Y, Shyy W, November 2003c, Assessment of modeling strategies for cavitating flow around a hydrofoil. *Proceedings of the Fifth International Symposium on Cavitation*, Osaka Japan, Cav03-OS-1-12.
- Wu J, Wang G, Shyy W, 2005, Time-dependent turbulent cavitating flow computations with interfacial transport and filter-based models. *International Journal of Numerical Methods in Fluids*, **49(7)**:739-761.
- Wu YT, Shin Y, Sues R, Cesare M, April 2001, Safety factor based approach for probability-based design optimization. *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference and Exhibit*, Seattle WA, AIAA-2001-1522.
- Yakowitz S, Szidarovsky F, 1985, A comparison of Kriging with non-parametric regression. *Journal of Multivariate Analysis*, **16**:21-53.
- Ye K, 1998, Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association*, **93**:1430-1439.

Ye K, Li W, Sudjianto A, 2000, Algorithmic construction of optimal symmetric Latin hypercube designs. *Journal of Statistical Planning and Inference*, **90**:145-159.

Zerpa L, Queipo NV, Pintos S, Salager J, 2005, An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates. *Journal of Petroleum Science and Engineering*, **47**:197-208 (also presented at *SPE/DOE 14th Symposium on Improved Oil Recovery*, 2004, April 17-21, Tulsa OK).

BIOGRAPHICAL SKETCH

Tushar Goel was born and brought up in the city of Sitapur in India. He completed his bachelor's education in mechanical engineering at the Institute of Engineering and Technology at Lucknow (India) in June 1999. He joined the Indian Institute of Technology at Kanpur for his graduate education. He worked with Prof. Kalyanmoy Deb in the area of evolutionary algorithms and received his master of technology in January 2001. Immediately after completing his master's education, he took an appointment with John F Welch Technology Center, research and development division of General Electric, as a mechanical engineer in the Advanced Mechanical Technologies group. He received his doctorate from the University of Florida at Gainesville in May 2007 under the tutelage of Prof. Raphael Haftka and Prof. Wei Shyy.

His research interests are design and optimization methods, surrogate modeling, and computational fluid dynamics.