UNIVERSITI PUTRA MALAYSIA

ROBUST OUTLIER DETECTION AND ESTIMATION IN RESPONSE SURFACE METHODOLOGY

MOHD SHAFIE MUSTAFA

IPM 2015 7

ROBUST OUTLIER DETECTION AND ESTIMATION IN RESPONSE SURFACE METHODOLOGY

By

MOHD SHAFIE MUSTAFA

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

October 2015

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

ROBUST OUTLIER DETECTION AND ESTIMATION IN RESPONSE SURFACE METHODOLOGY

By

MOHD SHAFIE MUSTAFA

October 2015

Chair: Professor Habshah Midi, PhD

Faculty: Institute for Mathematical Research

This thesis provides some extensions to the existing method of determining the optimization conditions in response surface design to cover situations with an unusual observations or outliers. It is shown how the presence of outliers have an unduly effect on the parameter estimation of response surface models and the optimum mean response. In real practice, the usual assumptions that the distribution of experimental data is approximately normal and constant variances are difficult to achieve.

The classical outlier diagnostic methods may not be suitable to correctly diagnose the existence of outliers in a data set. To rectify this problem, two procedures of robust diagnostic methods are proposed. In response surface optimization methodology, the parameters of the model are usually estimated using the Ordinary Least Squares (OLS) technique. Nevertheless, the classical OLS suffers a huge set back in the presence of outliers. In this situation, the optimum response estimator is not reliable. As an alternative, we propose using a robust MM-estimator to estimate the parameters of the RSM and subsequently the optimum mean response is determined. The results of the study reveal that our proposed method outperforms some of the existing methods.

This thesis also addresses the problems in the optimization of multiresponses, each of which depends upon a set of factors. The desirability function approach is commonly used in industry to tackle multiple response optimization problems. The shortcoming of this approach is that the variability in each predicted response is ignored. An augmented approach to the desirability function (AADF) is put forward to rectify this problem and to improve the practicality of the optimal solutions. Furthermore, the AADF can reduce the variation of predicted responses, as well as it is resistant to outliers.

In robust design studies, the usual assumptions of experimental data are approximately normal and there is no major contamination due to outliers in the data. In real practice, these two assumptions are difficult to meet. Hence we proposed Two-Stage Robust MM (TSR-MM based) method where it can remedy both problem of heteroscedasticity and outliers at the same time.

In order to make significant improvements in robust design studies, robust location (median) and robust scales estimates (Median Absolute Deviation (MAD) and Interquartile Range (IQR)) of the response variables are employed for dual response surface optimization. To get more efficient results, we proposed to adopt the robust MM estimator and the TSR-MM based method based on robust location and robust scales estimates when the problem of heteroscedastic errors comes together with outliers. The results of the study indicate that the robust location and scales estimates provide a significant reduction in the bias and variance of the estimated mean response.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENGESANAN TITIK TERPENCIL TEGUH DAN ANGGARAN DALAM METHODOLOGI PERMUKAAN SAMBUTAN

Oleh

MOHD SHAFIE MUSTAFA

Oktober 2015

Pengerusi: Professor Habshah Midi, PhD

Fakulti: Institut Penyelidikan Matematik

Tesis ini menyediakan beberapa sambungan kepada kaedah sedia ada bagi menentukan keadaan pengoptimuman reka bentuk permukaan tindak balas untuk menangani situasi pemerhatian yang luar biasa atau titik terpencil. Ia menunjukkan bagaimana kehadiran titik terpencil mempunyai kesan kelampauan pada anggaran parameter model permukaan sambutan dan min optimum sambutan. Secara realitinya, andaian yang biasa dalam eksperimen data bertaburan hampir normal dan varians malar adalah sukar untuk dicapai.

Kaedah tradisi dalam menentukan titik terpencil mungkin tidak sesuai untuk menentukan dengan tepat kewujudan data tersebut dalam sesuatu set data. Bagi mengatasi masalah ini, dua tatacara kaedah diagnosis teguh dicadangkan. Dalam kaedah pengoptimuman permukaan sambutan, lazimnya model parameter yang dianggarkan menggunakan kaedah biasa kuasa dua terkecil (OLS). Walaubagaimanapun, kaedah klasik OLS mengalami kelemahan dengan kehadiran titik terpencil. Dalam keadaan ini, penganggar sambutan optimum tidak lagi dipercayai. Sebagai alternatif, kami mencadangkan penggunaan penganggar teguh MM untuk menganggarkan parameter RSM dan seterusnya min optimum sambutan ditentukan. Keputusan berangka menunjukkan bahawa Optimum-MM adalah lebih cekap daripada Optimum-OLS. Keputusan kajian menunjukkan bahawa kaedah yang kami cadangkan mengatasi kaedah lain yang sedia ada.

Tesis ini juga menangani masalah pengoptimuman bagi sambutan berganda, setiapnya bergantung kepada satu set faktor. Pendekatan fungsi *desirability* biasa digunakan di dalam industri untuk menangani masalah pengoptimuman sambutan. Kelemahan pendekatan ini adalah variasi bagi setiap ramalan sambutan diabaikan. Pendekatan yang diperkukuhkan dengan Fungsi *desirability* (AADF) diperkenalkan untuk membetulkan masalah ini dan meningkatkan penyelesaian optimum yang praktikal. Selain itu, AADF dapat

mengurangkan variasi bagi setiap ramalan sambutan dan ianya tahan terhadap titik terpencil.

Dalam kajian reka bentuk teguh, kebiasaan andaian bagi eksperimen data adalah penghampiran normal dan tiada pencemaran utama disebabkan oleh titik terpencil di dalam data. Dalam keadaan sebenar, kedua-dua andaian tersebut adalah sukar untuk dipenuhi. Oleh itu kami mencadangkan kaedah Dua-Peringkat MM Teguh (TSR-MM berasas) di mana ianya pada masa yang sama boleh membetulkan kedua-dua masalah heterosedastik dan titik terpencil.

Dalam usaha untuk membuat penambahbaikan ketara dalam kajian reka bentuk teguh, anggaran lokasi teguh (median) dan skala teguh (Sisihan Median Mutlak (MAD) dan kuartil antara Range (IQR)) bagi pembolehubah sambutan digunakan untuk pengoptimuman permukaan dwi sambutan. Untuk mendapatkan hasil yang lebih berkesan, kami mencadangkan untuk menggunakan penganggar teguh MM dan kaedah berasaskan TSR-MM berdasarkan anggaran lokasi teguh dan skala teguh apabila masalah ralat berheterosedastik dengan titik terpencil muncul bersama. Keputusan kajian menunjukkan bahawa anggaran lokasi dan skala teguh memberikan pengurangan yang ketara dalam kepincangan dan anggaran varians bagi min sambutan.

ACKNOWLEDGEMENTS

In the name of Allah, The Most Merciful and The Most Compassionate. All praise is to Allah. With the blessings and Allah's guidance, I have completed my research and preparation of this PhD thesis.

I would like to express my utmost appreciation to the chairman of my supervisory committee, Professor Dr. Habshah Midi who consistently motivated and supported me with her insight, experience, and knowledge as well as for her invaluable guidance and advice during the course of my study. Under her supervision, I have learnt to be patient and optimistic in pursuing this study. I am equally appreciative of the advice extended to me by other members of supervisory committee, namely Assoc. Prof. Dr. Jayanthi Arasan and Dr. Md. Sohel Rana for their assistance and encouragement.

Acknowledgement is also due to all staff of the Institute of Mathematical Research (INSPEM) and the staff of the Department of Mathematics, UPM (academic and supporting staffs) for their help and kindness. Special thanks to Dr. Anwar Fitrianto for helping me understand the working and concept of response surface. I would like to express my sincere appreciation to all my friends, namely Luqman Hakim, Syarifah Nasrisya, Balqish Ariffin, Mohammad, Mazlina and Lim Hock Ann who had given me the moral encouragement and kind discussions during my study.

I also appreciate the financial support from the Ministry of Higher Education (MHE) provided throughout my study, via the SLAI programme.

Finally, my deepest gratitude goes to my parents (Mustafa Hj Mohd Salleh and Halimah Yaakob) whose love and prayers have made everything possible and have given me all the encouragement at all times, be it during good or difficult times. To my beloved wife, Fatimah Abdul Kadir and my daughter Nur Hafizah, both of you are my sources of motivation and inspiration for a better future. Thank you for the continuous support and prayers from all of you.

I certify that a Thesis Examination Committee has met on 20 October 2015 to conduct the final examination of Mohd Shafie Bin Mustafa on his thesis entitled "Robust Outlier Detection and Estimation In Response Surface Methodology" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Nik Mohd Asri Bin Nik Long, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohd Rizam Bin Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Noor Akma Binti Ibrahim, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

A. H. M. Rahmatullah Imon, PhD
Professor
Mathematical Sciences
Ball State University
Muncie, Indiana, USA
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Habshah Binti Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Jayanthi A/P Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Md. Sohel Rana, PhD
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: _____________

Name and Matric No.: Mohd Shafie Mustafa (GS23514)

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________________
Name of Chairman of Supervisory Committee: ______________________________

Signature: ______________________________
Name of Member of Supervisory Committee: ______________________________

Signature: ______________________________
Name of Member of Supervisory Committee: ______________________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

RSM	Response Surface Methodology
OLS	Ordinary Least Squares method
CCD	The Central Composite Design
FCD	The Face-Centered Design
AADF	Augmented Approach Desirability Function
LTB	Larger-The-Better
STB	Smaller-The-Better
NTB	Nominal-The-Best
BLUE	Best Linear Unbiased Estimators
r_{MM}	MM-Studentized Residuals
t_{MM}	MM-Deletion Studentized Residuals
r_{OLS}	OLS-Studentized Residuals
t_{OLS}	OLS-Deletion Studentized Residuals
Gt_i	Generalized Studentized Residuals
$MGRSt_i$	Modified Generalized Response Surface Studentized Residuals
WLS	Weighted Least Squares regression
IF	Influence Function
MSE	Mean Square Errors
LMS	Least Median of Squares
LTS	Least Trimmed Squares
IRLS	Iteratively Reweighted Least Squares
IWLS	Iterative Weighted Least Squares
ANOVA	The Analysis of Variance
SE	Standard Error
df	Degrees of Freedom
SS	Sum Square
MS	Mean Square
F	*F*-distribution
P	*p*-value
sd	Standard deviation
RLS	Reweighted Least Squares
TSR-MM	Two-Stage Robust MM based
MAD	Median Absolute Deviation
IQR	Inter-Quartile Range

CHAPTER 1

INTRODUCTION

1.1 Introduction and Background of the Study

Response Surface Methodology (RSM) was first developed and described by Box and Wilson in 1951 (Hill and Hunter, 1966). In a series of process optimization and experimental design, RSM consists of a group of mathematical and statistical techniques useful for modelling and analyzing a problem in which a response of interest is influenced by several variables. The main objective of RSM is to optimize the response and to find the combination of conditions that provides the highest response. RSM helps industrial world to realize how several input variables potentially influence some performance measures of a process and product. The relationship between a set of independent variables (also known as *control*, or *input* variables) and a response is determined by a mathematical model called regression model. Multiple regression analysis is one of the regression models useful for modelling and analyzing the relationship between a response and control variables required in RSM. In general, regression analysis is routinely applied in most applied sciences to observe the change in the response variable by changing any one of the control variables in the situation that the control variables are considered to be fixed. One of the predominant regression analysis techniques in RSM is Ordinary Least Squares Method (OLS). The popularity of OLS in industrial applications are due to its easy computation, universal acceptance, and elegant statistical properties. This method minimizes the errors sums of squares. Unfortunately, the OLS always depends on a number of restrictive and often unrealistic assumptions. Of all OLS assumptions, the normality of error distribution and the independency of explanatory variables are most common issues in linear regression (Montgomery et al., 2001, Psomas et al., 2007, Myers et al., 2009).

In applications, the normality of error distribution assumption will be inefficient in the presence of outlying observations in a data set resulting in less reliable estimates of the model parameters (Montgomery et al., 2001; Anderson, 2001; Kutner et al., 2004; Montgomery, 2009). The second assumption of OLS i.e. the implication of independency of explanatory variables can cause serious multicollinearity problems. This situation occurs when there are near-linear relationships between the explanatory variables which make up the columns of x.

Outliers can distort the regression results. When an outlier is included in the analysis, it pulls the regression line towards itself, which results the solution is more accurate for the outlier, but less accurate for the other cases in the data set. Outliers arise for many different reasons and appear in many different forms (Simpson and Montgomery, 1998). The existence of outliers include computational error, observation that is not part of the population being studied, result of keypunch errors or machine failure, or even transient effects. Not only the responses variable can be outlying, but also the explanatory part, leading to so called leverage points. Another issue in the presence of outliers is masking and swamping effects. There are many other discussion that described the presence of outliers in RSM (Myers et al., 1989, 2009; Morgenthaler and Martin, 1999; Montgomery et al., 2001; Park and Cho, 2003; Ching et al., 2005; Khuri and Mukhopadhyay, 2010).

1.2 Importance and Motivation of the Study

Outliers are often referred to the existence of a few anomalous points and empirical data set typically contains 10% outliers (Hampel et al., 1986). In multiple regression model, outliers may occur in *y*-direction (response direction) or outliers in the *x*-direction (regressor direction), which are also referred to as leverage points. Regardless of their sources, Simpson and Montgomery (1998) pointed out that the least squares estimation can be rendered useless by the presence of outliers. Unfortunately, many statisticians are not aware that outliers among the measurements will have a highly confusing effect and consequently leading to a wrong interpretation on response surface design. As such, outlier diagnostics are required to detect the existence of outlier in a data set. Many outlier diagnostics are based on residuals resulting from least squares method. However, in the presence of outlier, least squares estimator tries to accommodate the expense of the remaining observations. Therefore, an outlier may have small residuals, and consequently, diagnostics based on least squares residuals often fail to reveal such points. Myers et al. (2009) utilized the studentized residuals, R-students, and Cook's distance based on OLS estimates to detect outliers in response surface model. Unfortunately, to the best of our knowledge not much work has been devoted on detecting outliers in a second-order polynomial model for the response functions used in response surface methodology. This issue has motivated us to develop new diagnostics measures in RSM, namely, the studentized residuals, r_{MM}, and deletion studentized residuals, t_{MM} for the detection of outliers based on robust regression techniques. Since in RSM the *x* variables are fixed, our focus is only on the detection of outliers in the y-direction. The MM-estimator is incorporated in the establishment of almost all the developed methods in this thesis, as it is robust in *x* and *y* directions and has very high efficiency and high breakdown point. In this thesis, we also attempt to formulate another new outlier detection measures for detecting multiple outliers in response surface model. The developed method is called Modified Generalized Response Surface Studentized residuals ($MGRSt_i$). This work also has not been investigated, except for Imon (2005) who has developed such measure in multiple linear regressions model.

The first step in RSM is to fit a model between the controllable factors and the response variable. In response surface model particularly for one response variable, most of the estimation and regression analysis are generally constructed by the OLS method. However, it is well known that in the presence of outliers, the OLS may affect the optimization stage. The optimum conditions may be affected from the true optimum conditions we are actually looking for. Thus, a suitable approach needs to be developed. As such in this thesis, we employ a very high breakdown and high efficient robust MM estimator to estimate the model parameters. The advantage of using this measure is that, the robust regression techniques are not easily affected by outliers and subsequently will produce reliable optimum mean response.

This thesis also addresses the problem of optimization for multiresponse models in the presence of outliers. In many experimental situations, a number of responses need to be simultaneously optimized with respect to several criterions. Frequently, operating conditions need to satisfy several conditions or constraints on *m* responses. Outliers can wrongly show the optimum responses and are not reliable and may produce inefficient results. There are many good published articles in the literatures on the response surface (Myers and Montgomery, 1995; Morgenthaler and Martin, 1999; Park and Cho, 2003; Ching et al., 2005; Koksoy, 2008; Hejazi et al., 2010; Dellino et al., 2010). However, little attempts have been done in developing suitable robust methods for multiple response surface models in the existence of outliers in a data set.

The desirability function approach was introduced by Harrington (1965) and has been widely used extensively to simultaneously optimize several responses. The desirability function has generally been defined as aggregates multiple responses into a single dimensionless measure, so that a problem in the optimization of multiple responses is then converted into a single objective optimization problem. An overall, this desirability function technique assigns a set of responses and chooses factor settings that maximize the overall desirability function. However, an outstanding problem of this approach is that the variability in each predicted response is ignored. It is noted that the actual response sometimes may fall outside the acceptable region even though the predicted response at the optimal solution has a high overall desirability score (Fuller and Scherer, 1998, Chen et. al, 2012). Furthermore, Chen et al. (2012) stated that if the transformation into desirability does not cover the prediction interval, the optimal solution will not be acceptable for practical implementation. Chen et al. (2012) developed Augmented Desirability Function Approach (AADF-OLS based) to determine the factors settings and optimum mean response. Nonetheless their approaches are based on OLS estimator and Geometric Mean which are very sensitive to outliers. Their work has encouraged us to develop Augmented Approach Desirability Function (AADF-MM based) which is based on the MM estimator and Geometric Median which are outlier resistant.

Response surface methodology is designed to construct an approximation model for the response *y*. This approximation model is usually the second-order polynomial model to be fitted between the response variable (quality characteristics) and a number of input variables. The main aim is to find the best optimal settings of interest for the input variables or the best values of design parameters that optimize the response variable. Typically the main emphasis is on optimizing (minimizes or maximizes) the mean (location) value of *y* where the variance (scale) is assumed to be small and constant. These assumptions may not be valid in real-life practice. Nonetheless, only constructing a response surface model for the mean may not be adequate and optimization result can be misleading. Therefore, the dual response approach (developed by Myers and Carter, 1973) is used to tackle such problem (see Myers and Carter, 1973; Vining and Myers, 1990; Del Castillo, 1999; Park and Cho, 2003; Shaibu and Cho, 2009). Basically in dual response surface optimization, two models are established for the mean and for the standard deviation of the response *y*. Then the two fitted response models are optimized simultaneously in a region of interest. The experiments are repeated *m* times to measure the variability of *y*.

The OLS method is often used to estimate the parameters of the models. It is important to mention that the OLS regression estimates which are often used in RSM are also not appropriate for real-world industrial problems containing outliers. The problems get more complicated when outliers and heteroscedastic errors come together. Goethals and Cho (2011) employed the iterative reweighted least squares approach (RLS) method to estimate the model parameters when the assumptions of constant error variances are violated. Their work did not investigate the effect of outliers on the parameter estimates and consequently the mean optimal response will be affected. There is a strong evidence that the RLS is not reliable in this situation. The weakness of this estimator has inspired us to develop a new method that can rectify both problems simultaneously are call this method, the TSR-MM based method to estimate the parameters of the process mean model and the process standard deviation model in the dual response problems. Since the two fitted response surface models use the TSR-MM based method, we anticipated to get a more reliable estimated mean response.

In the classical dual response approach problem, the sample mean and the sample variance are used to fit the process mean and process variance functions based on the OLS method. However, these estimators are very sensitive to outliers or departures from the normality assumption (Lee et al., 2007). As a consequence, the optimum operating conditions may be located far from the true optimum values. Since contaminated data (or outlier) may reveal misleading results on sample mean and variance, Park and Cho (2003) proposed using sample median instead of the sample mean and the sample median absolute deviation (MAD) (and interquartile range (IQR)) instead of the sample variance of the responses. The results show that the new measures are less sensitive to contamination and departures from the normality assumption. However, they utilised the OLS estimator which is known to be sensitive to outliers to estimate the model parameters. The weakness of the Park and Cho (2003) approach has encouraged us to employ the MM estimator to estimate the parameters of the dual response surface models. The results of the study indicate that when using median and MAD of the response variables, it give the best optimal setting for the input variables. Due to the encouraging results of using robust location and scale, we investigate the performance of our developed TSR-MM based method using these measures. To the best of our knowledge this issue has not be explored.

1.3 Research Objectives

The foremost objectives of our research can be outlined systematically as follows.

1. To formulate an outlier detection measure for response surface model, polynomial regression model (single response *y*) by integrating the MM estimator in the studentized residuals, r_{MM}, deletion studentized residuals, t_{MM}, and modified generalized response surface studentized residulas, $MGRSt_i$.

2. To develop a new outlier detection measure $(MGRSt_i)$ for response surface model (single response *y*) to identify multiple outliers.

3. To employ the robust MM estimator to estimate the parameters of the response surface model for a single response variable and subsequently compare the estimated mean response based on OLS (Optimum-OLS based) and Optimum-MM based in the presence of outliers.

4. To develop a new augmented approach (AADF-MM based) to the desirability function based on MM estimator and geometric median for multiple responses.

5. To develop a new robust estimator (TSR-MM based) in the response surface design for repeated responses with heteroscedastic conditions.

6. To employ robust locations and robust scales measures and MM estimator and TSR-MM based estimator in dual response optimization approach for homoscedastic and heteroscedastic conditions.

1.4 Overview of the Thesis

In accordance with the objectives and the scope of the study, the contents of this thesis are organized in the eight chapters. The thesis chapters are structured so that the research objectives are apparent and are conducted in the sequence outlined.

Chapter Two: This chapter presents a literature review on response surface methodology (RSM) and its experimental design, modelling, and optimization techniques are highlighted. The methods developed in RSM to cope with multiresponses are also discussed. In the second part of the chapter, the OLS estimation of regression parameters and violations from its assumptions are described briefly in order to estimate the parameters of a second-order polynomial RSM model. Diagnostic methods of influential observations and outlier diagnostics are also reviewed. Basic concepts of robust regression and some important existing robust methods are also included. Finally, heteroscedasticity methods are reviewed briefly.

Chapter Three: This chapter presents several techniques for the identification of multiple high leverage points in response surface design (RSM) model and outlier diagnostics is defined following the basic ideas of linear regression diagnostics. The two procedures of detection of outliers are proposed in RSM. The first procedure is based on residuals of the MM estimates and the second procedure incorporated the $MGRSt_i$ MM-estimator. The performance of $MGRSt_i$ is also evaluated for real data set in detecting outliers or contaminated points and for simulation studies. Moreover, the regression diagnostic plots which are useful in detecting outlier points are also shown in this chapter.

Chapter Four: This chapter discusses the situations in optimum response when outliers (or without outliers) are present in the real data set. The existing analysis and optimization method in RSM which cause points to be contaminated are also investigated. The effect of outliers of experimental designs for fitting response surface models, optimization method and the performance of response surface in our proposed measures, namely MM-estimation are studied. The performance of the Optimum-MM is compared with the existing method, Optimum-OLS.

Chapter Five: The augmented desirability function approach is proposed to tackle multiple responses optimization problems which are discussed by Harrington (1965) and Derringer and Suich (1980). The newly proposed augmented desirability function (AADF) incorporated variability in each predicted responses and combined an overall desirability function using median. The method is formulated by adapting the MM estimator and logarithm of median. The effect of outliers on the AADF performance of optimization is investigated.

Chapter Six: This chapter involves two situations in regression parameters where heteroscedasticity errors come together with the existence of outliers in the data. The proposed method is called the Two-Stage Robust estimator based on MM-estimator (TSR-MM based) which can handle both the outliers and heteroscedasticity problem. The new proposed robust method is compared with some existing methods such as reweighted least squares based on OLS denoted as RLS. The comparison results of the performance of optimization based on our proposed method and classical method is discussed.

Chapter Seven: Incorporating the outlier-resistant estimators into robust design, namely the median and MAD or IQR is proposed in this chapter. The newly proposed estimators, TSR-MM based and MM-based estimator which is less sensitive to outliers in dual-response surface model employ the locations and scales measures.

Chapter Eight: This chapter provides summary and detailed discussions of the thesis conclusions. Areas for future research are also recommended.

REFERENCES

Ariff, R. M., Fitrianto, A., A. Manap, M. Y., Ideris, A., Kassim, A., Suhairin, A., Anis., S. B. M. (2013). Cultivation Conditions for Phytase Production from Recombinat Escherichia coli DH5α. *Microbiology Insights*. 6:17-28.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*.16:523-531.

Andersen, R. (2008). *Modern methods for robust regression*. The United States of America: Sara Miller McCune. SAGE publications.

Anderson, C. (2001). *A Comparison of Five Robust Regression Methods with Ordinary Least Squares: Relative Efficiency, Bias, and Test of the Null Hypothesis*, Unpublished Ph.D. thesis, University of North Texas, U.S.A.

Atkinson, A.C. (1981). Two graphical displays for outlying and influential observations in regression, Biometrika, 68(1):13-20.

Atkinson, A.C. (1982). Regression Diagnostics, Transformations and Constructed Variables, Journal of Royal Statistical Society, B 44(1):1-36.

Atkinson, A.C. (1983). Masking unmasked, Biometrika, 73(3):533-541.

Anscombe, F. J. and Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics*, 5, 141-60.

Ames A., Mattucci N., McDonald S.,Szonyi G., Hawkin D. (1997). Quality Loss Function for Optimization Across Multiple Response Surfaces. Journal Quality Technology, 29:339-346.

Armstrong, R. D. and Kung, M. T. (1978). Least absolute values estimates for a simple linear regression problem. *Applied Statistics*. 27: 363-366.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistcal Association*. 70:428-434.

Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*.16: 147-185.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York :Wiley.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*. 40: 318-35.

Box, G. E. P., and Behnken, D. W. (1960). Some New Three-Level Designs for the Study of Quantitative Variables. Technometrics, 2, 455-475.

Box, G. E. P. And Draper N. R. (1975). Robust Design. *Biometrika*, 62,2, p. 347.

Box, G. E. P. And Draper N. R. (1987). Empirical Model-Building and ResponseSurfaces. New York: John Wiley and Sons.

Box, G. E. P. And Draper N. R. (2007). Response Surfaces, Mixtures, and Ridge Analyses. 2nd Edition. Hoboken, New Jersey: John Wiley and Sons.

Box, G. E. P., Hunter, J. S. (1957). Multifactor Experimental Designs for Exploring Response Surfaces. The Annals of Mathematical Statistics, 28, 195-241.

Box, G. E. P., Wilson (1951). On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society, Ser. B, 13, 1-45.

Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*. NewYork:Wiley.

Chatterjee, S. and Hadi, A.S. (2006). *Regression Analysis by Example*. 4th edition. New York: Wiley.

Ching,C. K., Quah, S. H., and Low H. C. (2005). The MM-Estimator in Response Surface Methodology. *Quality Engineering*, 17: 561-565.

Chen, Wong, Xu (2012). An augmented approach to the desirability function. *Journal of Applied Statistics*, Vol.39, No. 3, 599-613.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*. 19:15-18.

Cook, R. D. (1979). Influential observations in linear regression. *Journal of American Statistical Association*. 74:169-174.

Cook, R. D., and Weisberg, S. (1983). Diagnostics for Heteroscedasticity in Regression. Biometrika, 70, pp. 1-10.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* London: Champan Hall.

Croux, C., Rousseeuw, P. J. and Hössjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association.* 89:1271–1281

Del Castillo E. (1996). Multiresponse Process Optimization via Constrained Confidence Regions. Journal Quality Technology, 26:61-70.

Del Castillo E., Fan S. K., Semple, J. (1999). Optimization of Dual Response Systems: A comprehensive Procedure for Degenerate and Nondegenerate Problems. *European journal Operation Research*, 112:174-186.

Dellino, G., Kleijnen, J. P. C., Meloni, C. (2010). Robust Optimization In Simulation: Taguchi and Response Surface Methodology. *International Journal Production Economics*, 125, 52-59.

Derringer G., Suich R. (1980). Simultaneous Optimization of Several Response Variables. *Journal of Quality Technology*, Vol. 12, No. 4, 214-219.

Draper, N. R. (1963). Ridge Analysis of Response Surfaces. Technometrics, 5, 469-479.

Edgeworth, F. Y. (1887). On observations relating to several quantities. *Hermathena*. 6:279-285.

Fang J. and He Z. (2010). Analysis of Response Surface Design to Outlier. 2010 International Conference on E-Business and E-government. IEEE International Conference.

Fuller, D., & Scherer, W. (1998). The desirability function: underlying assumptions and application implications. *Paper presented at the Systems, Man, and Cybernetics, 1998*. 1998 IEEE International Conference.

Fitrianto, A., Midi, H. (2010). Estimating Bias and RMSE of Indirect Effects Using Rescaled Residual Bootstrap in Mediation Analysis. *WSEAS TRANSACTIONS on MATHEMATICS*, 9(6), pp. 397-406.

Goethals, P. L. and Cho, B. R. (2011). Solving the Optimal Process Target Problem Using Response Surface Designs in Heteroscedastic Conditions. *International Journal of Production Research*, Vol. 49, No. 12, 3455-3478.

Greene, W. H. (2008). *Econometric analysis*. 6th edition. Upper Saddle River. New Jersey: Prentice Hall.

Habshah, M., Norazan, M.R. and Imon, A.H.M.R. (2009). The performance of Diagnostic-Robust Generalized Potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics.* 36(5): 507-520.

Habshah, M., Mohd, S. M., Anwar, F. (2012). The Performance of Optimum Response Surface Methodology Based on MM-Estimator. *International Journal of Mathematical Models and Methods in Applied Sciences*. Issue 6, Vol. 6, 757-764.

Hejazi, T. H., Bashiri, M., Noghondarian, K., & Atkinson, A. C. (2010). Multiresponse Optimization with Consideration of Probabilistics

Covariates. Quality and Reliability Engineering International, DOI: 10.1002/qre.1133.

Hill, W. J. and Hunter, W. G. (1966). A Review of Response Surface Methodology: A Literature Review. Technometrics, 8: 571-590.

Huat, N. K and Midi, H. (2010). Robust Individuals Control Chart for Change Point Model. *WSEAS TRANSACTIONS on MATHEMATICS*, 9(7), pp.499-508.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. *Computational and Statistical Data Analysis*. 14:1-27.

Hampel F.R., Ronchetti E. M., Rousseeuw P. J., Stahel W.A. (1986). Robust Statistics: The Approach Based on Influence Function. New York: 1986. John Wiley & Sons, Inc.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 69: 383-393.

Harrington (1965). The Desirability Function. *Industrial Quality Control*, 12: 494-498.

Gonda N. H., Oortmarssen G. J., Piersma, N. R. Dekker (2000). A framework for response surface methodology for simulation optimization. *Proceedings of the 2000 Winter Simulation Conference*. pp. 129-136.

Hoerl, A. E. (1959). Optimum Solutions of Many Variable Equations. *Chemical Engineering Progress*, 55:69-78.

Hoaglin, D. C and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *American Statistician.* 32:17-22.

Huber, P. J. (1964). Robust estimation of location parameters. *Annals of Mathematical Statistics.* 35:73–101.

Huber, P. J. (1973). Robust regression: asymptotic, conjectures, and Monte Carlo. *The Annals of Statistics*. 1: 799-821.

Huber, P.J. (1981). Robust statistics. Wiley:New York.

Hubert, M., Rousseeuw, P. J., and Verboven, S. (2002). A Fast Method for Robust Principal Components With Applications to Chemometrics. *Chemometrics and Intelligent Laboratory Systems*. 60:101–111.

Hu, Z., Cai, M., Liang H.-H. (2008). Desirability function approach for the optimization of microwave-assisted extraction of saikosaponins from Radix Bupleuri, *Sep. Purif. Technol*. 61(3), pp. 266–275.

Imon A. H. M. R. (2004). Weighted Mean Matrix on Weight Sequence Spaces. *WSEAS Transactions on Mathematics*, 3(4), pp. 789-793.

Imon A. H. M. R. (2005). Identifying Multiple Influential Observations in Linear Regression. Journal of Applied Statistics. 32:9, 929-946.

Imon A. H. M. R. and Hadi A. S (2008). Identification of Multiple Outliers in Logistics Regression. *Journal Communications in Statistics- Theory and Methods*., 37[11]: 1697-1709.

Kackar R. N. (1985). Off-Line Quality Control, Parameter Design, and the Taguchi method. *Journal Quality Technology*, 17:176-209.

Khuri, A. I. and Cornell, J. A. (1996). Response Surface, 2nd Edition. New York:Dekker.

Khuri , A. I. And Conlon, M. (1981). Simultaneous Optimzation of Multiple Response Represented by Polynomial Regression Functions. *Technometric*, 23, 363-375.

Khuri, A. I., Mukhopadhyay, S. (2010). Response Surface Methodology. *Advance Review*, Vol. 2, March/April 2010. John Wily & Sons, Inc.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2004). *Applied Linear Regression Models*. 5th edition. New York: MacGRAW-Hill.

Koksoy, O. (2008). A Nonlinear Programming Solution to Robust Multi-response Quality Problem. *Applied Mathematics and Computation*, 196, 603-612.

Lee, S. B. et al. (2007). Development of a High Efficient and Resistant Robust Design. *International Journal of Production Research*, Vol. 45, No. 1, 157-167.

Lind, E. E., Goldin, J., and Hickman, J. B. (1960). Fitting Yield and Cost Response Surfaces. *Chemical Engineering Progress*, 56, 62-68.

Lin D. K. J., Tu W. (1995). Dual Response Surface Optimization. Journal of Quality Technology, Vol. 27, No.1.

Mead , R., and Pike, D. J. (1975). A Review of Response Surface Methodology From a Biometric Viewpoint. Biometrics, 31, 803-851.

Midi, H. (1999). Preliminary Estimators for Robust Non-Linear Regression Estimation. Journal of Applied Statistics, 26(5):591-600.

Midi, H., Rana, S., Imon, A.H.M.R (2009a). The Performance of Robust Weighted Least Squares in Presence of Outliers and Heteroscedastic Errors. *WSEAS TRANSACTIONS on MATHEMATICS*, 8(7), pp. 351-361.

Midi, H., Norazan, M. R., Imon, A.H.M.R (2009b). The Performance of Diagnostics-Robust Generalized Potential for The Identification of

Multiple High Leverage Points in Linear Regression. Journal of Applied Statistics, 36(5):507-520.

Maronna, R. A., Martin, R. D., Yohai, V. J. (2006). Robust Statistics: Theory and Methods. John Wiley and Sons.

Montgomery, D. C. (2005). Design and Analysis of Experiments. 6th Edition. John Wiley and Sons, Inc.

Montgomery, D. C. (2009). Design and Analysis of Experiments. 7th Edition. John Wiley and Sons, Inc.

Montgomery, D. C., Peck, E. A., and Vining, G. G (2001). Introduction to Linear Regression Analysis. 3rd Edition. John Wiley and Sons, Inc.

Morgenthaler S., Martin, M. S. (1999). Robust analysis of a response surface design. *Chemometrics and Intelligent Laboratory System 47*, pp. 127-141.

Myers, R. H., and Carter, W. H. Jr. (1973). Response Surface Techniques for Dual Response Systems. *Technometric*, 15, 301-317.

Myers, R. H., Khuri I. A., Carter H. W. (1989). Response Surface Methodology. *Technometrics*, Vol. 31, pp. 137-157.

Myers, R. H., Khuri A. I., Carter, W. H. Jr. (1989). Response Surface Methodology: 1966-1988. *Technometrics*, Vol. 31, No. 2, pp. 137-157.

Myers, R. H., Khuri, A. I., Vining, G. G. (1992). Response Surface Alternative to the Taguchi Robust Design Problem. Am. Stat., 46, 131-139.

Myers, R. H., Montgomery, D. C. (1995). Response Surface Methodology. New York: John Wiley & Sons.

Myers, R. H., Montgomery, D. C. (2002). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 2nd Edition. New York: John Wiley and & Sons.

Myers, R. H., Montgomery, D. C., Anderson-Cook, C. M. (2009). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 2nd Edition. Canada. John Wiley and Sons, Inc.

Norazan, M.R. (2008). *Weighted Maximum Median Likelihood Estimation for Parameters in Multiple Linear Regression Model*, Unpublished Ph. D. Thesis, University Universiti Putra Malaysia, Malaysia.

Oehlert, G. W. (2000). Design and Analysis of Experiments: Response Surface Design. New York: W. H. Freeman and Company.

Palaniyappan, M., Vijayagopal, V., Viswanathan, R., Viruthagiri, T. (2009). Statistical Optimization of Substrate, Carbon and Nitrogen Source by

Response Surface Methodology for Pectinase Production Using Aspergillus Fumigatus MTC870 in Submerged Fermation. *African Journal of Biotecnology*, Vol. 8, (22), pp. 6355-6363.

Park, C. And Cho, B. R. (2003). Development of Robust Design Under Contaminated And Non-normal Data, Quality Engineering, 15:3, 463-469.

Pearson, E. S. **(**1931). The analysis of variance in cases of non-normal variation. *Biometrika*. 23:114-133.

Pignatiello J. J. (1993). Strategies for Robust Multiresponse Quality Engineering. IIE Trans, 25:5-15.

Psomas, S. K., Liakopoulou-Kyriakides, M., and Kyriakidis, D.A. (2007). Optimization Study of Xanthan Gum Production Using Response Surface Methodology. *Biochemical Engineering Journal*, Vol.35, pp. 273-280.

Ramli, N. M., Midi, H., Imon, A.H.M.R. (2009). Estimating Regression Coefficients using Weighted Bootstrap with Probability. *WSEAS TRANSACTIONS on MATHEMATICS*, 8(7), pp.362-371.

Ramsay, J.O.(1977). A comparative study of several robust estimates of slope, intercept, and scale in linear regression, *Journal of American Statistical Associations.*72:608-615.

Rao, C. R. (1965). Linear Statistical Inference. New York: Wiley.

Riazoshams, H., Midi, H., Sharipov. O. S. (2010). The Performance of Robust Two-Stage Estimator in Nonlinear Regression with Autocorrelated Error. *Communication in Statistics-Simulation and Computation*, 39: 1251-1268.

Rousseeuw, R. J. & Leroy, A. M. (2003). Robust Regression and Outlier Detection. John Wiley and Sons, Inc.

Rousseeuw, P. J. (1983). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*. Vol (B): 283-297.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of theAmerican Statistical Association.* 79: 871–880.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. *Mathematical and Statistical Applications*. B: 283-297.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute values. *Journal of the American Statistical Association*. 88: 1273-83.

Rousseeuw P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics.* 41:212–223.

Rousseeuw, P. and Van Zomeren, B. (1990).Unmasking multivariate outliers and leverage points. *Journal of American Statistical Associations*. 85: 633-639.

Rousseeuw, P. J. and Yohai, V. (1984). Robust regression by means of S-estimators, Robust and Nonlinear Time series Analysis. *Lecture Notes in Statistics*. 26: 256-272.

Simpson J. R., Montgomery D. C. (1998). A performance-Based Assessment of Robust Regression Methods. *Comm. In Stat.-Simulation and Computation*, 27(4), 1031-1 049.

Simpson, J. R. (1995). *New Methods and Comparative Evaluations for Robust and Biased-Robust Regression Estimation*. Unpublished Ph.D. thesis, Arizona State University, The United States of America.

Shaibu, A. B. And Cho, B. R., (2009). Another View of Dual Response Surface Modeling and Optimization in Robust Parameter Design. International Journal of Advanced Manufacturing Technology, 41 (7-8), 631-641.

Srikantan, K. S. (1961). Testing for a single outlier in a regression model. *Sankhya A*, 23:251-260.

Stromberg, A. J. (1993). Computation of high-breakdown nonlinear regression parameters. *Journal of the American Statistical Association.* 88:237–244.

Stromberg, A. J., Hossjer, O. and Hawkins, D. M. (2000). The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association*. 95: 853-864.

Tukey, J.W. (1960) A Survey of Sampling from Contaminated Distribution. In Contributions to Probability and Statistics; Olkin, I., Ghurye, S., Hoeffding, W., Madow, W., Mann, H., Eds.; Uni. Press: Stanford,448-485.

Taguchi G., Wu Y. (1985). Introduction to OO-Line Quality Control. Nagoya, Japan: Central Japan Quality Control Assoction; 1985.

Uraibi, H. S. (2009) .*Dynamic Robust Bootstrap Algorithm for Linear Model Selection Using Least Trimmed Squares,* Unpublished M.Sc. thesis, Universiti Putra Malaysia, Malaysia.

Vining, G. G., Myers, R. H. (1990). Combining Taguchi and Response Surface Philosophies: A Dual Response Approach. Journal of Quality Technology,22:34-45.

Vining, G. G. (1998). A Compromise Approach to Multiresponse Optimization. Journal Quality Technology, 30:309-313.

Vidmar, T. J., Mckean, J. W. (1996). A Monte Carlo Study of Robust and Least Squares Response Surface Methods. *Journal of Statistical Computation and Simulation*, Vol.54, pp. 1-18.

Vuchkov, I. N., Boyadjieva, N. L. (2001). Quality Improvement with Design and Experiment A Response Surface Approach. Kluwer Academic Publishers, The Netherlands.

Venables, W. N. and Ripley, B. D. (1999). Modern Applied Statistics with S-Plus 3rd Edition. Springer Verlag, New York.

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*. 2nd edition. The United States of America: Elsevier academic press.

Weisberg, S. (1985). Applied Linear Regression, New York: JohnWiley

Yohai V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates For Regression. *The Annals of Statistics*, Vol. 15, No. 20, 642-656.

Ypma, T. J. (1995). Historical development of the Newton-Raphson method. *SIAM Review*. 37 (4): 531-551.