Robust Design for Security and Humanitarian Support

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Edward B. Teague IV

May

2013

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

AUTHOR

The dissertation has been read and approved by the examining committee:

Prof Donald Brown

Advisor

Prof Steven Chase

Prof John Farr

Prof Gerard Learmonth

Prof Garrick Louis

Accepted for the School of Engineering and Applied Science:

James H. Ay

Dean, School of Engineering and Applied Science

May

2013

Acknowledgements

I am deeply indebted to my wife, Frances, my children, Maddie, Jack, Elena, & Frannie, my fellow graduate students, my advisor, my committee members, the faculty of the United States Military Academy Department of Systems Engineering, and Dr John Pemberton for the opportunity, time, patience, and support they provided during this research effort.

ABSTRACT

After conflict and disaster, social stability is a high priority strategic goal for stakeholders. Reconstruction and infrastructure development close capacity gaps, gain popular support for governments and institutions, and stave off illegitimate authority. Development allows the population to resume their daily lives and the government to demonstrate its reach and capabilities. It is a means to undermine support for insurgents and illegal activity while fomenting order. Infrastructure portfolios with carefully determined characteristics can be explicitly selected with this in mind, constituting a system. An optimal infrastructure portfolio for such a nebulous environment should include robust design features. It well satisfies design criteria and demonstrates resistance to exogenous factors. A systems approach using agent-based modeling, response surface methodology, robust parameter design, and a local optima filter provides leaders with statistically distinct, locally optimal choices better informing infrastructure decisions in a complex environment by using noneconomic measures to recommend settings in support of population stability.

The meta-model robust design process is introduced as a systems engineering methodology to address infrastructure decisions in complex, adaptive environments with exogenous factors. The process is comprised of a several subcomponents that trade accuracy (bias) for robustness. The robust features of the methodology include robustness in regression and robustness in parameter design. The local optima filter is able to differentiate between control variable recommendations when response confidence intervals and associated statistical tests fail to do so.

The meta-model is applied using known functions to demonstrate its properties. It is also applied to a post-combat infrastructure selection scenario in the city of Jalalabad, Nangarhar Province, Afghanistan. Finally, it is applied to infrastructure policy selection in Tijuana City, Baja California, Mexico. Though application, the meta-model robust design process provides stakeholders with recommendations that might not be otherwise selected due to the myriad permutations, the curse of dimensionality, heteroscedasticity, the desire for robustness, and the use of noneconomic assessment measures.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

"If a problem cannot be solved, enlarge it."

$\sim$ President Dwight D. Eisenhower

In February 1991, the Gulf War ended the Iraqi occupation of Kuwait. The aggressor's retreat eliminated security threats and fostered unfettered reconstruction. Coalition partners transitioned authority to the Kuwaiti government and departed. Despite hosting the world's largest military battle in 25 years, Kuwait quickly restarted its economy and resumed its prosperous, pre-war existence.

Conventional combat followed by a lasting accord is no longer the norm for the United States and its allies. Worldwide, opposition entities capitalize on the relatively low cost and high impact of persistent insurrection. They orchestrate protracted crusades, with the assistance of nonstate actors, without regard for realistic end states, capability calculus, or treaties and truces. Recent experience in Iraq and Afghanistan support this conjecture. [2] Both countries demonstrate the rise of insurgencies. After Saddam Hussein's regulars were defeated in 2003, it took nearly a decade for the United States and its partners to end large-scale security operations in Iraq. Presently, lasting peace in Afghanistan is a distant goal despite the North Atlantic Treaty Organization's (NATO) 2001 invasion.

Because an insurgency's center of gravity is a dissatisfied civilian population, a counter-insurgency (COIN) strategy under consideration by the United States Department of Defense (DOD) includes implementation of geographically tailored infrastructure projects. [3] With infrastructure, a nation demonstrates capacity and sets the conditions for peace and prosperity. In post-combat / disaster environs governments must serve the population but they are often fragile and lack depth. A policy of increasing infrastructure capacity can erode support for and indifference toward insurgents.

In order to limit unrest and foster long term stability, it is critical that leaders choose the greatest value-added infrastructure projects. However, the infrastructure is subject to dynamic and complex environments like social and economic systems. These environments present significant challenges when designing infrastructure systems, selecting project portfolios, and allocating resources to support the population and subdue discord. [4]

Reconstruction is not just necessary after violence and conflict. Natural and manmade disasters strike bluntly at a locale and also precipitate reconstruction. The contrast between Hurricane Sandy's impact on the northeastern coast of the United States in November 2013 and NATO reconstruction in Afghanistan illustrates the problem scope and research need. Storm damage in New York and New Jersey exceeded $70 billion (US), but leaders restored systems and services using historical references, codified processes, and trusted agencies. [5] In Afghanistan, $130 billion was spent in 2009 and 2010. [6] Rather than repairing pre-existing systems across a highly-evolved metroplex, the Afghan resources have been distributed at multiple levels via assorted stakeholders in a loosely affiliated country, with triple the population, thirty times the area, but without unified intent and purpose. [7] In addition, the resources are not limited to application to brick and mortar constructions. They are distributed across security, governance, development, and counter-narcotic functions, with the loin's share going toward security (¿

2/3). [6]. The absence of strong authority, stability, and a viable blue print due to decades of blighting warfare makes broad improvement unlikely given current methods of infrastructure improvement selection. [8, 9]

While it is not the only suitable application of the research, large scale warfare is of interest. Warfare involves premeditated, focused, and sustained efforts to disable critical infrastructure. In the time period following warfare, it is critical to quickly reestablish stability but this is a challenge for stakeholders due to capacity gaps. [10]

Viewed through the lens of systems engineering, the complexity of this problem relates to stakeholder needs and capacity delivery. It encompasses the challenge of an engineering solution for a labyrinthine environment. It demands robustness, a viable solution despite a range of exogenous, uncontrollable, factors. Design recommendations must account for locale and stakeholder behavior and justify the investment, long time horizons, and enormous resource allocation. A systems view of infrastructure selection provides leaders with insight they currently do not possess in order to implement better policies for political and popular stability.

Four steps, each with subcomponents, comprise a solution strategy to underpin infrastructure decisions, the Meta-Model and Robust Design Process (MMRDP)

- Create a model of relevant stakeholder behavior

- Create, evaluate, and select statistical models based on data from the behavioral model

- Find robust system settings that account for exogenous factors and achieve a desired end state

- Provide stakeholders a choice of settings based on resource constraints

The process estimates and analyzes poorly understood, high dimension, nonlinear response spaces that are subject to control and exogenous factors, Figure 1.1. Stakeholders require consistent, predictable end states to justify the effort required to enact solutions. In this regard, the estimated response must be robust, with quantifiable insensitivity to exogenous factor volatility and varied system design settings.

1.1 Problem Statement

The dissertation research goal is to provide policy makers with a method to guide municipal infrastructure selection policy in poorly understood environments. This could be in the time period after large scale military combat, after a regional disaster, or a situation requiring significant levels of previously unavailable capability. This effort is captured by the research questions and hypotheses.

Research Questions

- How does one support strategic stability goals using public infrastructure?

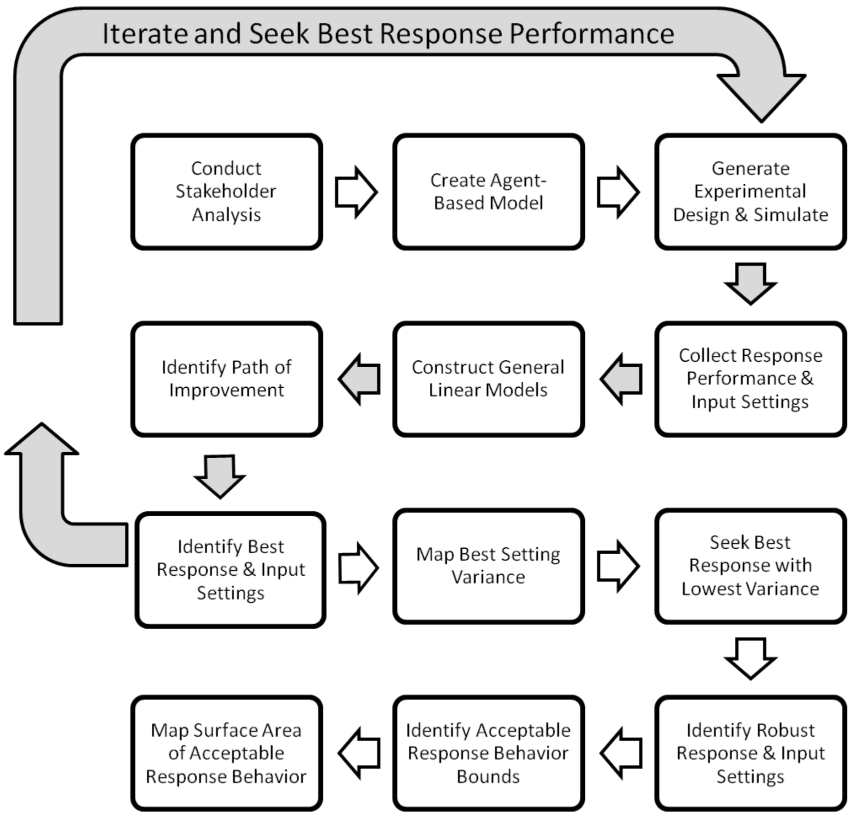


Fig. 1.1: The meta-model robust design process incorporates systems thinking, simulation optimization, and robust methods in order to improve infrastructure development policy.

- How does one identify settings for infrastructure investment in a particular locale that are optimal and robust?

Hypotheses

- It is possible estimate locally optimal response behavior that is insensitive to exogenous behavior in high dimensional space.

- It is possible to efficiently estimate relative probabilities for portfolios that achieve satisficing expected response behavior in high dimensional space.

Given an unstable and complex environment with infrastructure capacity gaps, apply noneconomic criteria to select from infrastructure projects, each with associated costs, and attributes, to discriminate between projects and select portfolios that support stability. Using a proxy for a measure of population stability for a particular geographic area, Y, maximize Y, as a function of first order functions of, second order functions of, and interactions between $1..N$ control factors (X), $1..M$ environmental factors. Constraints include but are not limited to control limits and a total budget, B, as determined by the stakeholders.

$$\text{Maximize } Y, Y \sim F(X, Z) \tag{1.1}$$

subject to

$$x_i \leq x_{U(i)} \tag{1.2}$$

$$x_i \geq x_{L(i)} \tag{1.3}$$

$$\bigvee i \in I, I = 1..N \tag{1.4}$$

$$Factor_j \sim G_j \tag{1.5}$$

$$\bigvee j \in J, J = 1..M \tag{1.6}$$

$$\sum \Psi X \leq B \tag{1.7}$$

$X_L, X_U -$ control variable upper and lower bounds.

$G -$ general exogenous variable distribution

$B -$ budget constraint

1.2 Dissertation Organization

The dissertation literature review addresses infrastructure systems and MMRDP components. The following chapter outlines MMRDP and its properties. MMRDP is then applied to city models of Jalalbad, Afghanistan and

Tijuana, Mexico. Results are presented. The dissertation concludes with significant findings and discussion.

2. LITERATURE REVIEW

”A successful society is characterized by a rising living standard for its population, increasing investment in factories and basic infrastructure, and the generation of additional surplus, which is invested in generating new discoveries in science and technology.”

∼ Robert Trout, American Radio Commentator

The literature review addresses steps within MMRDP, stakeholder analysis, infrastructure effects, measures to validate infrastructure selection, modeling techniques for data generation, and response surface methodology (RSM) to map and discover outcome behaviors of interest. Overall, the literature dictates the need for MMRDP due to the lack of a codified methodology to model a social environment in order to infer particular factors, interactions, and higher order effects for use in infrastructure selection, particularly when considering noneconomic measures.

2.1 Stakeholder Analysis

MMRDP is a systems engineering approach and as such requires a thorough understanding of the environment, stakeholders, factors, relationships, and measures. The philosophy of a systems view is well stated by Parnell et al. (2008). Stakeholders select infrastructure systems that close known capability gaps and underpin documented needs, such as security, governance, and development. A systems view demands analysis of measures, inputs, functions, linkages, interactions, and outcomes. In this way, one can identify and characterize potential solutions. Figure 2.1 shows the systems considerations when generating a basic framework for solution identification and scoring.

> *Systems thinking is a holistic mental framework and worldview that recognizes a system as an entity first, as a whole, with its*

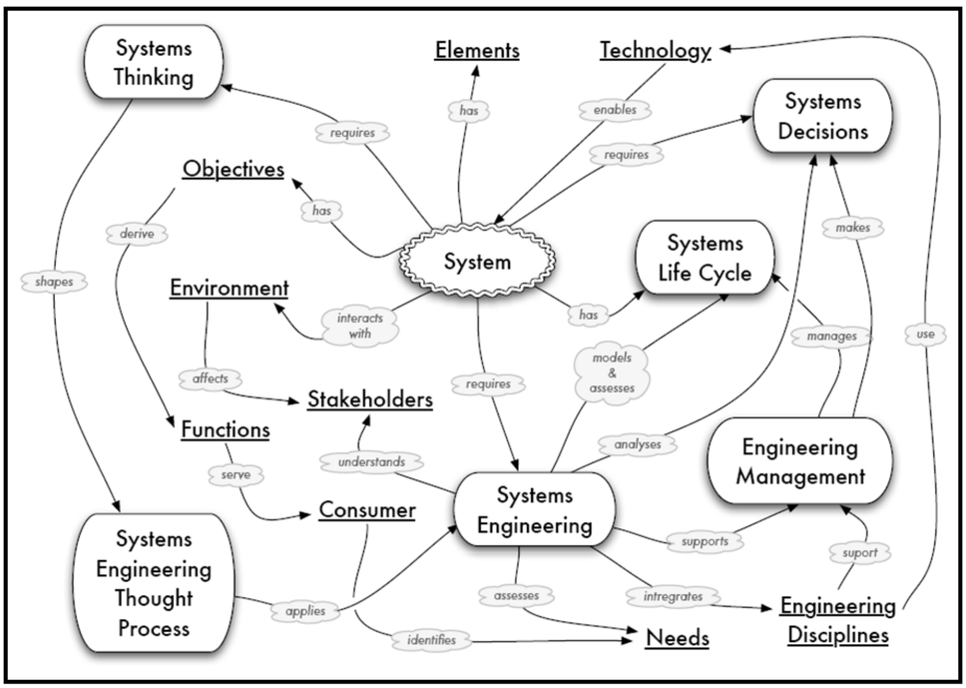


Fig. 2.1: Systems thinking as illustrated by Parnell et al.

> *fit and relationship with its environment being primary concerns. [11]*

Stakeholder analysis results in problem scope and centers on data collection. Three direct methods are available for collection, interviews, focus groups, and surveys. [11] One must assess the situation to determine the best method for their stakeholders using efficiency and the informative nature of the feedback as measures. For instance, the literacy and size of a stakeholder population may require large scale, face to face interviews rather than other, less labor intensive methods to collect data. Other considerations such as a stakeholder's willingness to participate, culture or morays, may have significant impact on collection methods. It the responsibility of the practitioner to be sensitive to these issues in order to perform a comprehensive stakeholder analysis.

Stakeholder analysis is a primary and critical step, as it informs all subsequent elements of MMRDP. The process of stakeholder analysis is unique to the problem at hand, however it includes the following steps in every case. [11]

Identify People and Organizations Relevant to the Problem: Compile the set of stakeholders

Determine the Stakeholder Wants, Needs, Desires: Identify the problem functions, objectives, measures, and constraints

Isolate Relevant Factors: Consider politics, economics, society, ethics, values, history, and technology

Refine the Problem Statement: Use stakeholder input to bound the system, scope, and redefine the problem

An analysis of the local environment can yield key insights about the functions that impact stakeholder objectives. One then develops stakeholder needs, outlining capability gaps that will drive system design. Those gaps have associated, quantifiable metrics. The metrics facilitate a basis of comparison when determining if needs and objectives are best met.

Parnell et al. (2008) define solutions in terms of their ability of the design to meet the stated need:

> *"There are three different types of system designs: satisficing, adaptivising, and optimizing. Satisficing, for an existing system, means the current performance is satisfactory. For a new system, satisficing means that any feasible solution will be satisfactory. ... The primary criteria for adaptivising solutions is cost effectiveness. Improved system performance is the goal, but only when it can be obtained at a reasonable cost. Satisficing accepts any solution that works. Adaptivising looks for "good" solutions. ... Optimizing solutions are better than or at least equal to all others. They are the best according to the performance measures. However, they may require much effort or expense."*

It is important that the practitioner and stakeholders use their assessment of the environment and resources to apply distinctions within the solution strategy.

2.2 Infrastructure Classification

In this study, public infrastructure research includes primary and secondary effects of infrastructure improvement or introduction of new infrastructure or systems of infrastructure on a population. Segments within the population and their characteristics are considered. The research addresses efforts to improve local environment attributes as determined by local, regional, national, or international offices or bodies.

Governments are inexorably responsible for and at the mercy of infrastructure. This relationship pre-dates written history and it defines cultures, expectations, and long term regional success. Whether provoked by disaster, warfare, moral imperative, or other needs, infrastructure investment is costly and must be justified. Infrastructure investment must explicitly address gaps by type.

Infrastructure definitions are subjective and can be classified in a number of ways. The American Society of Civil Engineers (ASCE) publishes a biannual report card on infrastructure within the US by type and subtype. Its primary categories (classes) are Water and Environment, Transportation, Public Facilities, and Energy. [12] Each type has a very specific description and assessment criteria.

Countries and other consortiums also categorize infrastructure according to functional decomposition. [13] For instance, the World Bank not only makes functional distinctions in infrastructure systems, it delineates between tanglible and more abstract systems, such as waste water versus public health. [14] Regardless of how systems are categorized, a thorough functional analysis will best outline needs that can later be mapped to systems within proposed solutions. [11]

2.3 Noneconomic Metrics for Infrastructure Selection

Measuring the effects of public investment on economic measures comprises over a century of research. In 1883, Adolph Wagner demonstrated the benefits to a state that invests in infrastructure, among many other services, for its people. [15] Based on analysis of 19th century Germany, *Wagner's law* posits that increasing revenues for the state should be plowed back into the population which then foment ever increasing productivity and revenue streams. The results of further exploration of his claim are mixed. Research on infrastructure benefits has primarily focused on measurable economics effects such as productivity and tax revenue.

Aschauer (1989) shows that government spending in the United States from 1949 to 1985 related to infrastructure gives a lift to the economy but other government spending showed no increase in productivity. [16] Other changes due to infrastructure investment are also difficult to specifically identify. As a component of infrastructure, transportation investment exemplifies this difficulty. [17] While it seems that public opinion generally assumes transportation networks have tangible benefit, Cohen (2007) [18] and Lakshmananan (2011) [17] both state that magnitude of benefit as well as the regional influence is not always conclusive. Benefits are more attributable at micro rather than macro levels due to obvious first order effects. [18]

These challenges appear to undermine the proposed research but in fact they propel it. Very few studies examine the relationships between the population, its perception of the environment, and infrastructure systems. Noneconomic, strategic arguments for infrastructure investment are philosophical, relying on engineering-based measures from singular projects to show how things will improve. There are no codified methodologies that attempt to capture second order effects of infrastructure systems on a population. This leaves organizations with interested in long term effects on the public, such as the United Nations (UN), NATO, Red Cross and Red Crescent, or national governments, without a method to broadly justify or inform decisions.

Along that vein, Perz et al. (2011) show the need for infrastructure in an area of low economic means. The southwest Amazon basin is challenged to capture the nature of all stakeholder needs across a region contained by three countries, with few modern conveniences and dozens of cultures. [19] They identify that infrastructure investment in the region is poorly measured by economic statistics thereby undermining the development opportunity for the local population.

Cavallo and Daude (2011) are also representative of developing nation infrastructure investment studies. Weakly performing geographic areas are not likely to get public infrastructure they need because the importance of philosophical and moral measures, such as quality of life, lag far behind economic measures when determining infrastructure feasiblity. This limits access to what many other parts of the world take for granted as essential services. [20]

2.4 Agent-Based Models

The high cost of infrastructure makes large scale trial implementation of multiple candidate solutions impossible. This constraint leads the practitioner to employ simulation to gain information and refine analysis. Given the high reliance on simulation and experimentation in this research, computer aided experimentation is a natural choice.

Computer simulation is ubiquitous within statistics, engineering, and experimentation. [21] It is also a tool for operations research and operations management researchers. [22] Kleijnen et al. (2006), Sanchez et al. (1996), and Sanchez (2000) are representative of the many surveys and papers that outline computer-based experimentation within a broad set of methodologies. The popularity of computer simulation is due to decreasing hardware costs as well as better and more intuitive software. These advances bolster extended applications as well, such as a broad infrastructure effects. [23]

Because infrastructure provides capability for a population, MMRDP estimates the effect of infrastructure on individuals. The variety of stakeholders, their interactions, high system complexity, and the desire for emergent behavior, make an agent-based model (ABM) appropriate for behavioral data for follow on within MMRDP. ABMs offer the combination of individual autonomous behavior, group aggregation, and emergent outcomes within a computer-based simulation. ABMs introduce autonomous agents in a prescribed environment and track changes to both over time. The agents interact with each other and the environment allowing study of the system as a whole. Cioppa et al. (2004) illustrate how ABMs are useful when testing tactics and strategy in dynamic environments where underlying causes and relationships are poorly understood. [24] These features, along with ABM flexibility, make them extremely attractive when encountering social problems without closed-form solutions.

The Department of Defense (DOD) and other federal agencies use ABMs for their breadth of application and depth of study. DOD sponsors a family of ABM tools with the New Zealand Ministry of Defence and the Naval Postgraduate School (NPS) collectively called Project Albert to aid research. [25] Project Albert simulations have modelled a variety of situations to include conventional warfare, logistics, security, and unconventional warfare. [26]

ABMs are common outside of the military purview as well. For instance

NetLogo, a mutli-agent programming language, is widely used freeware. It has demonstrated broad application in modeling complexity in nature and societies. [27] Designed and managed by the Center for Connected Learning and Computer-Based Modeling at Northwestern University, it includes a model library with dozens of systems that evolve over time such as wolf - sheep predation, enzyme kinetics, and motor vehicle traffic. [28] Netlogo and other similar languages are valuable due to their ease of use, their flexibility, and their large communities of diverse users.

Another, more specific, example of social modeling with ABMs, Young and Flacke (2010) present the logic and construct for an ABM to estimate population growth in Dar es Salaam, Tanzania. They apply existing data, such as UN development reports, and empirical data collected to server the needs of the simulation. [29]

A useful feature of an ABM is the ability to incorporate a Geographical Information Systems (GIS) layer. This layer significantly effects human behavior in reality and is germane to infrastructure planning. Isolation, avenues of travel, and other distinct features of geography are as important distance when planning infrastructure projects. The states of the agents reflect the geography's influence using GIS data imported into the model. At each time step, an agent's attributes can change as a function of model layers and interactions with other agents.

2.5 Response Surface Methodology

MMRDP steps include response surface methodology (RSM) given the desire to find an optimal response as a means to highlight certain input factors and recommend settings. Developed by Box and Wilson in 1951, RSM is a framework combining statistical models and numerical approaches in order to generate a relationship between a response, y and $k+1$ inputs, $(x_1, x_2, ...x_{k+1})$, where $(k+1) \in P = (1..p)$, $k \leq p$ factors or variables. [30] Typically, factors have constraints such as $x_{klower} \leq x_k \leq x_{kupper}$ or $x_k \in Z$.

RSM uses an estimate of a previously unidentified relationship to provide the user with a locally optimum response value and its associated factor settings. [31] Model 2.1 shows the general form of the relationship between the response and selected factors. [32]

$$y = \eta(\vec{x})\beta + \epsilon \tag{2.1}$$

Here, $\eta(\vec{x})$ and β comprise the true but unknown function of inputs of first order or higher. RSM attempts to estimate this relationship. The associated random error is ϵ, $\epsilon \sim (0, \sigma^2)$. [32]

The model used to estimate model 2.1 is typically restricted to low order, two or less with interactions, regardless of the complexity of the true relationship, $\eta(\vec{x})\beta$. Assuming that one defines a small enough neighborhood, RSM uses first order relationships, model 2.2, and second order relationships, model 2.3, to represent model 2.1. They are sufficient given assumptions about the error and response expected value $E[y|\vec{x}] = E[\hat{y}|\vec{x}]$ hold. [32]

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \epsilon \tag{2.2}$$

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum\sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \epsilon \tag{2.3}$$

By iterating experiments, RSM highlights a sequential path to an area on the estimated response surface where the performance is better than earlier addresses. RSM is complete when better performance cannot be found. Box and Wilson point out that RSM does not promise convergence and one may initiate RSM from multiple starting points to search for global optima. [30] Based on termination criteria the process provides relevant factor levels to achieve local optima driving system design and stakeholder decisions.

The exploration of the surface is fueled by successive experimental designs. It allows the practitioner, through a flexible but codified set of steps, to move through an solution space toward an area where the estimated response performance is may be better than earlier addresses. RSM segments the response surface by experimenting in small neighborhoods, Figure 2.2. RSM maps response surfaces using piecewise approximations using little a-priori knowledge of the true function, Figure 2.3. In this way, RSM can estimate and maneuver through a P-dimensional response surface one small neighborhood at a time evaluating the mean of the estimated response function and its variance, σ^2. [33]

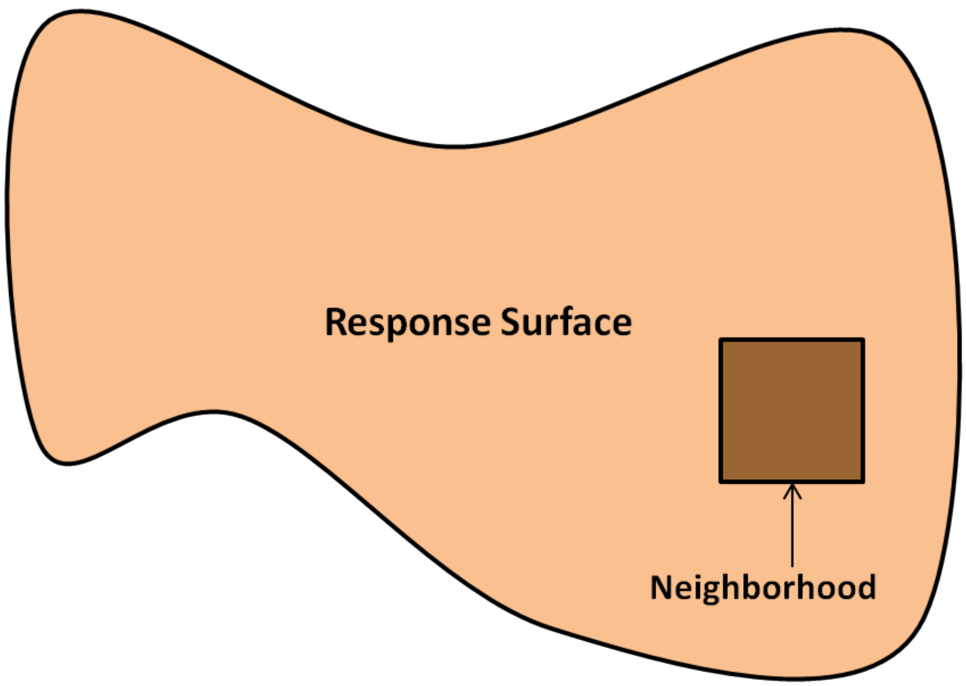


Fig. 2.2: The response surface experiments are in small neighborhoods.

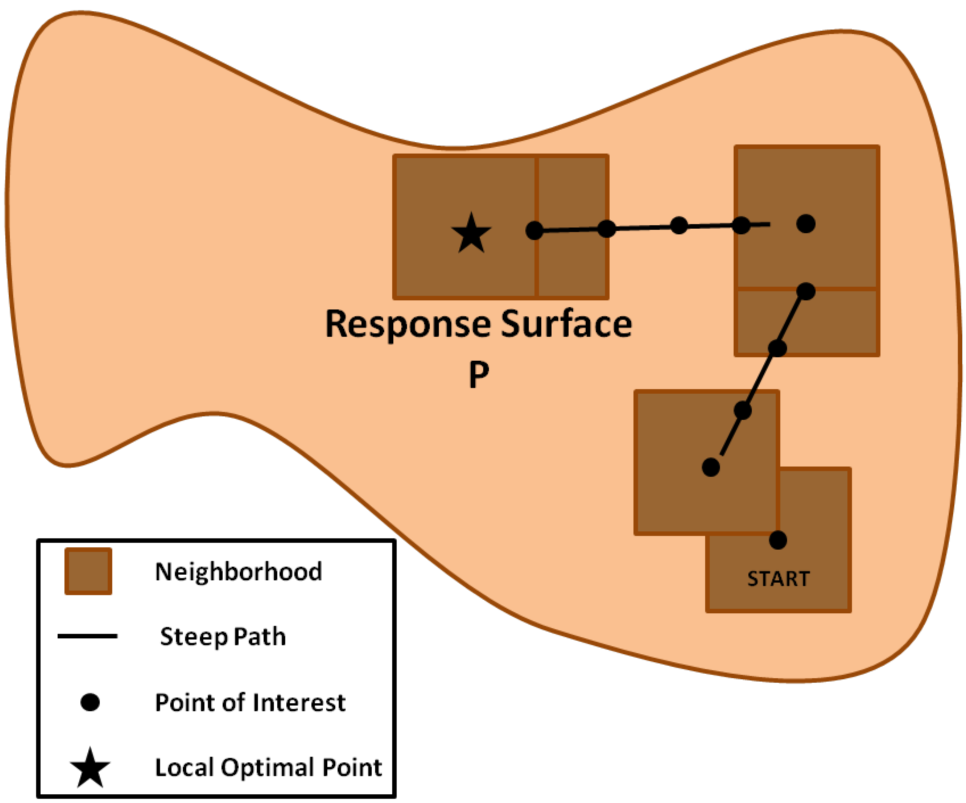


Fig. 2.3: Optimality is discovered by moving across the experimental neighborhoods of a response surface via the discovery of steep paths.

Between each successive better performing response point, a path of steepest ascent / descent (PSA / PSD) is constructed from $\hat{\beta}$. RSM then uses the PSA / PSD as an azimuth to move using arbitrary steps, toward a more desirable mean response performance as defined by the user, a minimum, maximum, or target response value, Figure 2.4. [31]

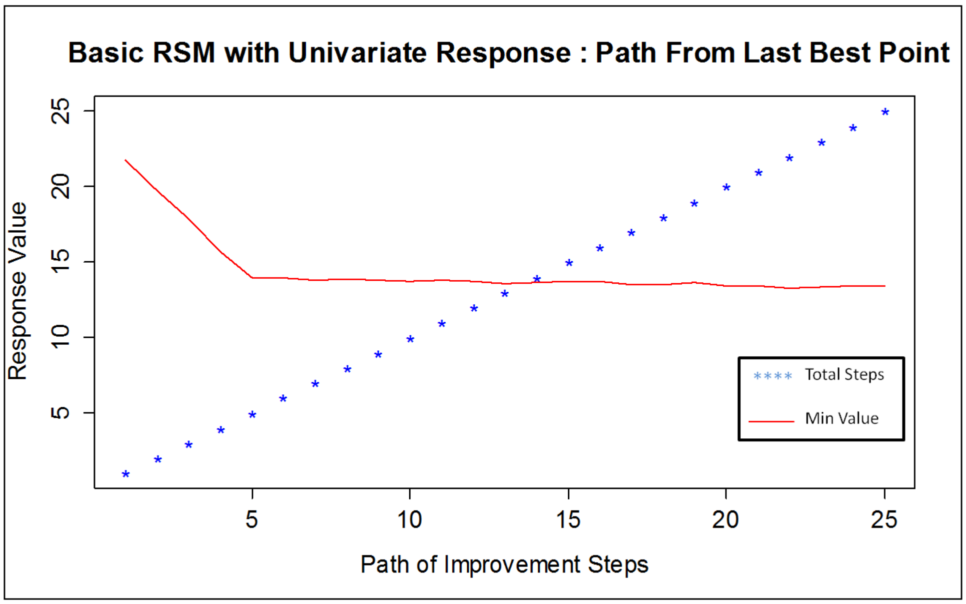


Fig. 2.4: RSM seeks response improvement along the path of steepest improvement.

Tests for curvature as well as general lack of response improvement alert the user to generate a new experiment, a new PSA or PSD, or to terminate the effort. If curvature is present, a second order model is fit. Using a system of equations generated by partial derivatives of the estimated second order model can then be solved in order to find the stationary, locally optimal, point. The optimal point classification is local because RSM cannot guarantee convergence. It is sensitive to the surface topology, initial point, PSA /PSD step size, and experimental design construct. [31] RSM includes the following steps.

- Selecting a starting point in P space
- Conducting experiments within a specified neighborhood using up to P factors
- Creating 1st or 2d order regression models from significant factors to estimate the response as appropriate
- Identifying and following a path toward better response performance
- Determining a new starting point or terminating the process with a recommendation of a factor address ($\leq P$ *elements)* and a relationship that yields a best local response performance

It is important to note that error within PSA increases as the distance from the latest "best point" increase because the error of each successive step maintains the error of the intial step and that of the regression model used to take the step. The more steps that one takes the greater the error. The error is re-baselined with each regression resulting from a new experiment.

2.5.1 Locally Optimal Point Selection

The primary motivation for RSM in this research is to identify optimal settings for control variables. Once $\overrightarrow{\beta}$ is found via regressive, additive, or other models, one can solve for the optimal input levels.

$$y = \widehat{\beta_0} + x'\widehat{\beta_1} + \sum\sum_{i<j} x_i'\widehat{\beta_{ij}}xj + \sum_{i=1}^{k} x_i'\widehat{\beta_{ii}}x_i + \epsilon \tag{2.4}$$

assuming that $\mathrm{E}[\epsilon] = 0$ and variance of the estimated response is constant, equation 2.5 shows the relationship between $\vec{x}$ and y.

$$\hat{y}(x) = \widehat{\beta_0} + x'\widehat{\beta_1} + \sum\sum_{i<j} x_i'\widehat{\beta_{ij}}xj + \sum_{i=1}^{k} x_i'\widehat{\beta_{ii}}x_i + \epsilon \tag{2.5}$$

While other methods exist to find $\vec{x}$, such as Hoerl's 1959 introduction of ridge analysis, later modified by Draper in 1963 and Khuri and Myers in 1979, one can use partial derivatives of $f\hat{x}$ to generate a system of equations in order to find the locally optimal setting, $\vec{x_*}$. [31]

2.6 Experimental Design

Within RSM, the exploration of the surface can be fueled by successive computer-based experiments. For this research, all experiments are full factorial, $resolutionIV$, designs with each of k factors run at $+1$ and -1 levels for n replications of 2^k points. Many other experimental designs are available but the anticipated cost due to computer simulation versus other design types is low. [31]

Experimental design considerations include orthogonality of the design, to avoid collinearity within $\vec{\beta}$ and rotatability, ensuring constant variance for all design points. Rotatable designs were developed by Box and Hunter in 1957 to fix the estimated response variance and that of the $\vec{\beta}$. [32] This allows for better comparisons of points in hyper-space and reduces violations of RSM assumptions.

When appropriate, second order models use an experimental design that adds a central composite design (CCD) to the 2^k design for a total of $n = 2^k + 2k + 1$ points. The greater number of points is due to the greater number of parameters to be estimated in the second order design as shown in model 2.3. In this case there are $p = 1 + 2k + \frac{k(k-1)}{2}$ parameters, the minimum number of design points required. [32]

The CCD includes the center point and $2k$ facial points to form a star design. Simply put, a 2^k design forms the corners of a hyper-cube and the CCD generates axial points in the center of each face but at the same distance from the center point as each corner. This way the design ensures constant variance for each element in $\vec{\beta}$ and by extension constant variance in the estimated response. [34] This distance, α, is chosen to ensure rotatability, thus $\alpha = F^{1/4}$ where $F = 2^k$, the number of points in the full factorial design.

For example, in a 2^3 design, there are 6 faces requiring $2 * 3 + 1$ additional points for the CCD when including the center point. Assuming that each corner point is a radial distance of $\sqrt{2}$ from the center, points centered on each face of the cube must be the same distance. In the 2^3 design this equates to a distance of $k^{1/2}$. Figure 2.5, shows the 2^3 full factorial, main effects design including a CCD. The 15 design points are sufficient to estimate the 10 parameters of a second order model. Though a full factorial is a greedy experimentation method, the use of computer simulation as a part of MMRDP facilitates these techniques. In addition, a full factorial design, retains sufficient degrees of freedom to include interactions within the model as necessary. In the event the simulation becomes too costly one can use nearly orthogonal latin hyper-cube sampling, (NOLHS).

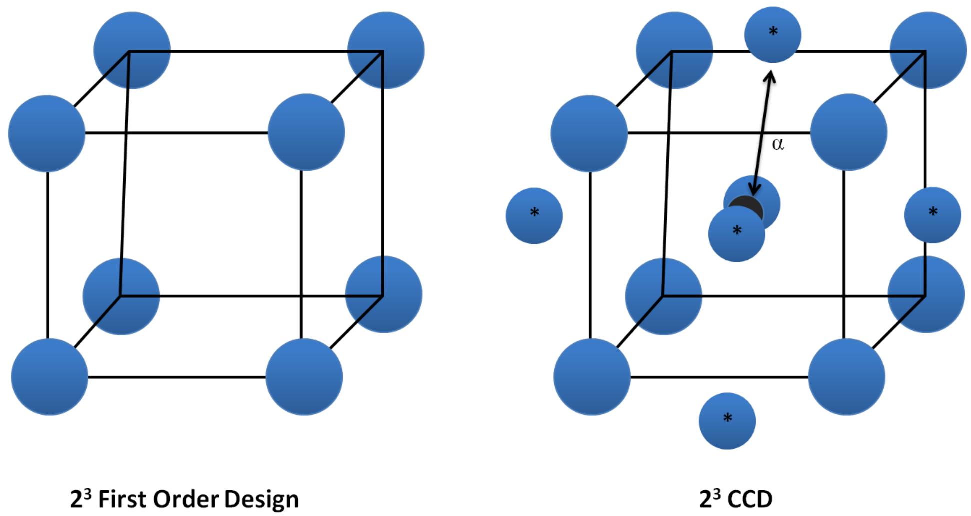


Fig. 2.5: The experiments in a 2^k first order function estimate full factorial design jump to $2^k + 2k + 1$ for a second order function.

NOLHS can reduce the computational burden due to simulation. It randomizes the levels of all factors considered by the ABM and reduces the number of design points while generates a space-filling set of experiments, ensuring a thorough exploration of the levels and factors at a discount. The NOLHS algorithm developed by Cioppa and Lucas, is a smaller design, induces orthogonality, and fits main-effect and interaction-effect models with nearly uncorrelated estimates of the coefficients for linear regression. [35] In this instance, relief from full factorial designs in P-space can result in experiments that are orders of magnitude smaller.

NOLHS has drawbacks. It saves time, money, and effort, but provides partial coverage of space as a tradeoff for efficiency. It must be carefully applied to show a distinction between integer and continuous variables. With continuous variables NOLHS forces its user to select alternate methods and interpolate response values across large gaps at times as all design points are not run. This can be problematic with complex functions but can be mitigated by viewing the function in small neighborhoods when applied as a component of RSM.

2.7 Statistical Model Evaluation

Of the the methods to uncover and describe structure that may exist in the data, sum of squared error (SSE) is natural choice to determine the accuracy of a model. It is immune to model nesting considerations. [36] SSE, mean squared error (MSE), root mean squared error (RMSE) are unbiased but intuitively, suffer from the scale of the input variables. Normalized MSE (NMSE) is attractive because it is scale insensitive. [37]

$$SSE = \sum (y_i - \widehat{y_i})^2 \tag{2.6}$$

$$MSE = \frac{\sum (y_i - \widehat{y_i})^2}{n - p} \tag{2.7}$$

$$RMSE = \sqrt{\frac{y'y - \hat{\beta}X'X}{n - p}} \tag{2.8}$$

$$NMSE = \frac{1}{\hat{\sigma}^2 n} \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 \tag{2.9}$$

The square of correlation coefficient, R^2, adjusted R^2_{adj}, predicted R^2_{pred}, and Mallow's C_p, are also tools to compare model fit.

$$R^2 = \frac{\hat{\beta}X'X - n\bar{y}^2}{y'y - n\bar{y}^2} \tag{2.10}$$

$$R^2_{adj} = 1 - (1 - R^2)(\frac{n-1}{n-p}) \tag{2.11}$$

$$R^2_{pred} = 100[1 - (\frac{PRESS}{y'y - n\bar{y}^2})]\% \tag{2.12}$$

$$C_p = \frac{SSE}{\hat{\sigma}^2} - n + 2p \tag{2.13}$$

However, they have drawbacks as well. For instance, R^2 and all of its variants measure linear correlation. Thus it is not suitable for second order models. Mallow's C_p penalizes the user for a higher numbers of factors but only for those models where the number of factors is greater than the expected value of Mallow's C_p, P. The Cp statistic is often used as a stopping rule for various forms of stepwise regression. Mallows proposed the statistic as a criterion for selecting among many alternative subset regressions. Under a model not suffering from appreciable lack of fit (bias), Cp has expectation nearly equal to P;

Models with large values of R^2, and R^{2adj}, and small values of RMSE are sought. R^2 increases with additional predictor variables regardless of how significant factors are. On the contrary, R^{2adj} may decrease if additional predictors do not contribute significantly to explaining the variability in the response. Thus, it is important to observe both statistics rather than R^2 alone.

In the actual, unknown relationship, $Y = \eta(\theta) + \epsilon = f(\theta) + \delta(\theta) + \epsilon$ where ϵ is random error and $\delta(\theta)$ is the bias. [31] Typically, $\delta(\theta)$ is unknown. In the event that $\delta(\theta))$ cannot be calculated, a more likely scenario, one can use MSE as a measure to evaluate a model. MSE can be easily calculated for a model without explicit knowledge of bias, and it accounts for bias and variance of the response. $MSE = \sigma^2 + \delta^2$.

Within RSM, the process repeatedly compares models; first order, first order with interactions, second order, and second order with interactions. Because the models are nested, MMRDP uses deviance as a method to compare and select. Deviance, a calculation of the log - likelihood ratio for a model, is suitable for Generalized Linear Models (GLMs) with continuous, discrete, and binary data. [37]

Agresti (2002), summarizes deviance and its calculation. Given a set of observations, $\boldsymbol{y} = (y_1, ..., y_n)$, $L(\mu;\boldsymbol{y})$ represents the log-likelihood function of the means $\mu = (\mu_1 ... \mu_n)$ The likelihood function is a monotonically increasing function whose maximum is at the same location as the maximum probability. This is implied by the relationship between the likelihood function of the parameters given factor observations and the probability mass function (pmf), $L(\Theta \mid x) = f(x \mid \Theta)$. [37]

Agresti (2002) denotes the maximum of the log-likelihood as $L(\hat{\mu};\boldsymbol{y})$ for a model. It is achieved via the saturated model, one where there is a separate parameter for each observation and perfect fit. However, the saturated model is useless with other data as it is not generalizable and it offers no feature reduction. It is the basis for comparison for all other models being fit. Any model other than the saturated model yields a likelihood function, $L(\boldsymbol{y}';\boldsymbol{y})$. Deviance, $-2[L(\hat{\mu};\boldsymbol{y})]$ can be calculated for each model under consideration. Additionally, deviance is χ^2 distributed, allowing for estimation of the statistical significance of changes in the ratio. [38] Within RSM, the model with the lowest level of deviance is selected for best fit.

2.8 Conclusion

The amalgamation of these well defined processes offer a framework for MM-RDP. They provide the practitioner a way to move through problem development and its scope, model the environment, generate data, estimate relationships, and seek optimality. Additional steps must be taken when model assumptions are violated or if one desires robust settings.

3. METHODOLOGY

"Thomsen and Cantrell both expressed some doubts about the future popularity of such approximate formulas for the hyperareas of hyper-ellipsoids."

∼numericana.com, May 2004

Since 2009, Knud Thomsen's approximation formulas have been cited 54 times.

∼scholar.google.com, March 2013

MMRDP is comprised of subcomponents arranged in a series of steps. The literature review addresses its basic elements. Using the process to estimate relationships in complex adaptive environments requires extensions within RSM, statistical models, and robust design. Robust solutions also demand greater clarification for inclusion in MMRDP. A detailed version of MMRDP is outlined in Figure 3.1.

3.1 Methodology Overview

The meta-model, $F(X, Z)$, initiates the process that inculcates control variables, X, and exogenous variables, Z. It represents a true function for output in terms of the environment, its inputs, and relationships, $\eta_{X,Z}$. It is rendered and validated using stakeholder analysis and represented via a computer simulation, Figure 3.2.

A combination of experimental design, regression, and RSM, join to form the estimate of the meta-model.

Characterize the Environment: Incorporate a systems approach and capture stakeholder needs and critical variables, processes, and relationships.

Create an Agent-Based Model: Represent key features of the environment, the stakeholders, and courses of action to provide a rich data set for analysis.

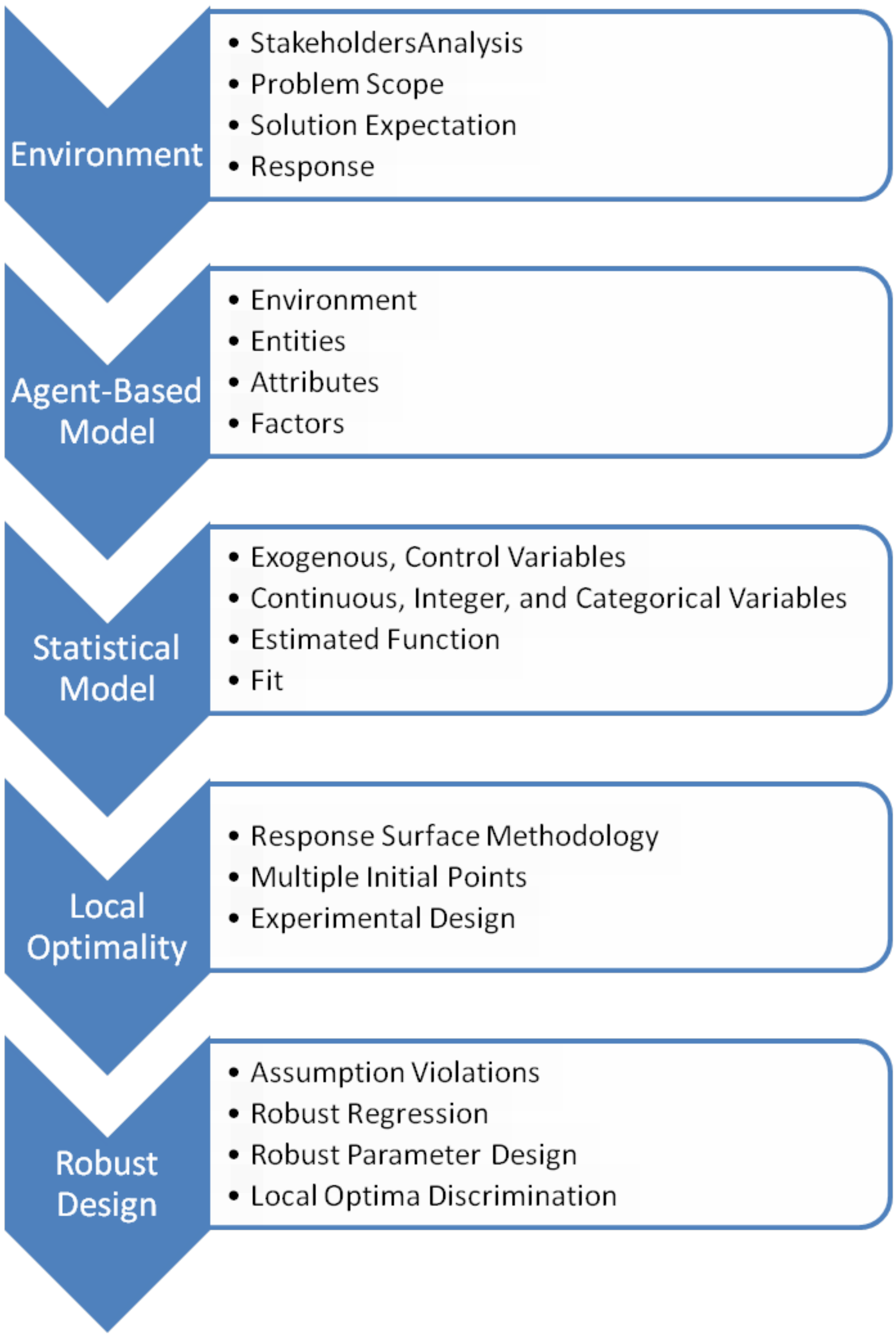


Fig. 3.1: MMRDP seeks optimality for infrastructure system selection in nebulous environments

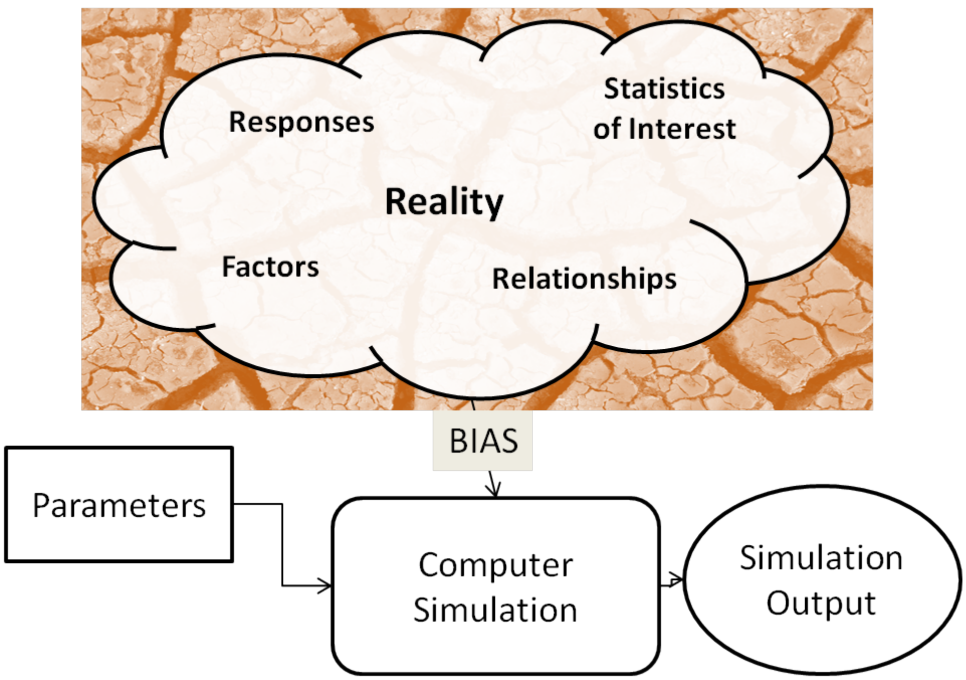


Fig. 3.2: The meta-model represents the environment and introduces bias

Seek Locally Optimal Settings: Employ RSM to provide a detailed view of the response surface in order to capture the effects of constraints and variance on optimal settings.

Use Statistical Modeling: Discover significant inputs and estimate low order relationships between inputs and the response in each experimental neighborhood.

Present Robust Design Alternatives: Provide stakeholders discernable alternatives based on optimal performance that are insensitive to system variance associated with exogenous variables.

Filter Robust Design Alternatives: Further discriminate between robust alternatives by comparing the relative probability of success for each alternative.

Recommend Pareto Optimal Solution: Provide stakeholders best candidate recommendations.

The combination of the meta-model and robust design offers users flexibility both in approach in general application. Despite the ground work necessary required to scope the problem, MMRDP is applicable across a wide range of environments. It is a systems approach and can use data from any model to further the process and serve as a basis for RSM. The process offer users the ability to trade speed and computational frugality for fidelity detailed exploration of the response surface. Output is further analyzed by assessing course of action likelihood of success before determining pareto optimality.

The general problem is stated formally.

Maximize Y $\sim f(\sum \beta X, \sum \xi Z)$

subject to

$X_i \leq B_{U(i)}$

$X_i \geq B_{L(i)}$

$\bigvee i \in I, I = 1..N$

$Z_j \sim G_j$

$\bigvee j \in J, J = 1..M$

$\sum \psi X \leq C$

B_L, B_U control variable upper and lower bounds

G exogenous variable distribution

C budget ceiling

A collection of techniques within MMRDP are required to address the problem.

3.2 Extreme Value Theory

Some environments experience influential but low probability events, such an improvised explosive device (IED) detonation, a tornado, or the discovery of valuable resource. It is natural to include such events in the ABM. However, the rarity demands study of extreme value theory.

Extreme values are not uncommon in complex dynamic systems. They highlight a two-fold problem. Usually, little data exists to fully characterize the environment or the relationships are very poorly understood to support confident estimates. [39] There are a number of ways to address this as the typical assumption of classic gaussian behavior for both inputs and outputs is not reliable. Selection of extreme value models is an exercise in tradeoffs. Gumbel models, Frechet models, Weibull models are a few goodness of fit and Likelihood ration tests are available for all [39]

Maximum likelihood estimation for parametric modeling can be completed based on the Generalized Extreme Value (GEV) Distribution outlined below. [40]

The model emulates the presence of

$$M_n = \max(X_1, X_2, ..., X_n) \tag{3.1}$$

where $X_1, ..., X_n$ is a sequence of independent and identically distributed (IID) output values. M_n represent the maximum value over a set time period of length n. The exact distribution of M_n is

$$P(M_n > z) = P(X_1 > z, X_2 > z, ..., X_n > z) \tag{3.2}$$

$$= P(X_1 > z) \times P(X_2 > z) \times ... \times P(X_n > z) \tag{3.3}$$

$$= (Fz))^n \tag{3.4}$$

Unfortunately $F(z)$ is rarely known. Estimation of $F(z)$ begins with its behavior as $n \to \infty$. Because $F^n(z) \to 0$ as $n \to \infty$, M_n must be transformed to make its probability insensitive to changes in scale.

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{3.5}$$

where $a_n > 0$, and b_n are constants chosen to stabilize the location and scale of M_n. As such the choices for a_n and b_n are more critical than M_n.

From Coles(2001), The Extremal Types Theorem outlines the range of limit distributions for M_n^*.

If there exist sequences of constants ($a_n > 0$) and (b_n) such that

$$P(\frac{M_n - b_n}{a_n \leq z}) \to G(z) \tag{3.6}$$

as $n \to \infty$ where G is a nondegenerate (stabile) distribution function, then G belongs to one of the following families

$$G(z) = \exp\{-\exp[-\frac{z-b}{a}]\}, -\infty < z < \infty \tag{3.7}$$

$$G(z) = \begin{cases} 0 & z \leq b \\ exp\{-(\frac{z-b}{a})^{-\alpha}\} & z > b \end{cases} \tag{3.8}$$

$$G(z) = \begin{cases} exp\{-[-(\frac{z-b}{a})^{\alpha}]\} & z < b \\ 1 & z \geq b \end{cases} \tag{3.9}$$

These three distributions, the extreme value distributions, are the Gumbel, Fréchet, and Weibull families. They will be sufficient to increase the fidelity of stochastic models as necessary.

3.3 Function Estimation

Often practitioners encounter ordinal and categorical variables essential to the model. In general, regression techniques abound to address all types of variables. However, as a component of RSM, ordinal and categorical inputs present a challenge. At the inception of RSM, Box and Wilson (1951) stated

> *We shall assume that within the region considered the derivatives of the response function are continuous.*

However, this can be overcome. Kleinjen et al.(2005) point out that discontinuities exist in highly complex environments and should be anticipated within the response surface. [21] Khuri and Mukhopadhyay (2010), Nelder and Wadderburn (1972), McColloch and Searle (2001), and Myers et al. (2002) point to GLMs as a suitable method to address function smoothness. [32,41–43] GLMs are favorable as they can fit discrete data, continuous data, and nongaussian distributions if necessary. [37]

Options for the practitioner are not limited to GLMs. Using models with discrete and continuous variables Moisen and and Frescino (2002) compared linear models, generalized additive models (GAMs), classification and regression trees (CARTs), multivariate adaptive regressive splines (MARS), and artificial neural networks (ANNs). While all showed suitability, MARS stands out for its computational speed and its performance in the presence of discrete variables. [44]

Multivariate Adaptive Regressive Splines

Complex, dynamic systems are often characterized by nonlinear relationships and high dimensional spaces. [45] Lack of deep understanding and the potential for broad application due to the high instance of complex systems has increased the interest and use of Generalized Additive Models (GAMs) in regression. [46] GAMs offer relevant utility within the research due to the strong potential for nonlinear relationships. The forward and backward selection of variables with GAMs is attractive because of dimension reduction. [46] A GAM with $X_1, X_2, ..., X_p$ as factors and Y as the response has the form

$$E(Y|X1, X2, ..., Xp) = \alpha + f_1(X_1) + f_2(X_2) + f_3(X_3) + ... + f_p(X_p) \tag{3.10}$$

The functions

$$f_j(X_j), j = 1..p, \tag{3.11}$$

assumed to be smooth, can modeled using sums of basis functions, h_m, and least square estimates for β values.

$$\sum_{m=1}^{M} \widehat{\beta_m} h_m(X) \tag{3.12}$$

The functions are a linear basis expansion in X. [46] They form elements of a piecewise linear function that can represent regular linear models, polynomial terms, and nonlinear transformations. In addition, the basis functions can act as indicators for regions of X_j to activate and deactivate functions and factors within those functions to best employ the GAM to approximate through least square the unknown function. [46]

MARS, an additive model, has been shown to efficiently and accurately estimate functions in high dimension regions. [47] Developed by Friedman in 1991, MARS offers benefit with high dimensional complex data sets. [48] This is true for estimating functions using data with and without error components. MARS, as a nonparametric method of approximation that uses forward/backward selection and multivariate basis function expansion takes the following form.

$$\widehat{f(X)} = \widehat{\beta}_o h_o + \sum_{m=1}^{M} \widehat{\beta_m} h_m(X) \tag{3.13}$$

The spline basis function construction of estimation allows for selection of different sets meaningful variables across different regions of the response surface. [47] Crino and Brown (2007) and Leathwick et al. (2006) showed that MARS, as a component of RSM, can outperform other methods of response approximation including GAMs, Neural Networks(NN), simulated annealing (SA) and genetic algorithms (GA). [49] MMRDP could potentially be applied across a broad range of systems and high dimensional spaces. The MARS model's ability to estimate linear and nonlinear relationships with its piecewise construction makes it very attractive.

Employment of regression models within RSM points to GLMs as sufficient. GLMs are well suited for simple, low order models. GAMs and MARS models are more flexible. They can address nonlinear response surfaces and collinear factors, in their response surfaces. [50] While this is a desirable feature when studying complex systems in general, the GAM and MARS approaches incur a cost. They are more difficult to interpret, and they demand a great deal of computational effort compared to GLMs. GLMs offer flexibility with their choice of link functions demonstrating robustness to a variety of types of error, binomial, guassian, exponential among others. The small neighborhoods of RSM validate simple models but they are not immune to the catalog of error. Because of this fact, the low computational effort, and interpretability, MMRDP will employ GLMs with RSM and robust RSM.

General Linear Model Goodness of Fit

GLM fit can be assessed using statistical deviance. Deviance is a comparison the log likelihood, $L(\hat{\mu}, y)$ of saturated model to that of the model under assessment. [37] The saturated model maximizes the $L(\hat{\mu}, y)$ using a the most general model. The most general model has a perfect fit, $\hat{\mu} = \bar{y}$ and parameter for every observation. It explains the relationship and all variation in the data. [37] Fitted models have fewer parameters, a worse fit, and an associated deviance, equation 3.14.

$$-2[L(\hat{\mu}, y) - L(y, y)] = 2 \log \frac{\text{maximum likelihood for model}}{\text{maximum likelihood for saturated model}} \tag{3.14}$$

GLMs nest with the saturated model and greater deviance exhibits a worse fit.

3.4 Incorporating Robustness

The solutions found with MMRDP are robust in that they provide predictable response behavior despite the presence of exogenous factors. The concept of robustness is not fixed in the literature. The definition varies based on point of view and application, Roy (2002). [51] One can seek model robustness Mulvey et al.(1995). [52] and Vincke (1999). [53], algorithm robustness, Sorensen(2004). [54] or solution robustness, Rosenhead et al.(1972). [55], Kouvelis and Yu (1997) [56], Wong and Rosenhead (2000). [57] Decision analysis dominates the literature. Wong and Rosenhead (2000) see robustness as a measure of the useful flexibility preserving many options for choices to be made in the future.

The robust design methods can add a degree of response insensitivity to the exogenous factors. One may pursue risk analysis with RSM, Chen et al.(2003), and Al-Omar(2002). But with complex adaptive environment, this amounts to an insurmountable task. One cannot fully account for all possible outcomes and assign probabilities, as in classic application of risk analysis.

Mosteller and Tukey (1977) present the definition of a robust estimate most relevant to the research in two parts. [58]

- A small change in the data will not result in a significant change in the estimate.
- Across a range of inputs the estimate is efficient

The first point involves robustness to atypical data. The estimate is sufficient for the majority of the data. The second point refers to robustness to violations of assumptions. The estimate is relatively accurate without regard for the distribution of the data. [59]

MMRDP has three facets of robustness, in regression, parameter design, and local optima discrimination. Robust regression addresses violations of gaussian, homoscedastic, or independence assumptions. [31] Robust parameter design generates a model where the response is less sensitive to exogenous factor variance. [60] Local optima discrimination compares already robust solutions using the relative probability of meeting minimum response performance criteria.

3.4.1 Robust Regression

Statistical model extensions within RSM center on alternate forms of data regression, particularly deviations from OLS regression assumptions. Robust regression is used when encountering gaussian, homoscedasticity, or independence assumption violations. [31]

Classic RSM subsumes the assumptions of ordinary least square regression and Taylor series approximation.

- Normal(gaussian) response behavior with noise $\sim$ IID N(0, σ^2)
- Neighborhoods where experiments occur are small enough to view the unknown function, η, to be estimated and optimized as second order functions or lower

Normal response behavior allows for ordinary least squares regression and yields best linear unbiased estimator (BLUE) of $\widehat{Y}$. It also generates regression coefficients, $\overrightarrow{\beta}$, with gaussian uncertainty. Small experimental neighborhoods allow the user to initially view η as linear justifying first order models and reducing regression complexity and computing cost. [31] Even when presented with curvature, this assumption allows for low order Taylor approximations of η and preserves degrees of freedom. [31] Preserving degrees of freedom allows for estimation of variance and other moments beyond average estimated response value and allows fractional factorial experimental designs. It should be noted that Taylor series approximation does not address autocorrelation nor categorical factors.

Nongaussian behavior can be determined in several ways, such as a QQ plot, fig 3.3 or more precisely with a Shapiro-Wilks test for normality.

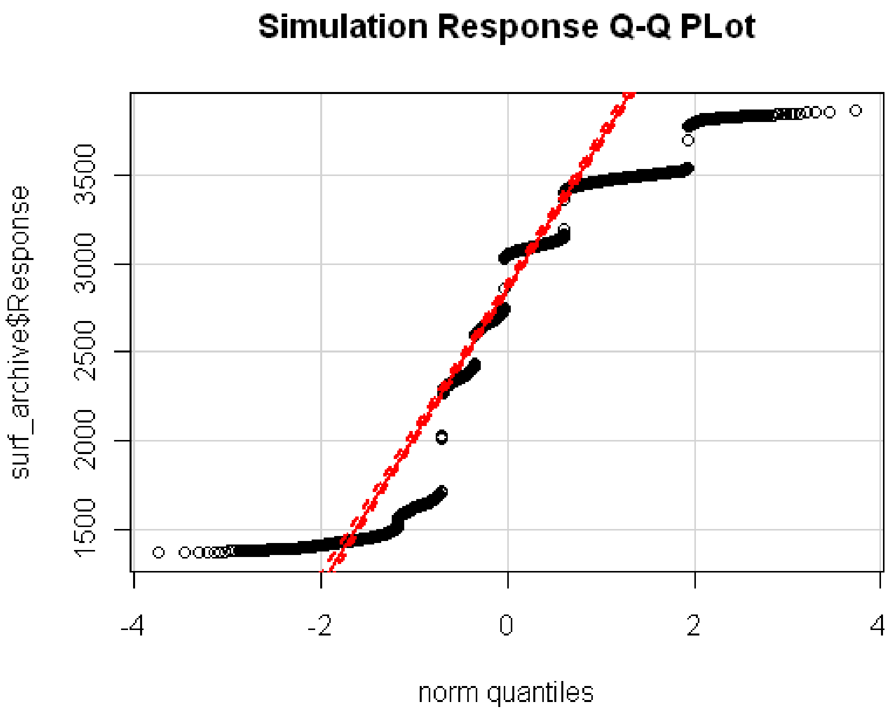


Fig. 3.3: A QQ plot shows nongaussian behavior.

The Shapiro-Wilks test null hypothesis is that data is normally distributed. Given an ordered set of random data , $x_{(i)}$, $I = 1..n$, a test is run using the W statistic. [61]

$$W = \frac{(\sum_i^n a_1 x_{(i)})^2}{\sum_i^n (x_{(i)} - \bar{x})^2} \tag{3.15}$$

$$a = (a_1, ..., a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}} \tag{3.16}$$

Here m' is the set of expected values of standard normal order statistics and V is the corresponding nxn covariance matrix.

W has no explicit form for a general level of n. However, tabular results are available for specific values of n. For instance, in the R Statistical Package, *stats*, the Shapiro-Wilks test function, $shapiro.test(x)$ work for $n = 3..5000$. [62] To apply the test with a given set of data, one compares a calculated statistic, W_{calc} to $W_{n,\alpha}$. If $W_{calc} < W_{n,\alpha}$, the null is rejected at the $1 - \alpha$ level, Figure 3.4.

```
 Shapiro-Wilk normality test

data:  surf_archive$Response[1:3000]
W = 0.899, p-value < 2.2e-16

>
```

Fig. 3.4: The results of the Shapiro-Wilk test on the response points classify it as nonGuassian behavior at the $\alpha = .05$ level.

When response data violates gaussian, homoscedastic, or independence assumptions, ordinary least squares regression is not the BLUE. Phenomena such as extreme values may be at play. Extreme values are not uncommon in complex dynamic systems and little data exists to fully characterize the environment or the relationships. [39]

In the presence of nongaussian behavior, once can use transformation of the response, such as a *log* transformation, bootstrapping, or robust regression. [31] Robust linear regression, does not minimize mean squared error like Ordinary Least Squares (OLS). It minimizes residual correlation. Common robust regression techniques are least trimmed squares (LTS), and least median of square (LMS). [63] These techniques lower estimated response variance compared to OLS but will not outperform OLS in the event of gaussian behavior. They are most suitable in the event of nongaussian behavior. [64]

Least Median Squares

LMS seeks to minimize the median, M of the residuals, $minM(e_i^2)$, instead minimizing the sum of squared differences in OLS, $min\sum_i (e_i^2) = min\sum_i (y_i - \sum x_{ij}\beta_j)^2$. [65] Application of the median within the optimization is in response the desire for less sensitivity to outliers within the data. [66]

Least Trimmed Squares

LTS is from the class of L-Estimators [59] It finds the minimum value of $\overrightarrow{\beta}$ by minimizing the summed, ordered residuals. LTS removes points, like extreme values, to reduce the variance of the set. When the sample set is not altered, the result from OLS is the same as LTS. LTS is preferred over LMS. LTS converges at rate $n^{1/2}$. LMS converges at a rate of $n^{1/3}$ [67]

3.4.2 Robust Parameter Design

In an environment with exogenous factors, poorly understood relationships, and lurking variables, the goal is to seek large regions of desirable system performance on the response surface in order to mitigate variance of the system. In robust parameter design, optimality is sought when some factors are not controllable, exhibit variance.

This research does not attempt to provide risk assessments to stakeholders. As a point of clarification, robust design is distinctly different than risk assessment. Risk assessment involves analysis of undesirable outcomes and their probability. [68] Robust design seeks to limit response sensitivity to variability in the factors. Given a complex system, one cannot purport to know all of the outcomes associated with it. [45] Amaral and Ottino (2004) stress that the nature of systems with emergent outcomes negate comprehensive risk assessment efforts. Thus, robust design is a more appropriate aim. It allows users to seek trade offs between variance behavior, and response behavior, Figure 3.5.

Robust design was developed by Genichi Taguchi to minimize process variance. Focusing on production, he splits factors that impact a system in two categories, control and noise. [60] Those things manufacturers can control, such as labor on hand or design dimensions are control factors. Those they cannot control, such as the weather or the time it takes for a chemical process to complete, are exogenous noise. For instance, variations in air density may impact the adhesion of parts in a bonding process. A familiar extension of Taguchi's work is the Lean Six Sigma process championed by General Electric and other companies. [69]

Taguchi's robust design provides solutions by modeling a process and the

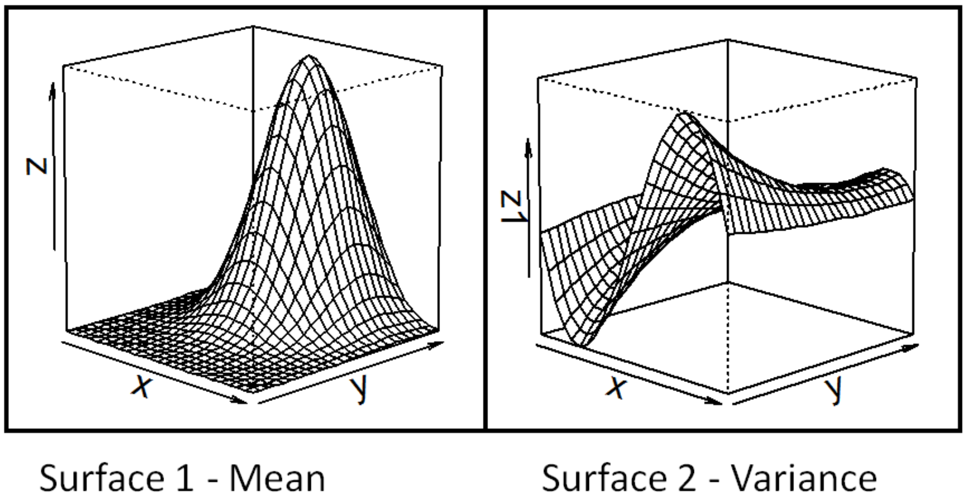


Fig. 3.5: Robust parameter design uses mean and variance surfaces to tune response behavior through control factor settings.

variables accordingly. His methods find optimal settings for desired outcomes then reduced noise sensitivity in that response by tuning the overall process. Taguchi's method centers on a two piece 2^k experimental design. [60] Set orthogonal to each other, an inner array contains control variables and an outer array holds the exogenous ones. Additionally, he combines response mean and variance into one response, signal to noise ratio (SNR). SNR is Taguchi's method to seek the estimated response mean and minimize variance. [32]

Smaller is better - response minimization

$$\text{SNR} = -10log_{10}[\frac{\sum y_i^2}{n}]$$

Larger is better - response maximization

$$\text{SNR} = -10log_{10}[\frac{\sum \frac{1}{y_i^2}}{n}]$$

Nominal the best - response target

$$\text{SNR} = 10log_{10}[\frac{\bar{y}^2}{n}]$$

Due to the unique construction of each SNR, each is maximized. While Taguchi's method was a leap forward in process control, it has its faults. Its failure to consider interactions, the high required number of experiments, and the SNR's insensitivity to factors that impact mean or variance separately motivated more research. [32]

Since Taguchi's development, Welch, et al. (1990) [70] and Shoemaker, Tsui, and Wu (1991) [71] found more efficient methods of experimental design to do the same, primarily by combining the noise and control variable arrays. Their research solidified the idea that one design can generate response surfaces for both the mean and the variance, a dual response problem. In this form, one can solve for a relationship using the response mean (minimize, maximize, or target) then use that as an additional constraint when minimizing the variance. [32] As a extension, RSM is applied using robust design, per Vining and Myers (1990) [72] and Myers, Khuri, and Vining (1992). [73]

As in classic RSM, function smoothness is not an issue in dual surface RSM. Nelder and Lee (1991) [74], and Engel and Huele (1996) [69] validated generalized linear models (GLM's) in robust design. One is not limited to GLM's, a number of other robust design techniques that are more specific to computer simulation been addressed by Sanchez (2000) [75], and Atkinson et al. (2007). [76]

As extension of model 2.1, one can separate the input variables into control, x, and exogenous(noise), z, variables. Accounting for control interactions and polynomial functions is g(x). β are the coefficients for the control variables, δ are the coefficients for the noise variables and $\triangle$ are the coefficients for interactions. [32]

$$\text{Estimated Function: } y(x,z) = \beta_0 + g(x)'\beta + z'\delta + g(x)'\triangle z + \epsilon \tag{3.17}$$

$$\text{Response Expectation: } E[y(x,z)] = \beta_0 + g(x)'\beta \tag{3.18}$$

$$\text{Response Variance: } Var[y(x,z)] = [\delta' + g(x)'\triangle]\mathrm{Var}[z][\delta' + g(x)'\triangle]' + \sigma_\epsilon^2 \tag{3.19}$$

Dual Response Surface Optimization Framework

Objective Function: minimize or maximize $E[y(x, z)]$

Subject to

$min_i \leq x_i \leq max_i, \forall i, i = 1..n$

Var[y(x,z)] $\leq \sigma^2_{ceiling}$

Optimization within MMRDP

Algorithmically, the optimization problem is separate from the RSM procedure. Once there is suitable data for both surfaces, it can be solved using software such as the Generalized Algebraic Modeling System,(GAMS), Gurobi, or ILOG CPLEX. [77–79] They allow for objective function relaxation and preprocessing for linear and nonlinear optimization including continuous, integer or binary variables.

3.4.3 Robust Setting Discrimination

Exclusively seeking optimality on a response surface is a limiting effort. Stakeholders, particularly in an environment with exogenous factors, may find it valuable to define the landscape more fully, not solely by locally optimal points. MMRDP extends RSM as a final step through optimal point discrimination.

Given a set of local optimal points found using RSM, A^*, each point $a_1^*...a_f^*$ has $P - 1$ dimensions and a corresponding univariate expected response value $y_1^*...y_f^*$ that comprise the P dimensional response surface. In a stochastic environment, the response value also has a confidence interval due to exogenous behavior.

Estimated $(1-\alpha)\%$ Confidence Interval

$$\widehat{\bar{y}}_i \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{s_{y_i}}{n}} \tag{3.20}$$

For gaussian behavior, Law and Kelton (1999) and Bland and Altman (1996) among many others recommend basic $(1-\alpha)\%$ confidence interval construction structured with the estimated mean, $\bar{Y}(n)$, and estimated variance $S^2(n)$ using n data points. Based on the data available, the *student-t statistic* or standard normal value, z, is used as appropriate when constructing the half width. [22]

The intervals assume gaussian and IID error due to OLS regression models. Despite the presence of nongaussian behavior, one can always frame confidence intervals based on gaussian behavior within MMRDP. The simulation generates j multiple independent replications per design point. As such, each optimal point has an average response value, $\bar{y}_i$. By the central limit theorem (CLT), $\frac{1}{s_i}\sum_{k=1}^{j}(y_{ij}-\mu_{ij}) \sim N(0,1)$ which implies that $\bar{y}_i \sim N(\mu_{y_i}, \sigma_{y_i}^2)$. [22]

MMRDP includes a robust design filter to incorporate a more sophisticated account of uncertainty in the RSM optimal estimates, aid in discriminating between the optima. Comparing points via confidence intervals,

$(1-\alpha)\%$ may imply lack of a statistically significant differences between courses of action. The confidence regions around each locally optimal point may overlap negating the use of paired-t or similar tests. [22]

Local Optima Filter Via Satisficing Bound Surface Area Estimation

The local optima filter is an alternate method that can be used to discriminate between courses of action with overlapping confidence intervals. Both classic and dual surface RSM can produce multiple local optima. All solutions form a set, each with a mean and variance, typically gaussian due to central limit theorem, Figure 3.6. Confidence interval overlap can occasionally be overcome via a paired-t or Tukey test in order produce an ordering within the set. [22] However, confidence intervals assume gaussian behavior, constant variance, and are not conservative in that they do not include the satisficing bound desired by stakeholders.

During stakeholder analysis, the desired response behavior will be quantified by bound. It is sensible to seek a range, rather than a point estimate of desired behavior. In the case of a maximization problem, the stakeholders should identify minimum, or satisficing, behavior that serves as both screening criteria and a lower bound for the desired response. RSM is a nonconverging sequence and can provide locally optimal multiple points by selecting varied initial points. [80] Assuming that more than one optimal point will be found through multiple applications of RSM within MMRDP, each point will have an associated satisficing area in the P dimensional space. From there,

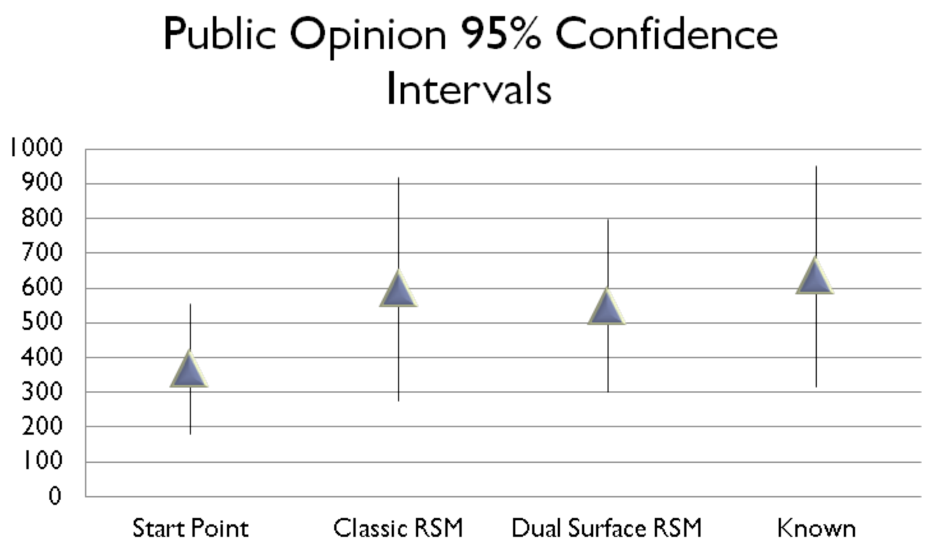


Fig. 3.6: The stochastic nature of the response can be represented with 2 sided, symmetric confidence intervals which often overlap and make them indistinguishable.

it is necessary to quantify the areas and compare them in order to rank order the estimates and associated settings.

This facet of robustness is valuable when there exists low fidelity system understanding. Figure 3.7, shows how the topology of a response surface can serve as an alternate method to organize locally optimal points by quantifying the size of the associated satisficing area.

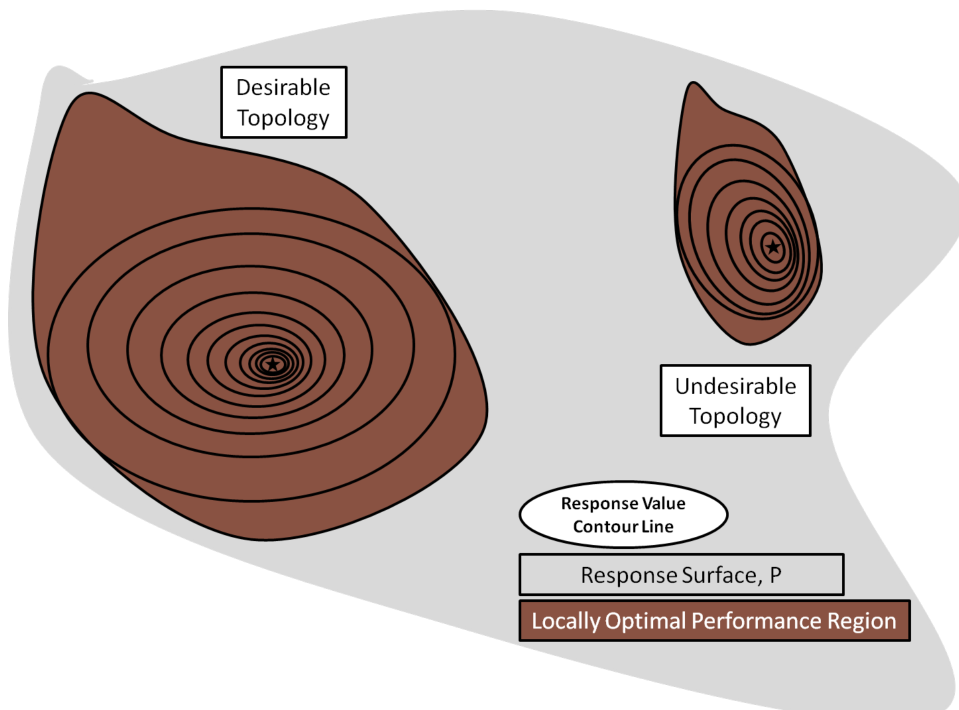


Fig. 3.7: Though both the left and right points are optimal, the left point is preferable due to the larger area of satisficing response behavior.

As an additional step within robust methods, one can consider the effective area of acceptable response behavior as a method to discriminate between optima in the presence of exogenous factors. On a shared response surface, each optima will have a perimeter outlined by acceptable response behavior. Among optima with overlapping confidence intervals, the area associated with each optima due to exogenous factors facilitates a ranking, where one can select the course of action (control factor settings) with the largest area as the best solution.

This method offers two benefits. It maximizes the probability of desired response behavior in the complex system of interest. It also provides an additional method to discriminate between solutions that may have overlapping response value confidence intervals.

Surface area is easy to understand and measure in three dimensions. This makes it a good tool to use for collaboration with stakeholders. Satisficing areas are the topological space equivalent of performance trade off. For example, it is easier to land a helicopter on a mesa than a spire. The spire may yield better response performance but that must be balanced against the relative probability associated with a larger versus a smaller area.

In this light, the probability of achieving success with a locally optimal point is directly proportional to its associated surface area. Large areas of suitable, not best, behavior should be sought when simulating poorly understood and complex systems because they represent greater a probability of success. MMRDP seeks to provide response behavior greater than a predetermined floor value as a substitute for point maximization. Conversely, one seeks response behavior less than or equal to a predetermined ceiling as a substitute for a point minimization.

The response surface will be composed of some estimate function, $y =$ *f(x,z)*; presumably, at a point of inflection. In the instance of seeking a maximum, ideally one is on a peak; for a minimum, the opposite. In general, intercept points for the desired satisficing behavior radius on *f(x,z)* can be found by solving the quadratic form of the function relative to the acceptable performance level. However, MMRDP only assumes functions are valid in as small neighborhood. Experimentation to define the space in necessary. Once complete, there exist the additional task of approximating another function to represent the satisficing bound for each optima.

Assume hyper-cube surface area was sought to approximate the satisficing area. The hyper-cube requires simulation of $2*N-1$ times to generate limits in $2*N-1$ directions of the $N-1$ exogenous dimensions. It requires $Ksteps$ from the optimal point $*2*N-1$ dimensions $*R$ replications. It is labor intensive but as a full exploration, it does not assume symmetry about the optimal point. The hypercube has the advantage of greater accuracy in that it is a result of fuller exploration of the space.

An alternative and quicker method to outline limits would be to follow the path of of Greatest Degradation (PGD) until the boundary of the acceptable response boundary (ARB) is found. The path greatest deterioration (PGD) is the opposite of the path of steepest improvement. The Euclidian distance between the optimal point and closest boundary point forms a radius, r. Surface area is determined by the surface area of a hyper-shape with radius R. This method assumes symmetry about the optimal point.

The PGD moves from the optimal point to the satisficing boundary uncovering the shortest path to the least desirable behavior. Using the β coefficients of *f(x)*, one can create the PGD and pursue it for some distance until suboptimal performance is found. The PGD is defined as the negative equivalent of the PSA / PSD as appropriate.

Using PGD, the practitioner will arrive at a point,G, at the boundary of satisficing performance. The difference between the start point, X^* and G is r, a $P-1$ dimensional vector of $P-1$ radii used to define the space.

Most obvious in 3-space, surface area of a function is one dimension less than volume. This holds for P dimensions as well. For smooth functions, volume is found via P integrals of the function and surface area one fewer. [81] In P dimensional space is there are P dimensional shapes with associated hyper-volumes and $P-1$ dimensional hyper-surface areas.

$$\begin{aligned}
\text{Volume (3)} &= \int\int\int (f(x_1, x_2))dx_1dx_2dy \\
\text{Volume (P)} &= \int_{x_1} ... \int_{y} (f(x_1, ..., x_{p-1}))dx_1...dx_{p-1}dy \\
\text{Surface Area (3)} &= \int\int (f(x_1, x_2))dx_1dx_2 \\
\text{Surface Area (P)} &= \int_{x_1} ... \int_{x_{p-1}} (f(x_1, ..., x_{p-1}))dx_1...dx_{p-1}
\end{aligned}$$

- Hyper-cube: $(-r_1, r_1)...(-r_{p-1}, rp-1)(-y, y)$

- Hyper-sphere: $\frac{x_1}{r} + ... + \frac{x_{p-1}}{r} + \frac{y}{r} = 1$

- Hyper-ellipsoid: $\frac{x_1}{r_1} + ... + \frac{x_{p-1}}{r_{p-1}} + \frac{y}{r_p} = 1$

Hyper-cubes (n-cubes) do not have an integrable function. Hyper-dimensional measures are based on the number of dimensions (P), edges($(P-1)2^{P-1}$), their lengths($2r_1...2r_{p-1}2y$), and the number of faces ($2P$). P space hyper-volume is $\prod_{i=1}^{P} r_i$ and hyper-surface area is summation of the hyper-area of each face. A 3-cube surface area is the sum of six 2-cube surface areas. A 4-cube (tesseract) surface area is the sum of eight 3-cube surface areas. A P-cube surface areas is the sum of $2P$, $(P-1)$-cube surface areas. Hyper-cubes can be problematic as an estimator due to their computational burden.

The orthogonal vertices of the hyper-cube can cause over estimation of hyper-area. [82] Hyper-spheres effectively interpolate cube edges and vertices providing a more conservative estimate of surface area. It is assumed that within the application of MMRDP, conservative recommendations are most desirable due to an awareness of complexity within systems.

Hyper-spheres have an integrable function that yields a P-space volume and surface area. [81, 83]

$P = 2k$ Hyper-sphere Volume

$$V_P Sphere = \frac{\pi^k}{k!} r^{2k} \tag{3.21}$$

$P = 2k$ Hyper-sphere Surface Area

$$SA_P Sphere = \frac{k-1!}{(2k-1)!} 2^{2k-1} \pi^{k-1} r^{2k-1} \tag{3.22}$$

$P = 2k + 1$ Hyper-sphere Volume

$$V_P Sphere = \frac{k!}{(2k+1)!} 2^{2k+1} \pi^{k} r^{2k+1} \tag{3.23}$$

$P = 2k + 1$ Hyper-sphere Surface Area

$$SA_P Sphere = \frac{k!}{(2k)!} 2^{2k} \pi^{k} r^{2k} \tag{3.24}$$

The radius r is defined as the euclidian distance from the local optimal point to the edge of satisficing behavior found via PGD. The radius, r, is a vector of dimension P. A radius represents an average length for all dimensions, Figure 3.8. A higher fidelity representative shape is the hyper-ellipsoid as it explicitly represents the behavior bounds in each of P dimensions figures, 3.9 3.10. Like the sphere, the ellipse function is smooth, continuous, and integrable. This method assumes symmetry about the optimal point. [84] [85]

While analytically tractable, the function is not reliably defined across broad areas given complexity, a lack of homoscedasticity, and the lack of fidelity from small neighborhood taylor series approximations. However, the presence of bounds facilitates substitute $P-space$ functions for $f(r)$. [81,83]

Figure 3.10

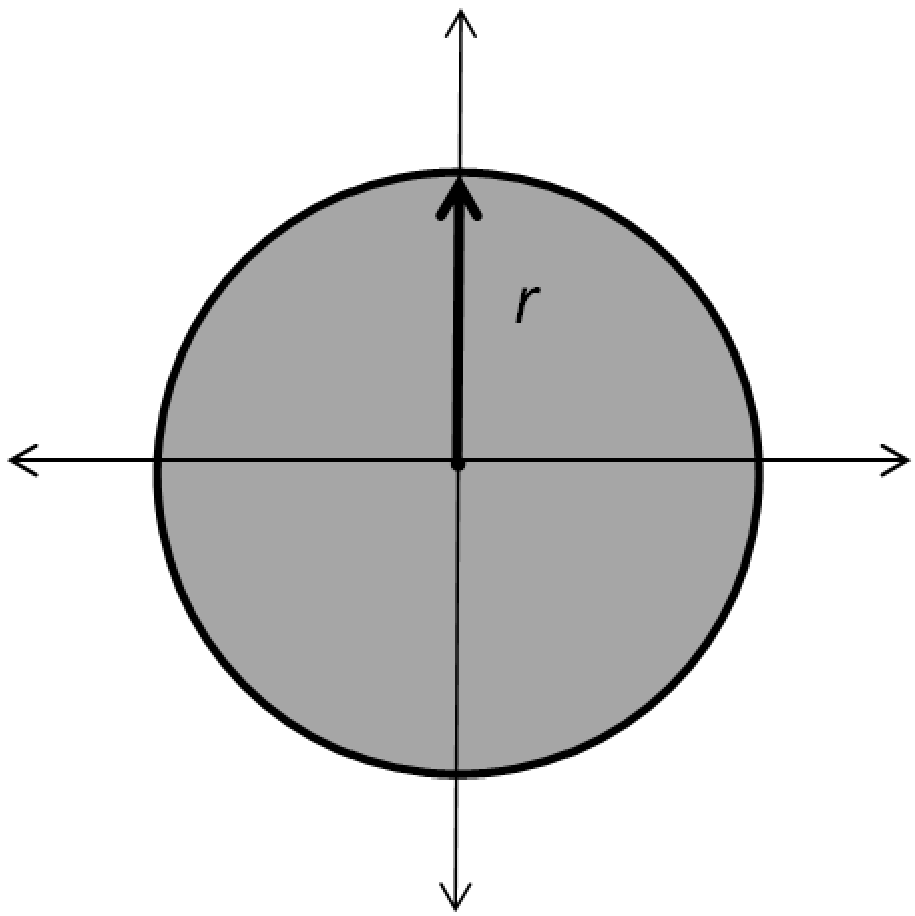


Fig. 3.8: Circle in two dimensions in standard form, $x^2 + y^2 = r^2$, with surface area, $\pi \times r^2$.

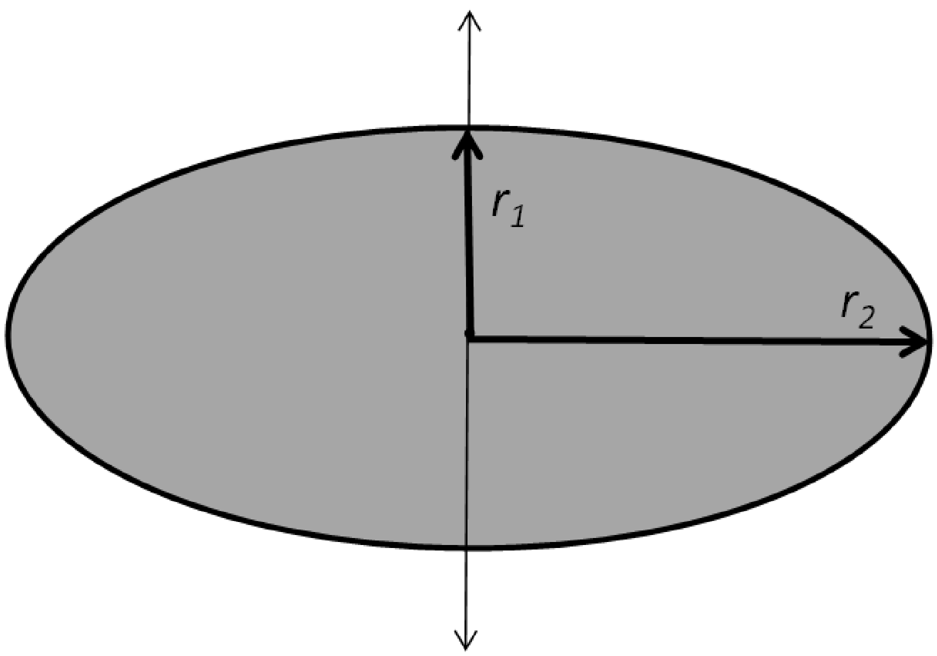


Fig. 3.9: Ellipse in two dimensions in standard form, $\frac{x^2}{r_1^2} + \frac{y^2}{r_2^2} = 1$, with surface area, $\pi \times r_1 \times r_2$

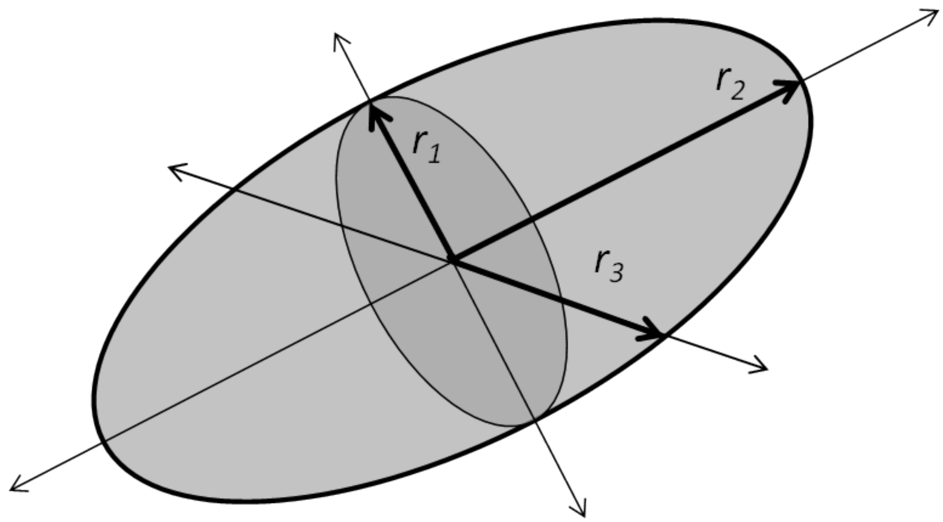


Fig. 3.10: Ellipsoid in three dimensions in standard form, $\frac{x^2}{r_1^2} + \frac{y^2}{r_2^2} + \frac{z^2}{r_3^2} =$, with surface area, $\int\int(1 + \frac{dz}{dx}^2 + \frac{dz}{dy}^2)^{\frac{1}{2}} dxdy$

3 Dimension Ellipsoid Function

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{3.25}$$

3 Dimension Ellipsoid Volume

$$\int\int\int(1 + \frac{dz}{dx}^2 + \frac{dz}{dy}^2)^{\frac{1}{2}} dxdydz \tag{3.26}$$

3 Dimension Ellipsoid Surface Area

$$\int\int(1 + \frac{dz}{dx}^2 + \frac{dz}{dy}^2)^{\frac{1}{2}} dxdy \tag{3.27}$$

N dimension Ellipsoid Function

$$\frac{x_1^2}{a_1^2} + ... + \frac{x_{n-1}^2}{a_{n-1}^2} + \frac{z^2}{a_n^2} = 1 \tag{3.28}$$

N Dimension Ellipsoid Volume

$$\int^{x_1} ... \int^{z} (1 + \frac{dz}{dx_1}^2 + ... + \frac{dz}{dx_{n-1}}^2)^{\frac{1}{2}} dx_1 ... dx_{n-1} dz \tag{3.29}$$

N Dimension Ellipsoid Surface Area

$$\int^{x_1} ... \int^{x_{n-1}} (1 + \frac{dz}{dx_1}^2 + ... + \frac{dz}{dx_{n-1}}^2)^{\frac{1}{2}} dx_1 ... dx_{n-1} \tag{3.30}$$

This closed form solution to find surface area is tractable but not computationally efficient due to the generation of up to $P-1$ differentiations of the functions as well and $P-1$ integrals of their sum.

Local Optima Filter Demonstration

As an example of the value of the optima filter, consider Montgomery's semiconductor process problem. [86] Montgomery (1999) provides function estimates, based on data, for the response mean and response variance of a process with two control variable (X_1, X_2) and three exogenous variables (Z_1, Z_2, Z_3).

Response Mean Model

$$30.37 - 2.92X_1 - 4.13X_2 + 2.6X_1^2 + 2.18X_2^2 + 2.87X_1X_2 \tag{3.31}$$

Response Variance Model

$$19.26 + 3.2X_1 + 12.45X_2 + 7.52X_1^2 + 8.52X_2^2 + 2.21X_1X_2 \tag{3.32}$$

General Response Model

$$\begin{aligned}30.37 - 2.92X_1 - 4.13X_2 + 2.6X_1^2 + 2.18X_2^2 + 2.87X_1X_2 \\ + 2.73Z_1 - 2.33Z_2 + 2.33Z_3 - .27X_1Z_1 + 100X_1Z_2 \\ + 2.58X_1Z_3 + 2.01X_2Z_1 - 1.43X_2Z_2 + 1.56X_2Z_3\end{aligned} \tag{3.33}$$

Assume there is a goal to minimize the response and that the stakeholders set a satisficing bound where acceptable behavior is less than or equal to 40. In this instance, the exogenous variables are distributed according to a gaussian distribution with a mean of 0 and a variance of 1. The mean response surface solution is the global maximum, 28.41, with a variance of 22.62. The dual (mean and variance) surface optimization solution minimizes the variance, 14.62, and finds a mean response value within the statisficing bound, 35.01, Figure 3.11.

The 95% confidence interval for both solutions overlap and cover the satisficing bound, Figure 3.12. By fixing the control settings and simulating

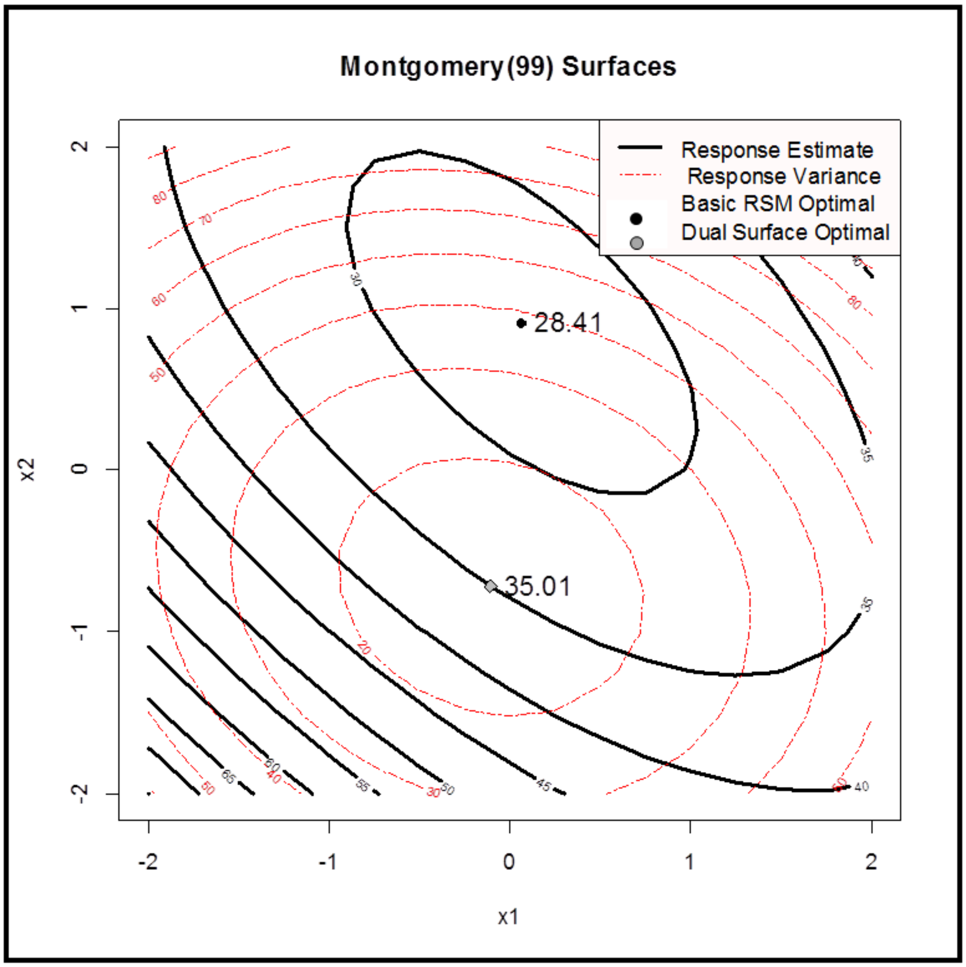


Fig. 3.11: In an overlay of the mean and variance contour plots, the single surface and dual surface optimal points are identified.

the general function with the range of the exogenous factors, PGD's are generated to find the minimum distances to the satisficing bounds via hyper-ellisoid surface area for each. The dual surface optimal point surface area is less than the single surface optimal point, 2.43 vs 5.35. In this instance, traditional methods would be sufficient to select the optimal point.

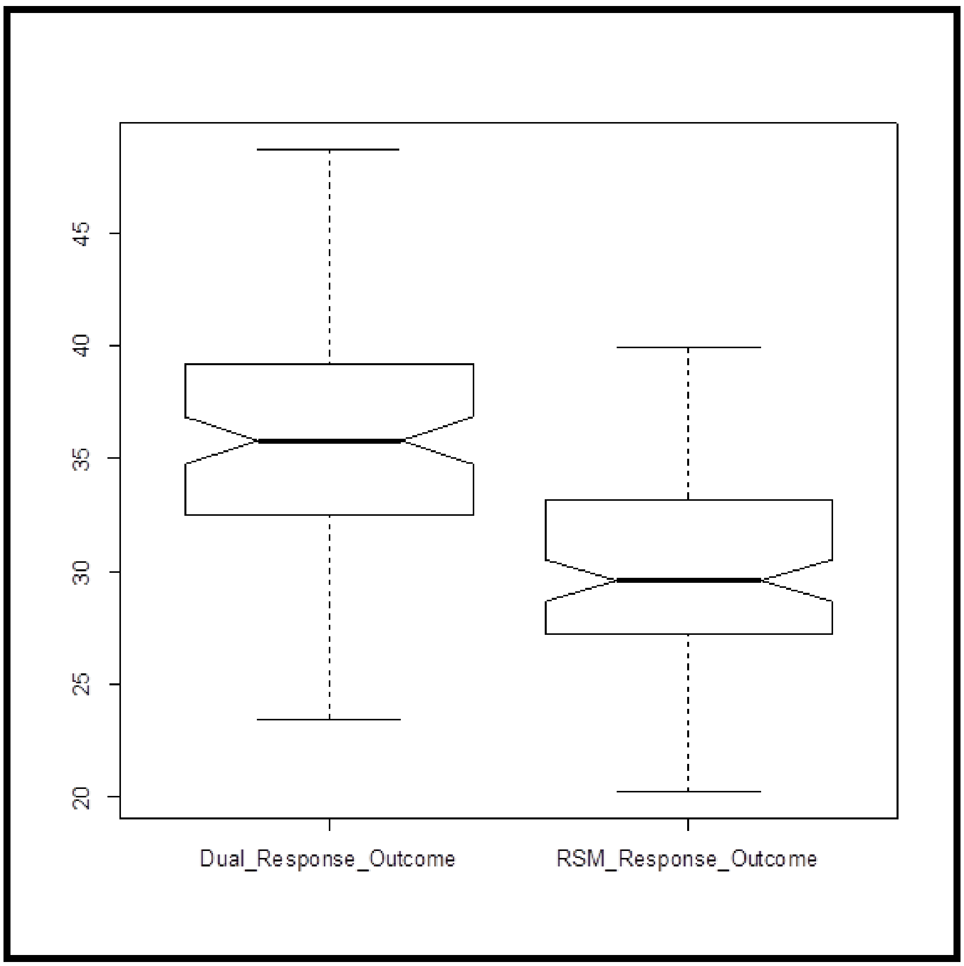


Fig. 3.12: The confidence intervals overlap.

However assume all parameters stay the same but the variance of the exogenous factors is increased from 1 to 10. The two optimal points are the same location but with higher variance than before, 28.41 vs 35.01, Figure 3.13.

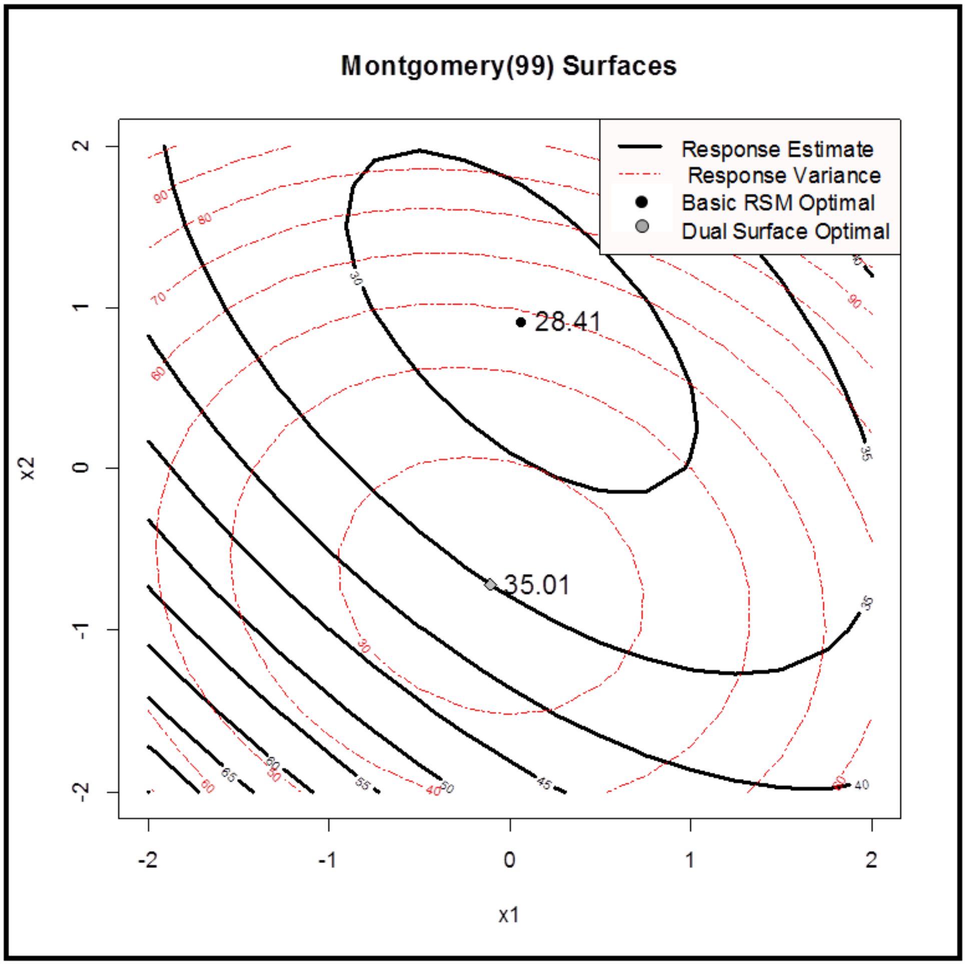


Fig. 3.13: In an overlay of the mean and variance contour plots, the single surface and dual surface control settings are the same as before but with higher variance.

As before, the 95% confidence interval for both solutions overlap and cover the satisficing bound, Figure 3.14. By fixing the control settings and simulating the general function with the range of the exogenous factors, PGD's are generated to find the minimum distances to the satisficing bounds via hyper-ellipsoid surface area for each. The dual surface optimal point surface area now much greater than the single surface optimal point, 596.75 vs 324.48. In this instance, traditional methods would not serve the conservative nature of stakeholders

The measuring the satisficing bounds is a valuable way to discriminate between local optima because is does not assume gaussian behavior, it uses direct simulation data to better encapsulate high order functions, and it does not assume constant variance. In the event that stakeholders strongly desire an understanding of how a course of action compares to another in terms of violating a bound because of exogenous variable behavior, the local optima filter is viable and valuable. Assuming the data already exists, it is rapid. This is and important and beneficial property because the filter shows no generalizable results. The combination of mean response value, response variance level, statisficing bound, and the underlying relationships drive the results of the satisficing surface area calculations and subsequent ordering.

Numerical Approximation of Hyper-Ellipsoid Surface Area

A P dimension ellipsoid surface area approximation developed by Knud Thomsen (2004) simplifies the calculation and is scalable. [87] The Thom-

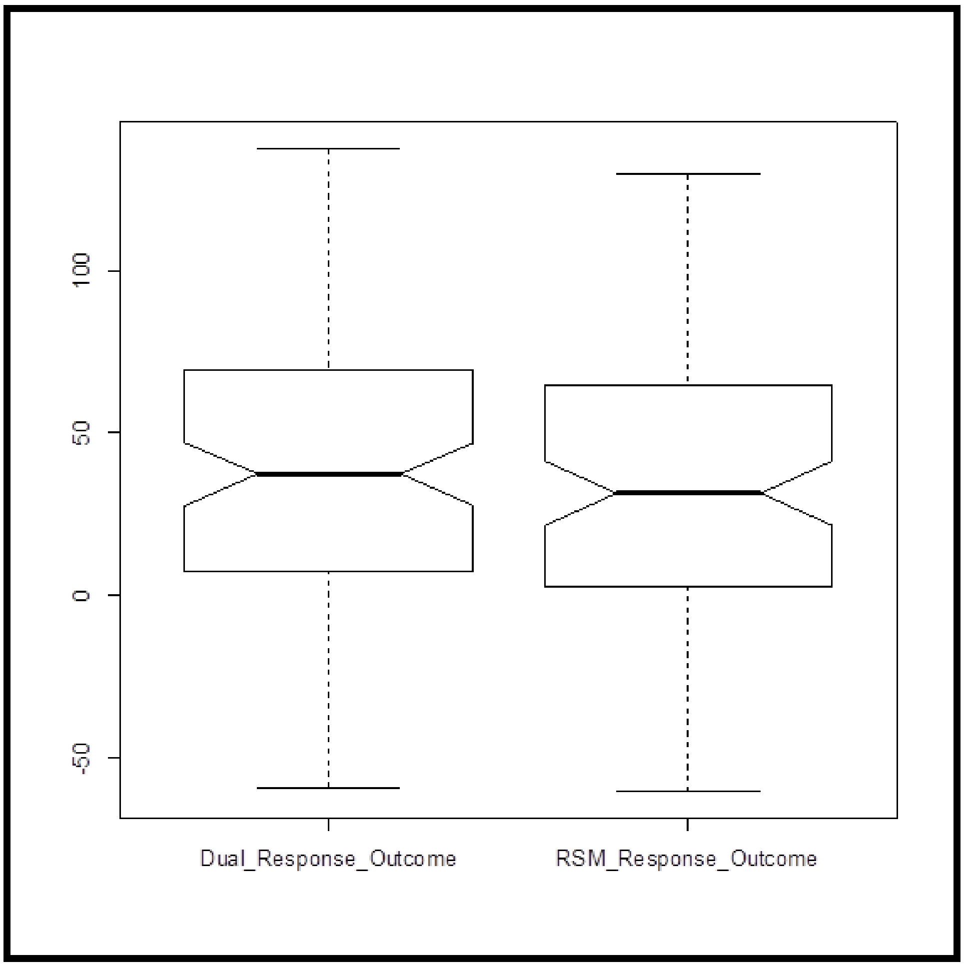


Fig. 3.14: The confidence intervals overlap.

sen (2004) numerical approximations are available based on the surface area function for a hyper-sphere and the Hölder mean.

Hölder Mean / Power Mean : for any exponent p the hölder mean is the quantity H whose pth power is the arithmetic mean of the p powers of the quantity under consideration. [88] Thomsen (2004)recommends a value for p. [87] where $p \to 2$ for large n.

$$H = H_p = (a_1^p + \ldots + a_{n-1}^p)^{\frac{1}{p}} \tag{3.34}$$

$$p = \log n / \log A \tag{3.35}$$

$$A = \frac{\sqrt{\pi}\Gamma(\frac{n}{2} + \frac{1}{2})}{\Gamma(\frac{n}{2})} \tag{3.36}$$

Case $N = 2k$ (n even)

$$A = \frac{\pi(2k-1)!!}{2^k(k-1)!} \tag{3.37}$$

Case $N = 2k + 1$ (n odd)

$$A = \frac{2^k k!}{(2k-1)!!} \tag{3.38}$$

The !! operator is a double factorial where $n!! = n(n-2)(n-4)...$, the product of all positive integers up to n, with the same parity as n. [89]

$0!! = 1, 1!! = 1, 2!! = 2, 3!! = 3, 4!! = 8, 5!! = 15, 6!! = 48, 7!! = 105...$

P = 2k Dimension Surface Area of a hyper-sphere

$$\frac{2\pi^k r^{2k-1}}{\Gamma(k)} \tag{3.39}$$

where $\Gamma(k) = (k-1)!$

The approximation of surface area for a hyper-ellipsoid substitutes the Hölder mean for the radius element (r^{n-1}) in the hyper-sphere surface area, equation 3.39.

Thomsen's approximation of hyper-spheroid surface area

$$\begin{gathered}\mathbf{N = 2k}\\ \frac{2H_{KT}\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\\ \mathbf{N = 2k{+}1}\\ \frac{2n^{\frac{1}{p}}H_{KT}\pi^{\frac{n-1}{2}}}{\Gamma(1+\frac{n-1}{2})}\end{gathered} \tag{3.40}$$

Hölder Mean for hyper-ellipsoid

$$H_{KT} = \prod_{j=1}^{n} a_j \left(\frac{\sum_{i=1}^{n} a_i^{-p}}{n}\right)^{\frac{1}{p}} \tag{3.41}$$

The Hölder Mean used in Thomsen's ellipsoid estimation (H_{KT}) is equivalent to the Hölder Mean (H_p) as previously defined.

$$H_p = \prod_{j=1}^{n} a_j (\frac{\sum_{i=1}^{n} a_i^{-p}}{n})^{\frac{1}{p}} = (\frac{\sum_{i=1}^{n} a_i^{p}}{n})^{\frac{1}{p}}$$

$$H_{KT} = \prod_{j=1}^{2} a_j (\frac{\sum_{i=1}^{n} a_i^{-p}}{n})^{\frac{1}{p}}$$

let n = 2

$$a_1 a_2 (\frac{a_1^{-p} + a_2^{-p}}{2})^{\frac{1}{p}}$$

$$(a_1^p a_2^p \frac{a_1^{-p} + a_2^{-p}}{2})^{\frac{1}{p}}$$

$$((\frac{a_1^p a_2^p}{a_1^p} + \frac{a_1^p a_2^p}{a_2^p})(\frac{1}{2}))^{\frac{1}{p}}$$

$$(\frac{a_2^p}{2} + \frac{a_1^p}{2})^{\frac{1}{p}}$$

let n = N

$$(\frac{a_n^p}{n} + ... + \frac{a_1^p}{n})^{\frac{1}{p}} \Rightarrow$$

$$H_{KT} = (\frac{\sum_{i=1}^{n} a_i^{p}}{n})^{\frac{1}{p}} = H_p$$

Pareto Optimality

A final tool to assist with the decision is a Pareto Analysis. Despite the noneconomic core of MMRDP, finances must be considered in conjunction with critical factors. A Pareto chart allows stakeholders to visually identify the Pareto efficient curve and select the desired combination of cost and performance. [11]

3.5 Properties of Methodology

MMRDP is best suited for poorly understood environments with complex features. It offers a range of solutions in the face of nebulous situations with long planning horizons and significant resource investment requirements. The timeline adds to complexity and large investments justify the extensive front-end cost of MMRDP full implementation. Stakeholder analysis, model development, and simulation demand indigenous knowledge of the problem, time, money and technical competency.

Curse of Dimensionality

MMRDP is used in high dimension spaces and subject to the curse of dimensionality. [90] In sample spaces, the volume grows with each added dimension. The sample landscape becomes sparse and variance inexorably increases unless the data set increases as well. For instance, normally distributed data exhibits reduced kurtosis and heavier tails as dimensions grow. [82] The growth in the space is exponential and has the immediate effect of requiring significantly more experiments to maintain proportional spacial and counter the growth, Figure 3.15.

This effect of high dimension is evident in linear predictor MSE($\hat{Y}$),

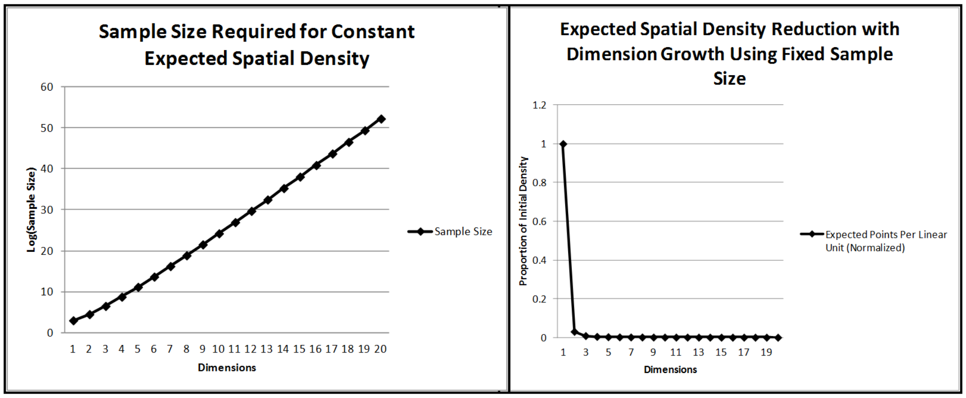


Fig. 3.15: As dimensionality increases, there is an exponential impact on sample size and spatial density.

$E[(\hat{Y} - Y)^2]$. OLS regression estimates are unbiased in univariate and multivariate linear relationships with gaussian error. [46] Despite lack of bias, MSE climbs proportionally with dimension in OLS regression. The unavoidable increase in variance is the cause, $\text{MSE}(\hat{Y}) = VAR(\hat{Y}) + Bias(\hat{Y}, Y)^2$, Figure 3.16. [46]

Due to the curse, stakeholders applying MMRDP must absorb experimentation cost, primarily time, due to the increasingly smaller neighborhoods where homoscedasticity and linear assumptions hold. Sparse experimental designs become less desirable as dimensions increase.

The response and factor relationship change across increasing dimensions. Application of local optima filter methods via surface area demand comparison using normalized values to negate scale influence and always across the same number dimensions.

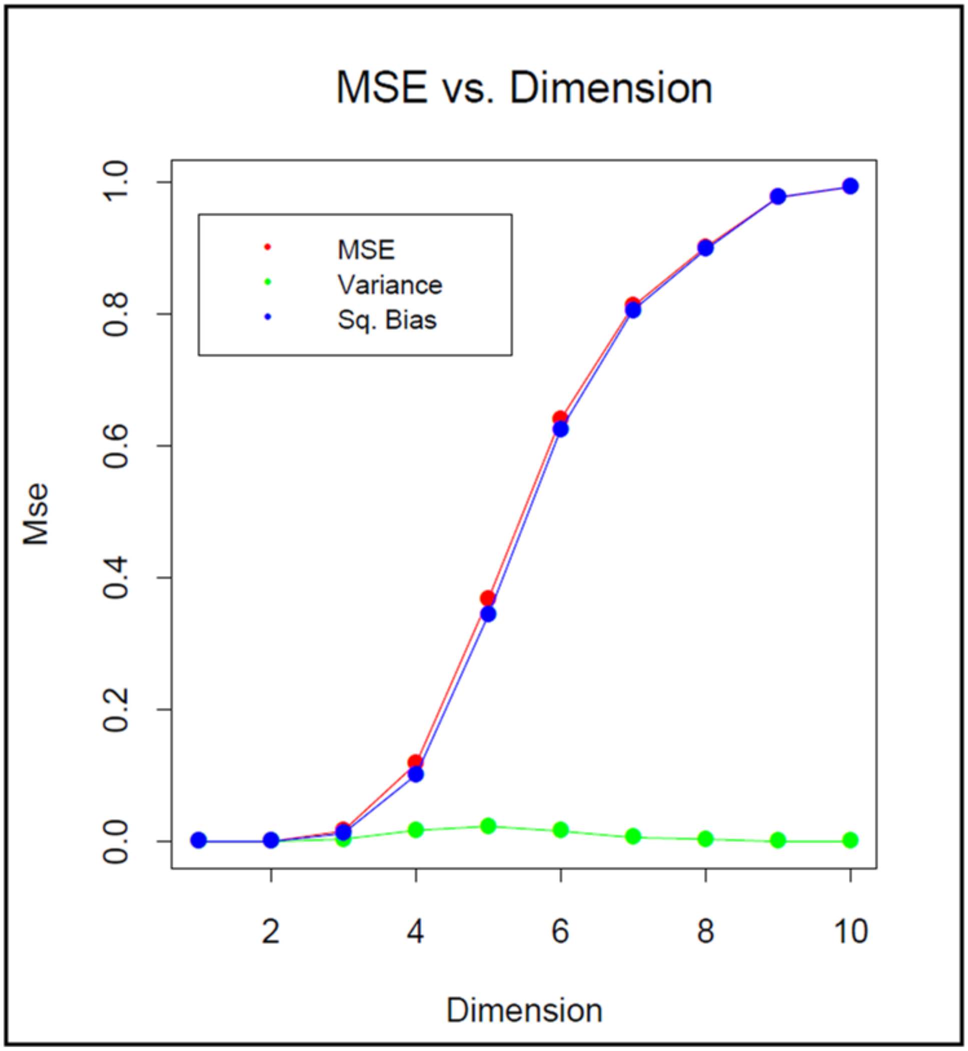


Fig. 3.16: Ordinary Least Squares estimators are unbiased. The variance increase from dimensionality increases mean squared error.

Local Optima Filter Conservative Performance Area Estimation

The strategy of selecting a function as an approximation for the surface area is problematic in that it deliberately induces error. However, it maintains the theme of robustness, via a conservative approach that reflects stakeholder desires. Given the type of environment suitable for application of MMRDP, underestimation is acceptable and tantamount.

Assume continuity exists on the response surface in the neighborhood surrounding the optimal point. The combination of PGD discovery of the satisficing boundary and hyper-ellipsoid function substitution will under estimate surface actual area.

An example is provided using PGD to estimate the area of satisficing behavior in two a dimension space given the desired minimization of a function, $f(x)$, and an optimal point, $X*$.

The 2 dimension symmetric case, Figure 3.17:

$x* = 10$

$f(x) = (x - 10)^2 + 50$

Satisficing boundary: $f(x) \leq 100$

Known area (orange and blue): 470.4

Estimated area (blue): 353.6

Area Underestimate: 116.8

The 2 dimension asymmetric case, Figure 3.18:

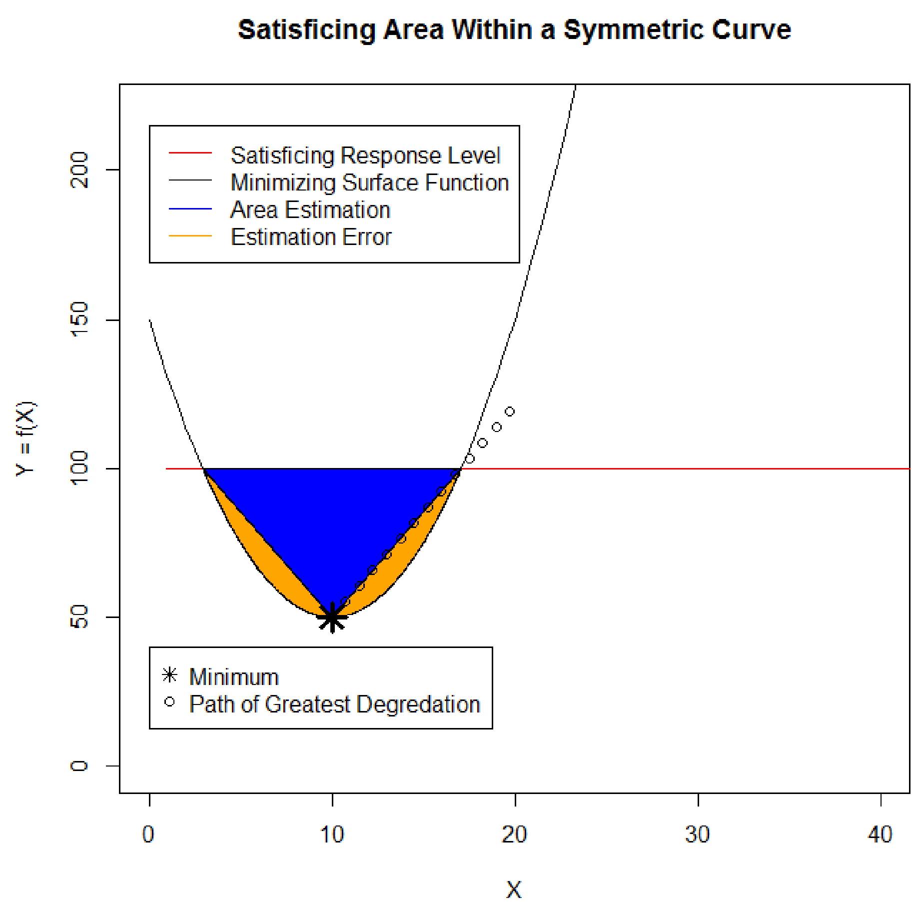


Fig. 3.17: The path of greatest degradation method ensures underestimation of surface area in the presence of symmetric functions.

$x* = 10$

for $x \leq x*$

$f(x) = (x - 10)^2 + 50$

for $x > x*$

$f(x) = \frac{(x-10)^2}{6} + 50$

Satisficing boundary: $f(x) \leq 100$

Known area (orange and blue): 813.1

Estimated area (blue): 353.6

Area Underestimate: 459.5

Hyper-Ellipse Surface Area Estimation error

When encountering very oblong, prolate, hyper-ellipsoids, Thompsen's method to estimate surface area can be up to $\frac{2^{1+\frac{1}{p}}}{\pi}$ times actual surface area. [87] However, smaller estimation error ,$\leq 1.061\%$, is more common. Exact ellipsoid surface area is analytically tractable, equation 3.42, but is an unnecessary effort. The surface area estimate is sufficient for MMRDP. In independent testing, Thompson's method yields very little error, Table 3.1.

$$\text{Ellipsoid Surface Area} = \int\int(1+(\frac{xr_3^{1/2}\sqrt{1+\frac{-x^2}{r_1}\frac{-y^2}{r_2}}}{r_1})^2+(\frac{yr_3^{1/2}\sqrt{1+\frac{-x^2}{r_1}\frac{-y^2}{r_2}}}{r_2})^2)^{\frac{1}{2}}dxdy \tag{3.42}$$

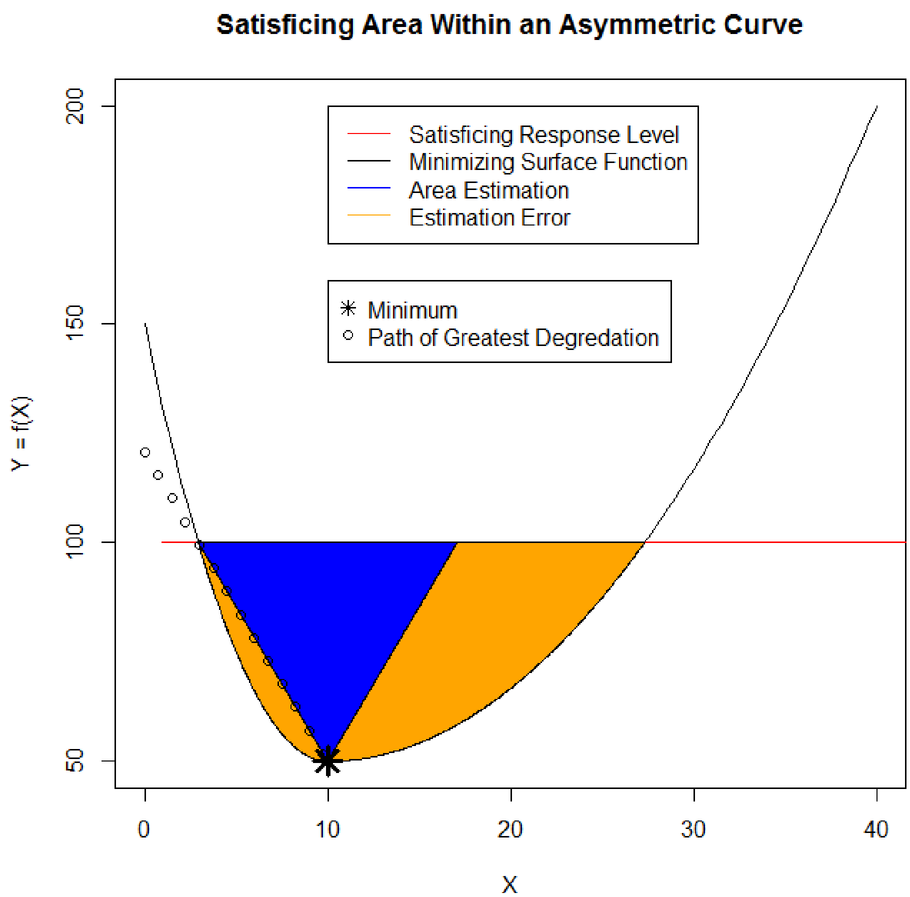


Fig. 3.18: The path of greatest degradation method ensures underestimation of surface area in the presence of asymmetric functions.

Tab. 3.1: Thomsen's surface area estimation method yields less than .25% error in three dimensions.

Radii	**known SA**	**KT Estimate**	**% diff**
1,1,1	12.56637061	12.56637	4.89E-06
10,5,1	333.946	334.3984	-0.135
1,8,7	369.097	369.9699	-0.236
20,10, 5	1586.92	1584.783	0.134
.01,.05,.4	0.132122	0.1319855	0.103

Application of MMRDP Robust RSM

MMRDP procedures were applied to four known functions, Appendix A. Each function is multidimensional, with known optimal values and variable constraints meant to simulate response, control, and exogenous variables, Table 3.2. Both basic and robust RSM were applied to data from the test functions exhibit stochastic behavior, both gaussian and nongaussian. Gaussian behavior is simulated using standard normal parameters and nongaussian behavior is simulated using exponential($\lambda = 1$), uniform (-2,2), and convolutions of pairs and tuples of the random variable types, Appendix B. The random variable generation meets the desired performance requirements.

Figure3.19 summarizes results from simulations of Dejong's first function in n dimensions and the robust methods. It shows the basic RSM and robust best local optimal value outcome given multiple start points and the varied combinations of control and exogenous factors. In general, bias increases with dimension and there is a slight estimation advantage in the presence of gaussian error. This is due to the use of OLS regression in the presence of gaussian error and LTS for nongaussian error. OLS is specifically optimized for gaussian error while LTS can be applied broadly to all complimentary stochastic behavior. Figure 3.19 shows the bias and variance tradeoff between robust solutions and more optimal ones when using the DeJong function data. Results from other functions, Rosenbrock's valley in n dimensions, Rastrigin's function in n dimensions, Schwefel's function in n dimensions, can be found in Appendix C.

Tab. 3.2: The four test functions have known minimum values across 1 to N dimensions.

Function Number	**Name**	**Dimensions**	**Optimal Value**
1	DeJong's First	N	0
2	Rosenbrock's Valley	N	0
3	Rastringin's	N	0
4	Schwefel's	N	-418.98

In general, MMRDP robust methods exhibit similar growth in MSE with dimensionality for all functions. MMRDP has greater success in terms of bias in the face of gaussian error than other types of error. This is true for all four functions, Figure 3.20.

3.6 Conclusion

MMRDP should not be conducted without contemplation. The overhead can be significant in terms of time and effort required to generate the ABM. The high fidelity simulation required by stakeholders, the more time required for

Func #	Exo Vars	Control Vars	Total Vars	Error Type	Known Opt Point	RSM Var	RSM Bias	RSM MSE	Robust Var	Robust Bias	Robust MSE
1	1	1	2	Gaussian	0.00	1.64	12.72	163.53	1.23	18.97	361.05
1	1	3	4	Gaussian	0.00	1.19	9.55	92.41	1.02	15.97	255.92
1	1	5	6	Gaussian	0.00	3.40	35.86	1289.34	2.42	48.03	2309.28
1	1	7	8	Gaussian	0.00	2.55	36.21	1313.89	1.60	46.51	2164.56
1	1	9	10	Gaussian	0.00	2.55	53.06	2817.91	1.85	71.91	5172.90
1	1	11	12	Gaussian	0.00	3.54	89.01	7926.32	2.30	112.18	12586.20
1	1	1	2	Non-Gaussian	1.00	8.42	12.89	174.51	7.86	20.55	430.33
1	1	1	2	Non-Gaussian	0.00	1.46	7.71	60.88	1.07	8.12	67.07
1	1	3	4	Non-Gaussian	0.00	1.58	24.29	591.73	1.32	32.16	1035.30
1	1	5	6	Non-Gaussian	0.00	1.42	35.90	1289.99	0.83	66.77	4458.46
1	1	7	8	Non-Gaussian	0.00	1.48	70.92	5031.74	1.45	121.93	14867.61
1	1	9	10	Non-Gaussian	0.00	1.54	79.52	6325.50	1.15	134.63	18127.54
1	1	11	12	Non-Gaussian	0.00	1.60	87.56	7668.94	1.46	93.17	8682.89
1	3	1	4	Non-Gaussian	1.00	8.87	5.62	40.46	4.49	5.91	39.43
1	3	3	6	Non-Gaussian	1.00	8.15	2.84	16.22	7.20	5.76	40.39
1	3	5	8	Non-Gaussian	1.00	3.46	3.96	19.16	2.13	6.17	40.23
1	3	7	10	Non-Gaussian	1.00	16.86	58.65	3456.68	15.93	74.55	5574.10
1	3	9	12	Non-Gaussian	1.00	8.72	83.51	6815.80	5.53	97.85	9581.07
1	3	11	14	Non-Gaussian	1.00	42.13	94.47	8965.76	27.65	154.55	23914.13

Fig. 3.19: Both RSM and the robust searches are subject to bias. However, the robust outcomes yield lower variance as a trade for increased bias. The marked difference in mean squared error between the two methods is due to the exponential relationship between bias and mean squared error.

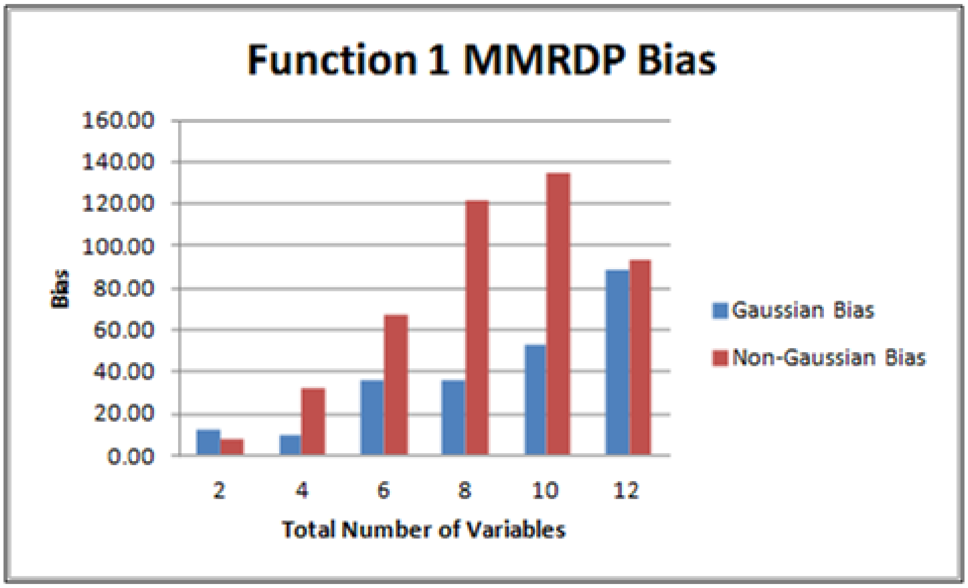


Fig. 3.20: While bias increases with dimension, it is less prevalent when estimating functions with gaussian error than those without.

construction and validation. MMRDP can suffer from lack of convergence and it hinges on the veracity of the ABM. The simulation is the link between stakeholder analysis and the statistical methods but it it cannot be validated. Complexity, human behavior, and resource allocation conspire to provide the modeling challenge. Currently, historical reference and the delphi method are most promising validation methods. However, the inherent subjectivity of both will not provide the practitioner guaranteed valid models.

Despite the risk, MMRDP can address the problem of locally optimizing a response using a computer simulation model by outlining an ordered, logical, and validated exploration of the space. [31] Though it lacks high precision, includes the assumption of symmetrical topography, and is subject to the curse of dimensionality, MMRDP is viable. It distills complexity into a managable representation and gives leaders a previously unused method evaluate courses of action. Within RSM, instills a lack of sensitivity within solutions to exogenous variable behavior with both gaussian and nongaussian stochastic data. When discriminating between locally optimal RSM outcomes, it offers conservative estimates of satisficing behavior area in order to eliminate the ambiguity of confidence interval. This method also eliminates numerous additional simulation runs required to use paired - t or Tukey testing.

4. APPLICATION OF THE ROBUST META-MODEL: JALALABAD CITY, NANGARHAR PROVINCE, AFGHANISTAN

"The first condition of understanding a foreign country is to smell it."

~ Rudyard Kipling, British Author, Poet, and Nobel Laureate

4.1 Introduction

Consider the application of MMRDP with respect to the current situation for the United States and NATO in Afghanistan. The United States Army Corps of Engineers (USACE) funded an effort to model the impact of infrastructure portfolios on public opinion in support of COIN strategy. Key stakeholders include the International Security Assistance Force (ISAF), the Government of the Islamic Republic of Afghanistan (GIRoA), local governments, and civilians. Analysis of the local environment yields key insights about the relationships that meet the stakeholders' objective, improved population opinion of the government in Afghanistan. Initially, the analysis identifies infrastructure capability gaps and associated quantifiable metrics. With the help of an ABM and RSM, locally optimal and robust infrastructure solutions are recommended to improve public opinion of government.

4.2 Background

Large-scale civil reconstruction accompanies catastrophe. Natural and inadvertent disasters occur but here, civil recovery as a function of violence and conflict is of interest. Warfare involves a premeditated, focused, and sustained effort to disable critical infrastructure. Afghanistan, a prime example, requires a large investment in reconstruction after the 2001 invasion by US and Coalition Forces (CF). The conflict region offers no feasible benchmarks for reconstruction. Applying western and NATO values and ideals to the problem is a naive strategy. [8]

The NATO mission in Afghanistan faces a tenacious insurgency. [2] To assist in achieving NATO's goal of stability, COIN strategy includes fostering local support for the government by meeting basic population needs. [91] With this in mind, regional and local system of systems design should bridge capability gaps and enhance rule of law through selected project portfolios. However, there are few methodologies that capture higher order effects on populations and include interactions within infrastructure system portfolios. [92] This motivates the use of MMRDP to improve public support, Figure 4.1.

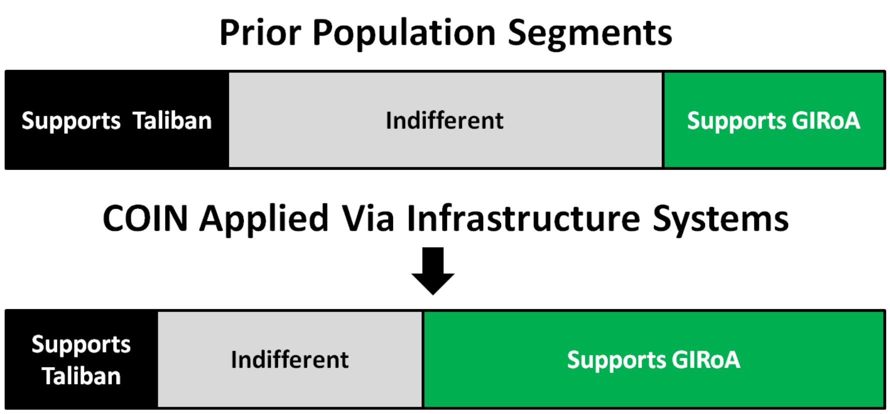


Fig. 4.1: MMRDP selects local infrastructure systems to augment counter insurgency strategy and improve public opinion.

The United States military uses reconstruction as a foundational element of stability and a critical task within COIN. Though expensive, it is beneficial to all stakeholders, tangible in nature with a direct and positive impact on the community, and fosters long term stability. Countries with loose national cohesion, porous borders, limited governmental influence, minimal revenue streams, and rudimentary infrastructure development present military commanders with a problem, mission creep. Is the charge to bring peace by merely brokering an end to large-scale conflict, or grant a more robust peace opportunity, built upon stability and the people's access to critical resources and services? The US stability strategy addresses the debate by outlining the constituents of stability. [93]

- A safe and secure environment
- Established rule of law
- Social well-being
- Stable governance
- A sustainable economy

Worldwide, DOD is collecting lessons learned with an increased emphasis on COIN doctrine. [3] The military is best suited to address and insurgency because it has the "integral mobility and protection capability" necessary to operate and win the true prize of an insurgency, the population. [94] Infrastructure reconstruction is a lynch pin for a successful COIN strategy. It is an accessible tool for large military organizations, governments, and coalitions. It is a necessary and moral obligation, and it is a means to undermine insurgency by bridging capability gaps and demonstrating to the population the sanctioned authority's reach. [3]

The US Army and US Marine Corps COIN strategy is comprised of four steps. [3]

Shape Set the conditions to plan and execute operations.

Clear Establish security in the locale for all participants.

Hold Affirm the government's authority and control.

Build Gain support from the population.

The steps foster incremental improvement for the population in a region and beget less sanctuary for insurgents. This, in turn, invites and sustains stability. It is not purely a military effort. Sustainable gains against insurgency are multi-faceted. They reflect a grander cooperative effort between the population, governments, and agencies. Efforts by NATO in Afghanistan and the Coalitions Forces in Iraq have had varying levels of success due to naivety or blatant disregard for holistic solutions by civil and military leaders. Military units that fail to plan for life after large scale combat is over will have a large problem with security. That will bleed into reconstruction and stability failures as well.

Security as a Basis for Stability

Infrastructure is not always tangible. In violent situations, infrastructure that foments stability must include security. [91] More traditional systems are then viable projects. Without security, access to other infrastructure systems is limited, negating their existence. [95] Maslow's Hierarchy of Need offers a method to prioritize societal infrastructure needs. [96] The hierarchy is based on that which motivates human behavior, from the base physiological needs to less tangible goals such as esteem, Figure 4.2.

Mapping Maslow's hierarchy to regional needs can offer decision makers a plan for resource allocation. [97] In Afghanistan, security is the base needs that all gains build from. [91]

- Physiological needs – required to survive: Water, food, shelter
- Safety needs – ensure order and predictability: Public health, security, justice, information
- Love and belonging – friends and family: Awaken culture and community institutions
- Esteem – the need for respect: Participation in governing, inclusive decision making
- Self-actualization – need to realize full potential: Regional center for other countries to emulate

4.3 Problem Statement

Stakeholders wish to maximize public opinion of the government in the city of Jalalabad, Nangarhar Province, Islamic Republic of Afghanistan based on

Fig. 4.2: Maslow's Hierarchy of Need demonstrates security as a base human need.

infrastructure policy decisions and a fixed budget, Figure 4.3. Specifically, within the class of problems that encompass optimization of a univariate response in a complex, data-poor environment, MMRDP is used to estimate a mathematical relationship between relevant factors and the response. It explores the relationship within factor bounds and budget constraint in order to discover locally optimal response levels insensitive to exogenous factor uncertainty. In the event multiple optimal portfolio options exist without statistically significant differences, the MMRDP optima filter uses an alternate method to discriminate. The relationships and the recommended significant factors levels constitute an improved robust solution to this problem.

Given a set of infrastructure projects, each with associated life cycle costs and attributes, selection criteria are necessary to discriminate between them, group them, and select them. Maximize public opinion of government, O, as a function of projects, and environmental factors for a population across a specific geographic region. Constraints include control factor minimum and maximum levels and a total budget, B.

Fig. 4.3: The impact of infrastructure on Jalalabad population opinion is the focus of the study

$$\text{Maximize } O, O \sim f(X, Z) \tag{4.1}$$

$$\text{subject to} \tag{4.2}$$

$$x_i \leq x_{U(i)} \tag{4.3}$$

$$x_i \geq x_{L(i)} \tag{4.4}$$

$$\bigvee i \in I, I = 1..4 \tag{4.5}$$

$$Factor_j \sim G_j \tag{4.6}$$

$$\bigvee j \in J, J = 1..3 \tag{4.7}$$

$$\sum \Psi X \leq B \tag{4.8}$$

X_L, X_U – control variable upper and lower bounds.

G – exogenous variable general distribution

B – budget constraint

4.4 Stakeholder Analysis

Input for stakeholder analysis consists of feedback from GIRoA, NATO, USACE, the UN, civilians from the region, and the US military. It considers the myriad needs of the stakeholders, emergent properties of the environment, interactions, and planning documents such as the Afghan National Development Strategy (ANDS). ANDS outlines the GIRoA plan to establish governance, rule of law, human rights, and economic and social development. [98]

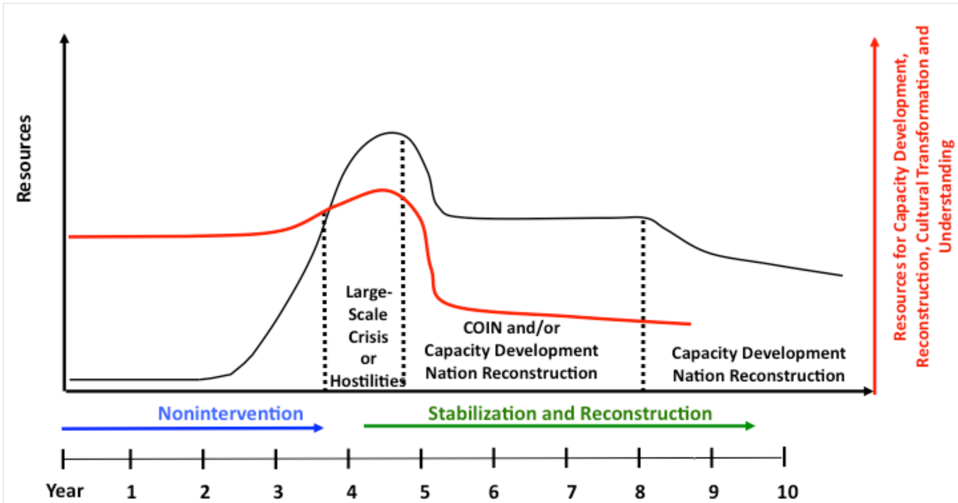


Fig. 4.4: The Center for Nation Reconstruction and Capacity Development post combat development timeline shows that resources available for reconstruction after hostilities are finite and they decrease.

It also serves as the azimuth for infrastructure development in Afghanistan. Given available resources, a general set of projects can be selected as candidates for a region.

Insurgent activity, ethnicity, gender, religion, and regional norms make it extremely difficult to measure the impact of individual or systems of infrastructure projects. Include graft, politics, geography, and the challenge grows exponentially. It is therefore very difficult for GIRoA to understand how to gain traction with the population through infrastructure investment.

In Afghanistan, failure to improve the lives of the population lends credence to the Taliban movement. The civilian population support those that give them the basics they so desperately need. [99] This leads to a tacit, but real, legitimacy for Taliban rule, extends the insurgency indefinitely, and renders goals of security, governance, and development unattainable.

Operations in Afghanistan have cost ISAF over 3200 lives since 2001. [100] The civilian death toll is at least 5 times greater. [9] The 43 nations of ISAF have a strong moral and strategic imperative to stabilize this "graveyard of empires" because the government cannot generate and maintain peace on its own.

Protracted efforts by the NATO and GIRoA have achieved only minor changes in the status quo. As a result, Afghan insurgents continue to operate with near impunity and civilian patience with foreign troops wears thin. [101] Coalition partners in Afghanistan are reconsidering their support given great sacrifices and cost. Beyond forces, they provide billions of dollars a year in foreign aid. Though expected after large scale hostilities, reconstruction resources quickly peak, then wane. Figure 4.4, from the Center for Nation Reconstruction and Capacity Development (C/NRCD), shows the resource decline after hostilities dissipate. In Afghanistan, resources and resolve are dwindling. This limits needed infrastructure projects that might serve as tools to turn the tide toward peace.

Afghanistan's strategic importance and diversity has led to countless wars within and across its borders. It constitutes the route from the Middle East to Asia, comprising much of the famed Silk Road. [7] It is a tapestry of distinct ethnicities and religions making the country more of a collection of tribes and races rather than a unified nation, complicating efforts for country wide stability. [102]

Afghanistan's history of warfare, the ISAF invasion, and subsequent security operations in this fractious nation make the need for infrastructure capacity and capabilities great. Few roads are paved. Electricity, sewer, and water are luxuries even within the capital, Kabul. A steady exodus has all but drained any intellectual capital along with the stewardship of institutions that most other countries take for granted. [102] Only 15% of Afghans are literate and they have the second lowest life expectancy of any nation in the world, 45 years. [7] Among many others, USACE, and the C/NRCD are working to design, install, and improve infrastructure across the nature in order to underwrite a more promising future.

Precedents illustrate the benefits of reconstruction. The Marshall Plan, 1947–51, provided a means for the United States to help rebuild and redevelop Europe. It was lauded for the positive impact on individuals as well as on entire countries. The Allies worried that newly peaceful Western Europe would not recover unless substantial programs for support of the people and industry, such as training and construction were not enacted. [103] By 1952, the economies of each member nation of the Marshall Plan surpassed prewar levels by double digit margins. [104] Governments exploited stability to maintain rule of law and foster greater prosperity.

Things are slowly improving in Afghanistan. However, resource allocation failures still occur due to a lack of understanding. Unforeseen second and third order effects can make bad situations worse despite good intentions. ISAF and USACE want to provide infrastructure that the Afghans need. As

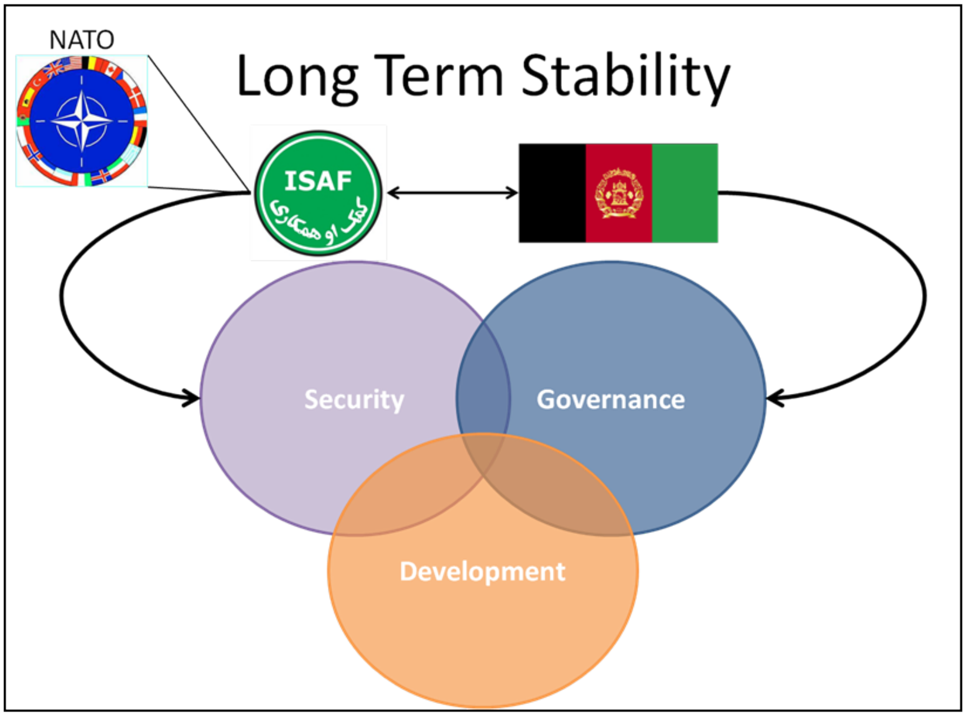


Fig. 4.5: ISAF and GIRoA use common lines of operation, Security, Governance, and Development, to drive operation, strategy, and tactics.

an overall strategy for stability, ISAF and GIRoA jointly focus on improving three lines of operation (LOOs): Security, Governance, and Development, Figure 4.5. However, there are currently few mechanisms to indicate the effects of infrastructure on popular support for GIRoA or changes in the status of the LOOS.

Insurgent activity, ethnicity, gender, religion, and regional norms make it extremely difficult to measure the impact of individual or systems of infrastructure projects. Include graft, politics, geography, and the challenge grows exponentially. It is therefore very difficult for GIRoA to understand how to gain traction with the population through infrastructure investment.

Marshalling stakeholders and analyzing the local environment yield key insights about functions that meet the stakeholders' objectives. Further analysis generates capability gaps and associated quantifiable metrics. The gaps drive infrastructure solution recommendations while the metrics allow for a basis of comparison screening criteria for needs and objectives.

Figure 4.6 shows a functional diagram from the stakeholder analysis. Each LOO was decomposed by a group of experienced military officers with experience in Afghanistan into its critical functions. [91] Those functions were dissected as necessary until metrics were appropriate. Strong performance on the infrastructure metrics will support the LOOs and draw support away from the insurgents by improving public opinion of GIRoA. This vertical and horizontal nesting of effort and understanding reflects stakeholders' desires and allows for comparison of alternatives.

Public Opinion of Government

The meta-model, through an ABM, generates a measure of public opinion as an outcome of an infrastructure portfolio selection. Generating public opinion adds a greater level of complexity to the process because it is key to understanding the the relationship between insurgent support and infrastructure selection. Factors that impact public opinion of government in Muslim countries include domestic economic issues, education, and health services. [105] Weak government institutions frustrate citizens while poorly

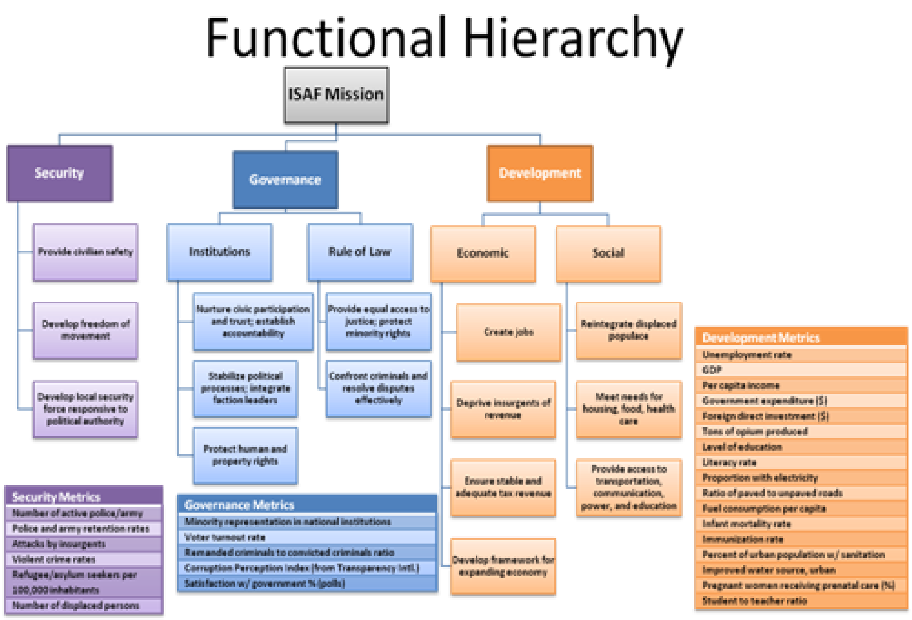


Fig. 4.6: The functional decomposition of the Lines of Operation links each line to metrics. It identifies capacity gaps and their associated metrics.

representing all races, tribes, religions, and classes inviting greater unrest and discord. This can lead to the, typically large and poor, segments of the population seeking alliances with power in order to increase their chances of achieving self determination. Over time, the Taliban are becoming less popular due to their lack of ability improve economic conditions and public service. But support for the Taliban will propagate if the authority cannot reliably provide the same. [105] In the eyes of the people, there is no distinction between government institutions and government actors meaning that all get the blame for the poor performance of anyone or anything associated with it. [106]

Though westerners see the the long ISAF commitment in Afghanistan as noble, locally, it is seen negatively. The impression is that ISAF has changed from liberator to occupier. [105, 107] Until ISAF ends it mission, GIRoA and Afghan National Security Force (ANSF) led operations can alleviate tensions through greater inherent cultural savvy. They limit foreign involvement and validate domestic capability as well. In Afghanistan, a high tolerance for insecurity given the past 30 years has provided patience for operations that negatively impact the public in the short term in order to achieve long term gains. [108] However, this bit of goodwill is nearly gone. Polling indicates that populations desire stability as a means to fight extremism because a stable environment is less likely to exploit or ostracize a minority. This leads to greater social cohesion, fewer disruptions, and a higher opinion of police. [109] In order to improve opinion, a multifaceted approach that looks beyond on economics and includes security is necessary to improve public opinion of government. This philosophy was applied within the ABM to generate the opinion measure.

The public opinion measure in the meta-model is the result of portfolio settings, agent behaviors, characteristics, attitudes, and interactions. Some outcomes are intuitive while others are not. For instance, if deaths increase due to substandard health care, public opinion suffers. However, a population encountering strict security measures or a large instance of ISAF troops, may not increase their opinion despite a decrease in insurgent attacks. The role of cultural awareness is another complicated feature of the model. While it would seems that the population would be more tolerant of ANSF patrols compared to ISAF patrols, the great variability in the quality of ANSF agents can provide surprising opinion results.

The stakeholder analysis also includes an infrastructure impact assessment. Key infrastructure needs for Jalalabad include transportation, water, power, sanitation, agriculture, and public health. An impact analysis for each addresses user benefits, socio-economic effects, and the impact on the civilian population as well as the insurgency, Appendix D.

The stakeholder analysis indicates that for Jalalabad, security, public health, and unfettered access to water are most important to the population. Consideration for ISAF presence, ethnicity. and socio-economic factors must also be made within the meta-model to determine opinion.

4.5 Jalalabad Agent-Based Model

The meta-model process for infrastructure portfolio selection is a systems approach incorporating stakeholder analysis as an initial step. In addition to ANDS and the ISAF and GIRoA LOOs, the model uses measures of performance (MOP) and measures of effect (MOE) for each level of governance and military responsibility. The framework the model is based on is periodically reviewed and revamped to adjust to new players and changes in the environment. [6] This offers not only current but future stakeholder analysis needs making the use of computer simulation desirable.

Infrastructure systems have long planning horizons and require tremendous investment. An agent-based model is desirable for data collection. It is the best opportunity to experiment and observe individual autonomous behavior, pseudo-emergent outcomes, and model complexity within a computer-based algorithm. Participatory modelling is another way to increase the fidelity of a social model and inculcate complexity. However, Afghanistan's lack of widespread information technology coupled with the poorly educated population negate this option. Only small fractions of the population would be able to participate in such a model introducing bias.

The stakeholder analysis informs a representative ABM of a local area. Data from the simulation is used within response surface methodology to generates a robust, locally optimal portfolio, Figure 4.7. Drawing from the stakeholder analysis, MMRDP explores portfolios of water wells and health

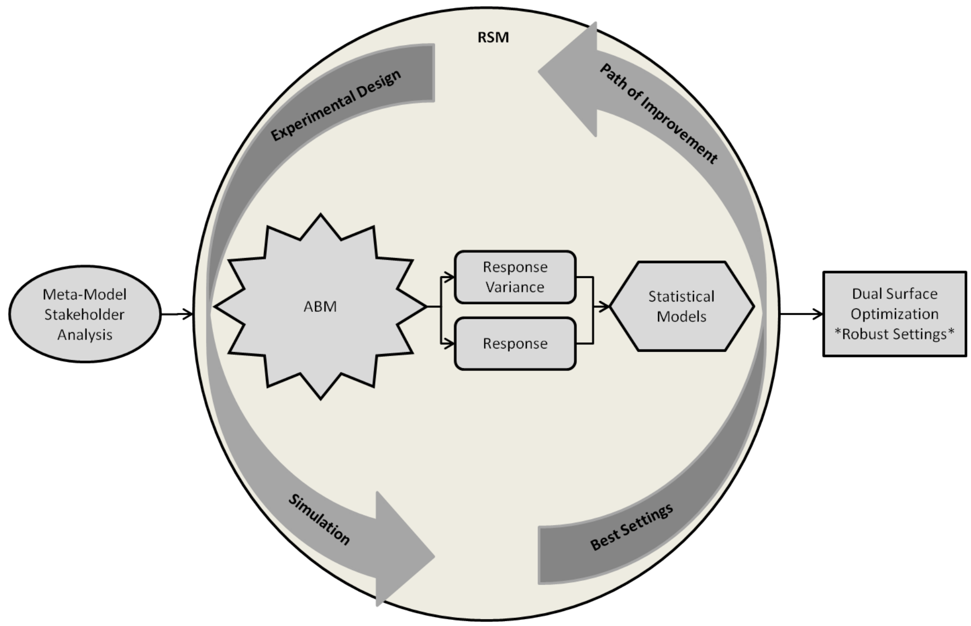


Fig. 4.7: The meta-model process incorporates stakeholder analysis, ABM simulation, and dual surface RSM for a robust solution.

care while instituting security policies and ISAF:ANSF patrol troop ratios in Jalalabad.

The ABM represents the Jalalabad city center, in Nangarhar Province, Afghanistan, total population of approximately $200,000$ people, Figure 4.3. It considers the impact of public health options (hospital capacity), access to municipal water supplies (public access wells), security levels (on the ground presence and capability), and ANSF and ISAF patrol mixes on the population. The ABM does not aspire to discern small differences in geographic arrangement of infrastructure courses of action but contains geographic features (terrain and distance) to add fidelity. The model demonstrates the impact of gross infrastructure decisions through the lens of the population opinion. Insurgents follow directives such as *Harass the authority and attrit its agents* and *Maintain unrest in the civilian population*. It is in their interest to prolong conflict and undermine security, governance, and development for as long as possible. The ubiquity of the struggle can demonstrate any organization's reach and highlight government failings. [3] The geospatial and demographic data in the ABM comprise exogenous factors that are beyond the control but of interest to decision makers.

The ABM used in the Jalalabad study was developed in Repast Simphony version 1.2 and is based on the RepastCity project. [110] The RepastCity project serves as a starting point for the Jalalabad ABM because it features agents moving around a city between their homes and other buildings. Agents in the simulation represent individuals whose attributes and traits are sampled from a demographic model collected from a variety of sources summarized in Table 4.1. The agent-based model simulates behavior using 200 agents and their public opinion over a 60 day period for a given infrastructure portfolio. 200 agents are the limit of the simulation therefore the representative geographic area of Jalalabad is sized to an appropriate proportional level, Figure 4.8. Each iteration requires 5-10 minutes on a quad core system. As an extension of stakeholder analysis, the model uses empirical data in the simulation, Table 4.1.

The model represents a small portion of the city with the maximum number of agents. While the agents are spread equally between them, each neighborhood has its own characteristics in terms of exogenous variables.

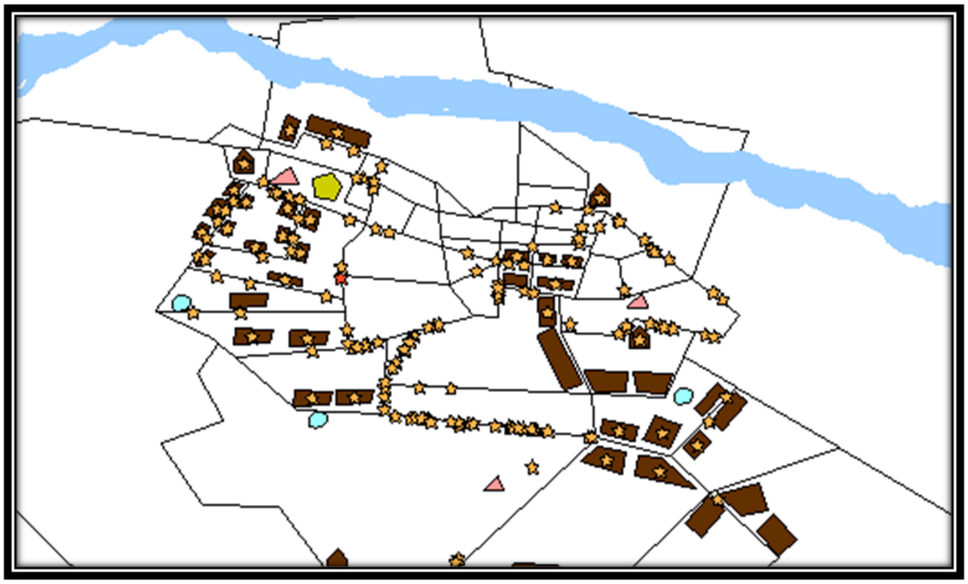


Fig. 4.8: Agent-Based Model of Jalalabad (Pink Triangle - Hospital, Blue Circle - Well, Brown Square - Home, Green Pentagon - Police Station, Orange Star - Agent)

Tab. 4.1: The Jalalabad ABM uses local and national empirical data.

Afghanistan Data	**Value**	**Source**
Life Expectancy	48.3	World Bank, 2010
Poverty Rate	36%	World Bank, 2008
Rural Access to Improved Water	42%	World Bank, 2010
Jalalabad Data	**Value**	**Source**
Annual Instances of Outpatient Sickness	4,792	ANHSR, 2004
Population	205,000	UNICEF, 2009
Pashtun	85%	CIA Factbook, 2011

The geospatial and demographic data input capability of Repast provides the flexibility to increase the model's fidelity and generalize the approach for other applications by adding necessary layers or changing the background environment. Jalalabad road data from the National Geospatial Intelligence Agency (NGIA) and building data provide routes of travel within the model and the initial regional infrastructure levels within the city at the outset.

Within the model, agents interact with each other and receive benefits from local infrastructure projects. Due to ethnic divides and an ingrained culture of corruption, not every agent receives the same level of service from the infrastructure projects. Based on their wealth and ethnicity, rich Pashtuns receive immediate health care and do not risk being turned away from a well or a hospital. Alternately, poor agents and minorities wait longer than usual for health care and may not receive care at all. This is the case in times where the level of security is low. Security forces are built into the simulation to help preserve equal access to infrastructure projects. Security forces are comprised of a mixture of ANSF and ISAF patrols.

The level of security also influences the probability of a mass-casualty event. A low frequency random distribution generates mass-casualty events where the parameter is tied to the level of ground security. With low security, there is an increased chance of an event compared to times when high security is in place. The large events simulate the effect of capacity stressors on the health system by natural disasters, epidemics, or improvised explosive devices (IEDs). The ration of ISAF:ANSF troops impacts the population.

A high ratio of ISAF to ANSF patrols may be more effective in stopping insurgent gatherings and preventing mass casualty events, but the ISAF patrols are more costly than their ANSF counterparts. ISAF troops run more efficient military operations but have won little trust. ANSF troops can have high cultural sensitivity yielding better intelligence and local knowledge but at times they fall prey to ethnic prejudices, regional differences, poor preparation, and graft.

Within the model, agent health is modeled as a health score between 1 and 4. All agents are begin at level 1, good health. Based on proximity to a water source, agents have a probability of becoming sick. [111] When they become sick, health scores change to 2 and agents move along the roads between their homes and the nearest hospital. If not treated quickly enough, an agent's health condition may deteriorate to a score of 3. Once they arrive at a hospital, they enter a patient queue and eventually receive care when it is their turn. Health scores of 2 require some primary outpatient services. Health scores of 3 require longer (inpatient) stays at the hospital before they can be released. In some cases, agents with inpatient sicknesses cannot be treated fast enough and they die. Death is the final stage in an agent's existence and it corresponds to a health score of 4. Occasionally, mass casualty events occur. These events represent IEDs or other insurgent strikes and subject civilians to harm. Such events can lead to severe injury (health score 3) or death (health score 4).

The statistic of interest for each simulation is *Public Opinion*. It is dependent on the type of infrastructure portfolios in the city as well exogenous factors. The number of population deaths, injuries, illness, and insurgent attacks influence *Public Opinion*. Each replication has terminal level of the statistic. Those portfolios with the greatest level of *Public Opinion* are most desirable.

Verification of the agent-based model is not a simple as discrete event simulation. However it is possible. For instance, long runs of the replications yield estimations of death and illness rates. In a portfolio with no additional infrastructure, these rates met expectations from embedded rates. Validation is more difficult, if not impossible. In the presence of a complex, adaptive system like Jalalabad, some data exists but validation is best sought by polling the expertise within the stakeholders.

4.6 Evaluation of Infrastructure Portfolios

The Jalalabad ABM has four control factors and three exogenous factors, each with various levels, Tables 4.2 and 4.3. The control factor levels also have an associated costs, Table 4.4. [95] The Jalalabad simulation has a very small surface space yielding only 48 design points, Table 4.5.

While MMRDP was run without a priori knowledge of a global optimum, the small response surface size makes full control factor enumeration possible and informs measurement of bias induced by MMRDP. The estimate of the optimal control settings is known, Table 4.6.

Tab. 4.2: The Jalalabad ABM has integer control factors.

	Wells	**Hospitals**	**Security**	**ANSF**
Definition	Wells to Construct	Hospitals to Construct	Security Presence	Patrol Units
Unit	Wells	Hospitals	Level	Level
Levels	4	3	2	2
Min	0	1	Low	ANSF Heavy
Max	3	3	High	ISAF Heavy

Tab. 4.3: The Jalalabad ABM exogenous factors are continuous.

	Wealth	**Ethnicity**	**Deaths**
Definition	Population Fraction Above the Poverty Line	Pashtun Fraction of Population	Deaths Per 60 Days
Unit	%	%	Agents
Mean	29.53	85.01	13.75
Variance	148.48	10.12	17.47

Tab. 4.4: Infrastructure systems each have an associated, scalable cost.

Project Cost	
System	**Cost (USD)**
Well Network Construction (135 wells)	$150,000
Status Quo Well Coverage	$0
Low Well Coverage (1 network)	$150,000
Mid Well Coverage (2 networks)	$300,000
High Well Coverage (3 networks)	$450,000
Hospital Construction	$25,000,000
Low Hosp. Coverage (1 hospital)	$25,000,000
Mid Hosp. Coverage (2 hospitals)	$50,000,000
High Hosp. Coverage (3 hospitals)	$75,000,000
Train and equip Afghan soldier	$25,000
Pay for Afghan soldier per month	$240
Total yearly cost of Afghan soldier	$25,240
Train and equip ISAF soldier	$50,000
Pay for ISAF soldier per month	$1,500
Total yearly cost of ISAF soldier	$51,500
Low Security ISAF-heavy (40 soldiers)	$898,700
High Security ISAF-heavy (160 soldiers)	$2,696,100
Low Security Afghan-heavy (40 soldiers)	$636,100
High Security Afghan-heavy (160 soldiers)	$1,908,300

Tab. 4.5: 48 Courses of action are available for Jalalabad between three types of infrastructure.

Project Portfolios					
Portfolio	**Wells**	**Hospitals**	**Security**	**Security Lead**	**Cost (USD)**
1	Status Quo	1	Platoon	ISAF	$ 26,797,400.00
2	Status Quo	1	Platoon	Afghan	$ 26,272,200.00
3	Status Quo	1	Company	ISAF	$ 32,189,600.00
4	Status Quo	1	Company	Afghan	$ 30,088,800.00
5	Status Quo	2	Platoon	ISAF	$ 51,797,400.00
6	Status Quo	2	Platoon	Afghan	$ 51,272,200.00
7	Status Quo	2	Company	ISAF	$ 57,189,600.00
8	Status Quo	2	Company	Afghan	$ 55,088,800.00
9	Status Quo	3	Platoon	ISAF	$ 76,797,400.00
10	Status Quo	3	Platoon	Afghan	$ 76,272,200.00
11	Status Quo	3	Company	ISAF	$ 82,189,600.00
12	Status Quo	3	Company	Afghan	$ 80,088,800.00
13	1 Network	1	Platoon	ISAF	$ 26,947,400.00
14	1 Network	1	Platoon	Afghan	$ 26,422,200.00
15	1 Network	1	Company	ISAF	$ 32,339,600.00
16	1 Network	1	Company	Afghan	$ 30,238,800.00
17	1 Network	2	Platoon	ISAF	$ 51,947,400.00
18	1 Network	2	Platoon	Afghan	$ 51,422,200.00
19	1 Network	2	Company	ISAF	$ 57,339,600.00
20	1 Network	2	Company	Afghan	$ 55,238,800.00
21	1 Network	3	Platoon	ISAF	$ 76,947,400.00
22	1 Network	3	Platoon	Afghan	$ 76,422,200.00
23	1 Network	3	Company	ISAF	$ 82,339,600.00
24	1 Network	3	Company	Afghan	$ 80,238,800.00
25	2 Networks	1	Platoon	ISAF	$ 27,097,400.00
26	2 Networks	1	Platoon	Afghan	$ 26,572,200.00
27	2 Networks	1	Company	ISAF	$ 32,489,600.00
28	2 Networks	1	Company	Afghan	$ 30,388,800.00
29	2 Networks	2	Platoon	ISAF	$ 52,097,400.00
30	2 Networks	2	Platoon	Afghan	$ 51,572,200.00
31	2 Networks	2	Company	ISAF	$ 57,489,600.00
32	2 Networks	2	Company	Afghan	$ 55,388,800.00
33	2 Networks	3	Platoon	ISAF	$ 77,097,400.00
34	2 Networks	3	Platoon	Afghan	$ 76,572,200.00
35	2 Networks	3	Company	ISAF	$ 82,489,600.00
36	2 Networks	3	Company	Afghan	$ 80,388,800.00
37	3 Networks	1	Platoon	ISAF	$ 27,247,400.00
38	3 Networks	1	Platoon	Afghan	$ 26,722,200.00
39	3 Networks	1	Company	ISAF	$ 32,639,600.00
40	3 Networks	1	Company	Afghan	$ 30,538,800.00
41	3 Networks	2	Platoon	ISAF	$ 52,247,400.00
42	3 Networks	2	Platoon	Afghan	$ 51,722,200.00
43	3 Networks	2	Company	ISAF	$ 57,639,600.00
44	3 Networks	2	Company	Afghan	$ 55,538,800.00
45	3 Networks	3	Platoon	ISAF	$ 77,247,400.00
46	3 Networks	3	Platoon	Afghan	$ 76,722,200.00
47	3 Networks	3	Company	ISAF	$ 82,639,600.00
48	3 Networks	3	Company	Afghan	$ 80,538,800.00

Tab. 4.6: The global optimal point is known given the small decision space of the Jalalabad study.

Wells	Hospitals	Security	ANSF	Public Opinion	Variance	Cost
3	2	1	1	633.62	100544.32	$ 55,538,800.00

RSM can begin at any point in the control space. The study used three initial points. Each represents a policy of monolithic infrastructure investment; maximum health care, maximum Afghan-led security presence, and maximum wells, Table 4.7. Each yielded a locally optimal and robust control setting recommendation via MMRDP.

Tab. 4.7: Each initial point represents a full investment in one infrastructure type.

Start Points	Wells	Hospitals	Security	ANSF	Public Opinion	Variance	Cost
All Security	0	1	1	1	387.90	119460.47	$ 30,088,800.00
All Wells	3	1	0	0	501.33	88695.94	$ 27,247,400.00
All Hospitals	0	3	0	0	365.57	34992.84	$ 76,797,400.00

Within the RSM, the Jalalabad agent-based model was run iteratively using 2^4 full-factorial designs for each experiment. The *resolution-IV* experimental design, derived from a foldover *resolution-III* design creates orthogonal, experimental sets free from colinearity. [31] Within the crossed array, each design point of interest had 13 replications due to the stochastic nature of the exogenous variables, *Wealth, Ethnicity, and Deaths.* Each experimental design run required 48 hours on a dedicated host server. The total computation time for all initial points was over 2 weeks. The R Statistical Package, version 2.15, was used estimate the response surface and paths of greatest improvement leading to further experiments and three locally optimal points. [62] The average response value for each design point was fit using using a GLM regression model due to the integral factors. The appropriate fit was determined by measuring deviance of first order, second order, and interaction models. Without exception, the second order model with interactions produced the best fit at locally optimal points.

After reaching classic RSM local optimality, the response variable, *Public Opinion*, was maximized using dual response surface optimization. To facilitate this step, satisficing behavior was determined to be response values greater than or equal to the median, ≥ 449.1. In the neighborhood surrounding each locally optimal point, the factors were fit to variance data. In this case, LTS regression models were used as the response was distributed according to a χ^2 distribution. [22] The best fit for the variance models was found via second order models with interactions. Both regression models were combined along with a budget constraint of $60,000,000 and the control factor limits within an optimization problem. The objective function maximized the *Opinion* response subject to reduced variance levels and the constraints. The resultant nonlinear, mixed integer problem was solved using GAMS software, see Appendix F. [77] In each of the three instances, it yielded a robust optimal control setting recommendation. Finally, the local optima filter was used to discriminate between the robust course of action recommendations to provide a ranking in terms of likelihood of achieving suitable behavior.

4.7 Results

MMRDP yielded three robust optimal recommendations each with bias but variance less that associated with the global optimal point, Table 4.8.

Tab. 4.8: MMRDP identifies three local optima that are less sensitive to exogenous factors than the global optimal point.

Local Optima	Wells	Hospitals	Security	ANSF	Public Opinion	Cost	Variance	Bias
1	1	2	0	1	538.65	$ 51,422,200.00	26486.23	94.97
2	2	1	0	1	476.67	$ 26,572,200.00	32349.54	156.95
3	1	2	0	0	432.43	$ 51,947,400.00	46423.84	201.19
Global	3	2	1	1	633.62	$ 55,538,800.00	100544.32	0.00

Each robust optima, 1, 2, and 3, sacrifice bias for lower variance compared to classic RSM results, Table 4.9. Of note, optima point 2 found via classic RSM was the global optimal point but robust RSM abandoned it for a point with less variance.

Tab. 4.9: Local optima 1, 2, and 3 arrive at lower variance points via dual surface optimization.

All Security (1)	Wells	Hospitals	Security	ANSF	Public Opinion	Bias	Variance	Cost
Start Point	0	1	1	1	387.90	245.72	119460.47	$ 30,088,800.00
Classic RSM	2	1	1	1	584.10	49.51	42092.70	$ 30,388,800.00
Dual Surface RSM	1	2	0	1	538.65	94.97	26486.23	$ 51,422,200.00
All Wells (2)	**Wells**	**Hospitals**	**Security**	**ANSF**	**Public Opinion**	**Bias**	**Variance**	**Cost**
Start Point	3	1	0	0	501.33	132.28	88695.94	$ 27,247,400.00
Classic RSM	3	2	1	1	633.62	0.00	100544.32	$ 55,538,800.00
Dual Surface RSM	2	1	0	1	476.67	156.95	32349.54	$ 26,572,200.00
All Hospitals (3)	**Wells**	**Hospitals**	**Security**	**ANSF**	**Public Opinion**	**Bias**	**Variance**	**Cost**
Start Point	0	3	0	0	365.57	268.05	34992.84	$ 76,797,400.00
Classic RSM	0	2	1	0	511.03	122.59	108261.20	$ 57,189,600.00
Dual Surface RSM	1	2	0	0	432.43	201.19	46423.84	$ 51,947,400.00

MMRDP demonstrates its value in its ability to sift through the courses of action using modeling and simulation. Intuitive solutions and extreme resource allocations are inferior to both the RSM and robust optimal solutions, Table 4.10.

The three optima are robust in terms of the effects from exogenous factors however, they are statistically the same due to their overlapping confidence intervals, Table 4.11 and Figure 4.9.

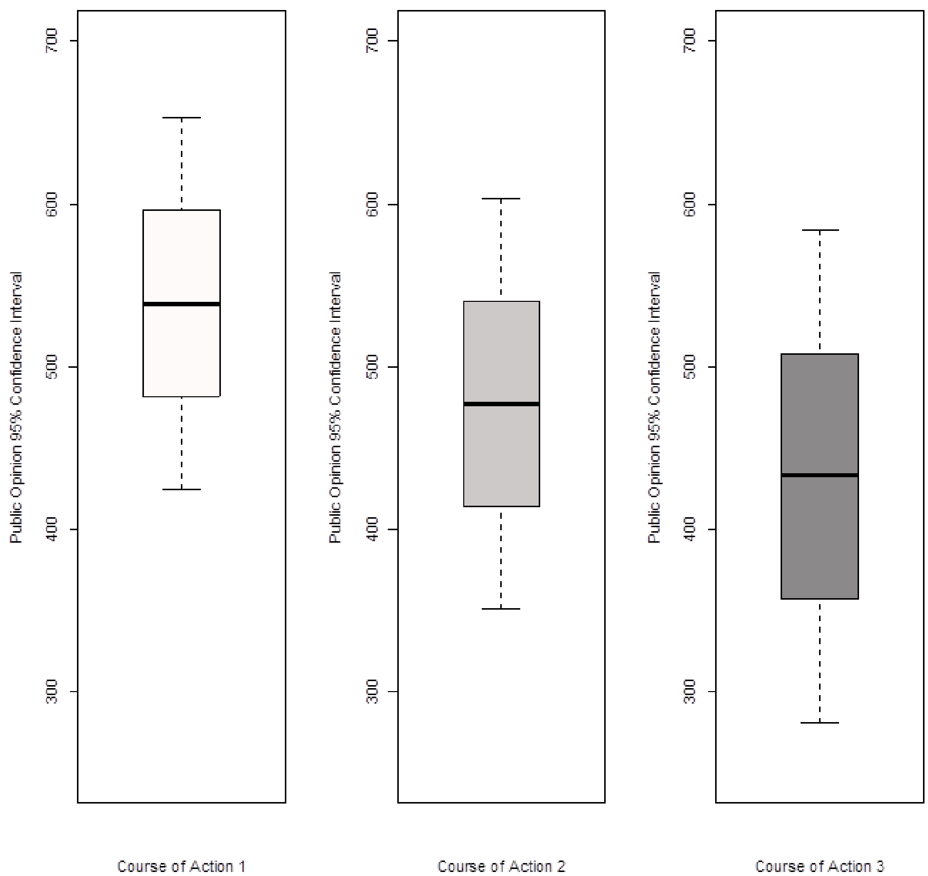


Fig. 4.9: The robust optimal points show no statistically significant differences.

Tab. 4.10: MMRDP provides solutions that beat intuitive solutions via both expected response outcome and response variance.

COA	Wells	Hospitals	Security	ANSF	Cost	Expected Public Opinion	Variance	Greater Than Median
All Hospitals	0	3	0	0	$76,797,400	365.57	34992.84	No
Minimum Cost	0	1	0	1	$26,272,200	339.84	56622.43	No
All Security	0	1	1	1	$30,088,800	387.90	119460.47	No
Robust Optimal	2	1	0	1	$26,572,200	476.67	32349.54	Yes
All Wells	3	1	0	0	$27,247,400	501.33	88698.94	Yes
Maximum Cost	3	3	1	0	$82,639,600	478.24	97003.38	Yes
RSM Optimal	3	2	1	1	$55,538,800	633.62	100544.32	Yes

Tab. 4.11: The 95% confidence intervals for the optima overlap and do not provide a clear recommendation.

Local Optima	Public Opinion	Variance	Half Width	Upper Limit	Lower Limit
1	538.65	26486.23	114.32	652.96	424.33
2	476.67	32349.54	126.34	603.00	350.33
3	432.43	201.19	9.96	442.39	422.47

Local Optima Filter Results

The local optima filter within MMRDP addresses this situation by determining the surface area of the ellipsoid that approximates satisficing response performance. Larger surface areas are directly correlated to higher probabilities of achieving desired response behavior. In this study, the ellipsoid is defined by the exogenous variable behavior, Wealth, Ethnicity, and Deaths, about the recommended control factor settings, *Public Opinion* $\geq$ 449.1. Through experimentation around each robust optima, the PGD to the satisficing was determined, providing radii for the corresponding ellipsoid and subsequently, its surface area, Table 4.12. Course of action 2, provides the greatest probability of success when compared to the others using ellipsoid surface area estimation as a proxy for probability comparison.

Tab. 4.12: Course of action 2 presents the greatest probability of success compared to the other optima using ellipsoid surface area estimation.

Satisficing Radii	**Wealth**	**Ethnicity**	**Deaths**	**Surface Area**
1	1.74	0.99	0.13	11.20
2	9.74	2.51	2.63	252.79
3	4.77	1.01	1.13	51.15

As additional inference for stakeholders, simulation data yielded more information about the response behavior in relation to the exogenous factors. By their nature, exogenous factors are uncontrollable. However, stakeholders can use simulation to determine their estimated effects on the control setting response. With optima 2, *Civilian Deaths* dominates other exogenous factors, Figure 4.10. While the factor of *Civilian Deaths* interacts with the other exogenous factors, the response is undesirable when *Civilian Deaths* exceed 23 in a 60 day period.

The result achieved by MMRDP has its greatest advantage in its analysis. The complexity of the COIN strategy for Jalalabad, the long time lines, the cost, the actors and the exogenous factors complicate decisions beyond the reach of tractional mission analysis methods. Security based decisions would be ignorant of other critical factors. Economic approaches would be none more helpful. Neither the most expensive nor the least expensive courses of action yielded better results in the expected level of public opinion or its variance. MMRDP offers insight about infrastructure' impact on COIN that leaders do not have.

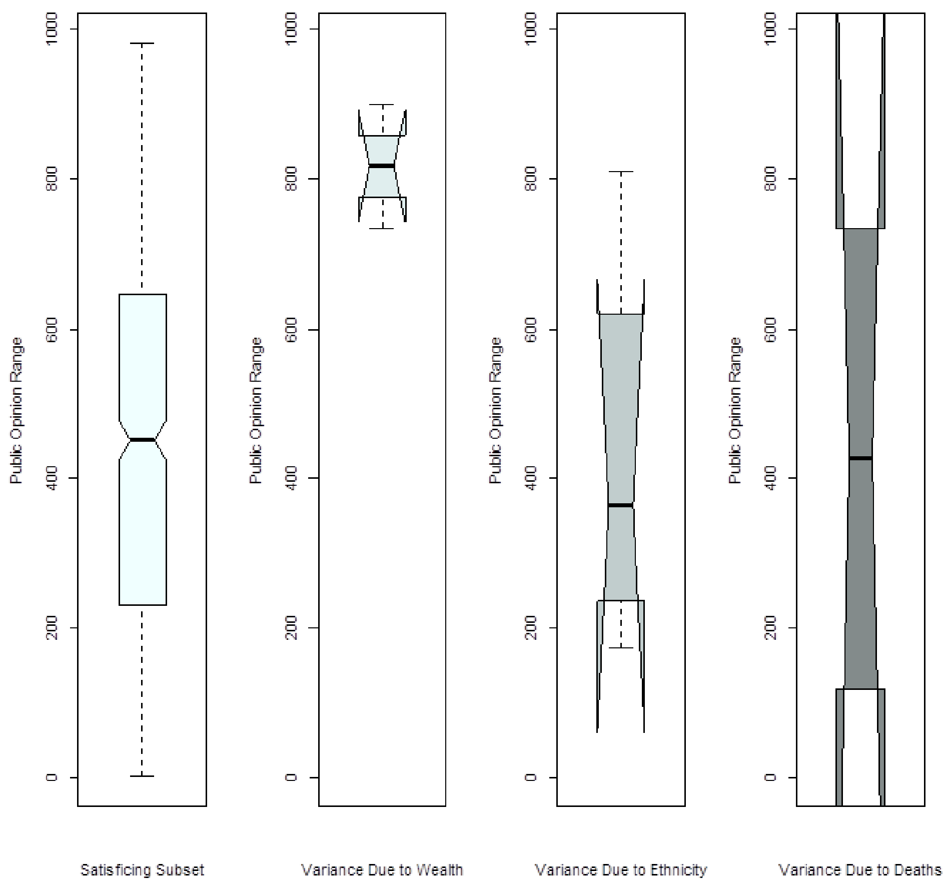


Fig. 4.10: Civilian deaths effect public opinion more than any other exogenous factor.

Pareto Efficiency

Pareto efficiency must be considered carefully within MMRDP. Robustness can invite deliberate suboptimal response performance to reduce variability. The naive (nonrobust) approach generates a curve that outperforms the robust approach, Figure 4.11. The robust method recommends optima 1.

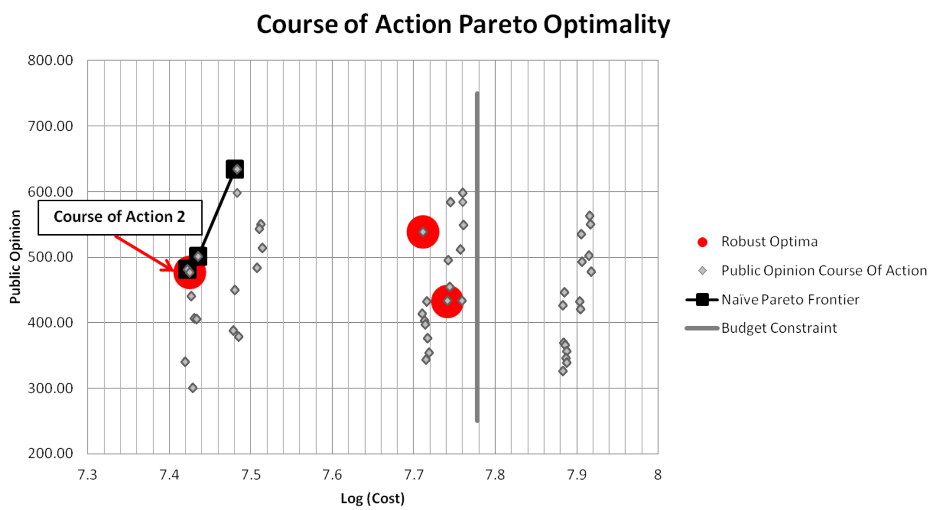


Fig. 4.11: The Pareto efficiency curve dominates Robust courses of action.

In this study, the pareto efficiency curve and the ellipsoid surface area estimation generate conflicting recommendations. The surface area calculation provides a clearer separation between the recommended courses of action. Optima 2 is the best recommendation, Table 4.13.

Tab. 4.13: Optima 2 is the MMRDP recommended setting.

Local Optima	Wells	Hospitals	Security	ANSF	Public Opinion	Cost	Variance	Bias
2	2	1	0	1	476.67	$ 26,572,200.00	32349.54	156.95

4.8 Conclusion

In the Jalalabad study, the recommendation highlights the benefits of cultural fluency as a combat multiplier. While it recommends construction of 2 well systems and one hospital, it does not recommend high levels of visible security forces. It indicates that the population is tired of the troop presence despite ISAF's perception that more security is better. It also indicates that the efficiency of highly trained foreign troops is "lost in translation" making them a reactive force compared to Afghan forces. Ideally, stronger investment in Afghan security will yield more predictable long term results at a lower cost compared to other more intuitive courses of action. Excess infrastructure capacity does not compare to just enough capacity with adequate security that best satisfy local stakeholder needs.

To improve the Jalalabad meta-model it may be worthwhile to move away from the Repast City ABM. Repast is not built for large-scale ABMs. It is slow, requiring approximately 20 minutes per replication with only 200 agents. It also very troublesome with upgrades, a common problem with freeware. Repast makes it difficult to build a freestanding, hardware and software independent simulation. Transfers between machines and software upgrades should not be points of friction as they appear to be.

Further analysis is required in reference to ANSF models. It is necessary to determine the impact of training on reported high corruption. The impact of nonlocal or minority Afghan security elements in Jalalabad is also undetermined. Within the ANSF, the contrast between Afghan National Army (ANA) performance and Afghan National Police (ANP) performance is of interest. In this way, ANSF can be deployed more precisely, cultural awareness can be exercised more effectively, and the imminent draw down of ISAF troops can be better mapped based on ANSF preparedness. Security, as a task, should underpin, not disrupt, reconstruction.

A noneconomic measure of stability provides stakeholders a picture of Afghanistan in the light used when the ISAF mission began in 2001. While strong economic performance is important to stability, is it not what compelled ISAF to enter Afghanistan. In that light, economic measures will not indicate the success or failure of the ISAF COIN mission. Public opinion of government, a harbinger of regional stability and insurgent support, is a viable means to guide infrastructure capacity development in a COIN environment.

Success of the the meta-model process hinges on the veracity of the response mean and variance, a very difficult task to validate. The process can also suffer if local data or expertise is unavailable. However, in the case of Jalalabad, ISAF has an eleven year history in the area. Many of the local issues are understood and well documented. ISAF employs human intelligence assets, human terrain mapping, and maintains a lessons learned repository available to units in the region. Dual surface RSM as a mechanism to achieve robust solutions, pairs well with the ABM and is an attractive feature for USACE, ISAF, GiROA, and other stakeholders with long time horizons and significant resource investments.

5. APPLICATION OF THE ROBUST META-MODEL: TIJUANA, ESTADA BAJA CALIFORNIA, MEXICO

"America's energy sector is just one part of an aging infrastructure badly in need of repair. Ask any CEO where they'd rather locate and hire: a country with deteriorating roads and bridges, or one with high-speed rail and internet; high-tech schools and self-healing power grids. The CEO of Siemens America – a company that brought hundreds of new jobs to North Carolina – has said that if we upgrade our infrastructure, they'll bring even more jobs."

∼ President Barack Obama

State of the Union Address, February 2013

"Smart businesses do not look at labor costs alone anymore. They do look at market access, transportation, telecommunications infrastructure and the education and skill level of the workforce, the development of capital and the regulatory market."

∼ Janet Napolitano, Secretary of Homeland Security, 2009-2013

MMRDP is applicable when seeking municipal stability in the absence of violence and organized conflict. Consider the the manufacturing (maquiladora) industry in Tijuana, Mexico. Tijuana's maquiladoras represent those in many of the Mexican cities along the United States' southern border. Mexican immigration to Tijuana City has produced steady population growth in response to decades of Mexican maquiladora proliferation without corresponding investment in shelter and basic services. Social and municipal stability can be put at risk due to capacity gaps and lack of reliable service access for laborers. In this study, an ABM informed by stakeholder analysis about Tijuana and the maquiladora industry. The stakeholder analysis identifies key factors and classifies them by type, control and exogenous.

MMRDP, through classic and robust RSM, identifies policy courses of action to minimize the maquiladora city immigrant laborers' long term rate of basic service exclusion. Low rates of exclusion can reduce the potential for unrest by improving quality of life. Informed infrastructure investment can provide policy recommendations for a sustainable maquiladora system underpin long term stability for municipalities and industrial partners.

5.1 Introduction

Operating under an arrangement between the United States and Mexican governments, raw materials and finished products arrive and depart, duty-free, from Mexican maquiladoras, domestic factories owned and operated by foreign companies. [112] For over forty years, labor needs climbed. [113, 114] In exchange for business friendly practices, these manufacturing centers garner a motivated labor force and a beneficial supply chain construct.

High employment in maquiladora cities is a double-edged sword. Industrial growth benefits cities financially but it is difficult for them to absorb the influx of new arrivals and provide them basic services in a timely manner. [115] As a result, ready access to potable water, electricity, and transportation is unlikely for newcomers. New immigrants cluster in outlying areas on the assumption that they will find suitable shelter there. Most settle in ramshackle communities on the city outskirts that often overlap with areas used to illegally dispose of waste. [116]

As a part of the permissive business environment, municipal authorities seldom review industrial practices or update exogenous policies. Industrial growth goes unchecked resulting in unregulated air pollution and inadequate water treatment. [117] The population's elderly and children are at the greatest risk, another potential source of volatility.

Beyond the pace of growth, the arid environment of the region limits the water supply, recasting growth from ally to antagonist. [118] City infrastructure investment increases via a woefully reactive policy stance that fails to address infrastructure capacity gaps and pace growth. For sustained investment and quality of life, alternate policies must be considered. The reliable Mexican labor force and economics of the maquiladora system justify their expansion but the substandard social aspects of the system can undermine it long term.

5.2 Background

During the early 1940's the United States developed the bracero program to address a labor shortage in the southwest stemming from World War II enlistments. [114] The bracero program eased travel for Mexican laborers travelling to and from US border state farms during the harvest season. In the off season, the labor force would take up station in northern Mexico waiting for employment. The pool of human capital was a opportunity for both the Mexican government and large businesses like General Motors and Ford. [119] The laborers' reputation for quality and dedication, as well as their high unemployment rate, set the stage for the maquiladora program to begin. US businesses built factories in Mexico. They transported raw materials in and exporting finished goods out duty-free; as long as they used Mexican labor. This policy increased inter-Mexican immigration to the border region due to the allure of work and in order to meet labor demand in cities such as Tijuana, Chipapa, Notana, Hoplili, and Nogales, Figure 5.1. [120, 121]

Maquiladora cities in the northern Mexican border zone have thrived for decades. Tijuana, a representative case study in the western state of Baja California, has quadrupled its population since 1980. [122] Located adjacent to San Diego, CA, the Pacific Ocean, and the busiest international border

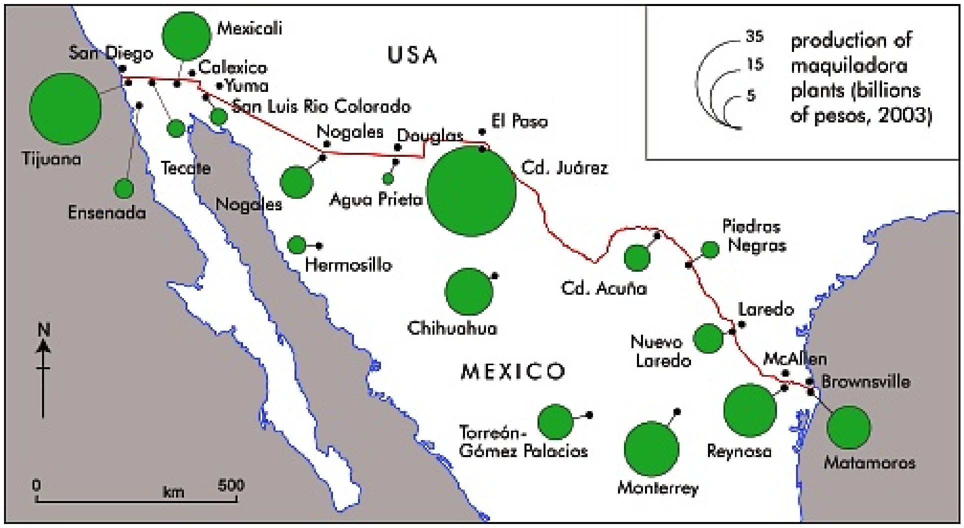


Fig. 5.1: Tijuana boasts the 2d largest annual Maquiladora output in Mexico.

crossing in the world, Tijuana is an ideal trade center. Not surprisingly, the maquiladora program expanded from solely US to international participation. Today, Tijuana hosts maquiladoras for over 100 multinational companies including, Sanyo, Sony, Volkswagon, and Panasonic, Table 5.1. [123]

Tab. 5.1: Mexican Maquiladora Border Towns have seen consistent grow for over 30 years.

	Number of Operating Plants	Employment (000)	Total Exports ($Billion)
1980	620	123.9	2.52
1985	758	217.5	5.08
1990	1,920	446.3	14.09
1995	2,267	681.3	31.1
2000	3,703	1,310.00	79.47
2002	3,240	1,090.00	57.8
2006	2,813	1,210.00	81.9
2010	2,454	1,400.00	117.6
2011	2,448	1,450.00	132.2

World market growth along with genial Mexican national and maquiladora city industrial licensing has fostered steady employment. The increase in citizens has not yielded similar expansion of suitable living environments for its laborer inhabitants. The pull of strong employment has led to overcrowding and a large infrastructure capacity gap. Access is outstripped by growth. [124] Regionally, 40% of all water consumed is untreated with 10% of the population without any immediate access to water. [125] This lack of access to water can lead to widespread disease, malnutrition, and public health degradation. [118]

The immigrants' limited access to basic services has secondary effects as well. Lack of electricity disincentives growth, keeps quality of life artificially low, limits access to information, and extends the "working masses" myth where businesses and municipalities are insensitive to labor force quality of life, i.e. increased risk due illegal electrical networks and service theft. The poor conditions threaten stability and may reduce the probability of continued investments.

Despite decades of growth, labor unrest is common in Mexico. Recent history points to an unforseen agent, the North American Free Trade Agreement (NAFTA). NAFTA was designed to reduce or eliminate tariffs and trade barriers between Canada, the US, and Mexico. [126] It further incentivised investment and growth in Mexican maquiladoras from the date of its signing, 17 December 1993. Unexpectedly, it was not well received in Mexico. Its signature coincides with the beginning of the influential and pro-labor

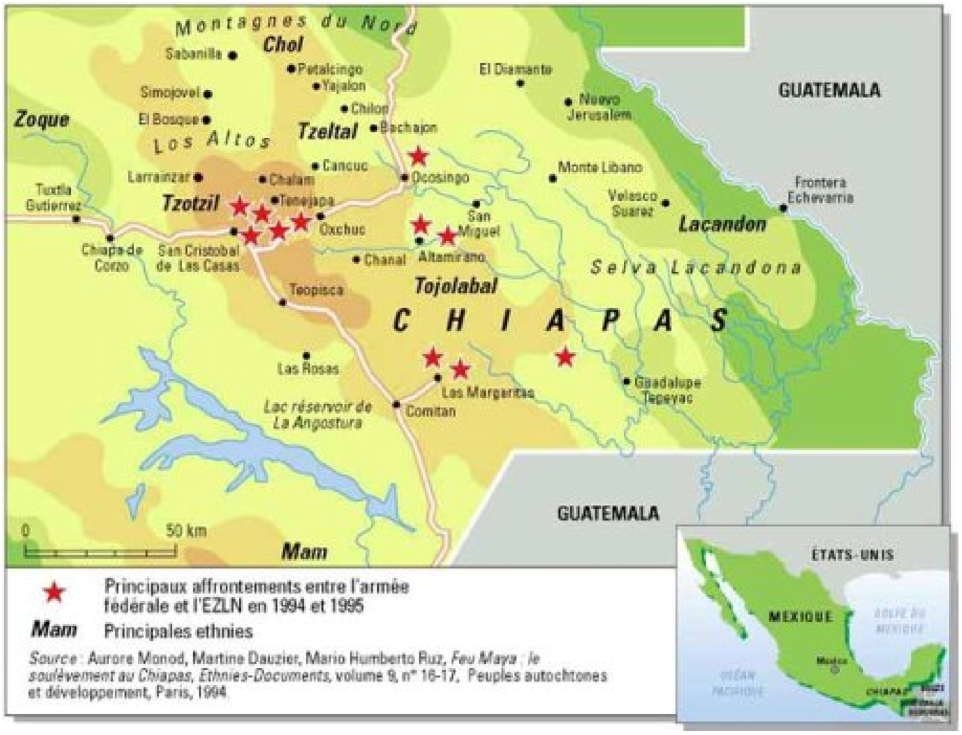


Fig. 5.2: Zapatista protests in Chiapas led to armed conflict in 1994 and 1995. [1]

Zapata movement. The movement began the same day as the signature in the southern state of Chiapas. [127] Within two years of its inception, there were over 200 Zapatista led strikes and protests in Chiapas alone. They led to direct, armed conflict with the Mexican National Army, Figure 5.2. [127]

Chiapas has interesting parallels to Afghanistan in that it has high ethnic diversity in a small area where more than half of the population does not speak the national language. It is likely the unrest began due to subpopulations feeling marginalized. It took hold with laborers who felt that greater trade would not show requisite improvements in laborer quality of life. [128]

The Zapatista movement was spread by interstate immigration and unified dissatisfied labors in all 31 states by 2006. [129] Throughout Mexico, laborers claimed that the government colluded with industry to hold wages low to hold down costs and spur more investment. [128] Laborers demanded social and economic upward mobility and the ability to organized unions without interference. [128] Tijuana was not immune to the unrest seeing a series a strikes in the mid-1990s and early 2000s. [130, 131].

During the past few decades, China's economy has experienced similar growth to Mexico due to manufacturing expansion and a ready labor force. [132] As with Mexico, China's winnings have come with a social cost. Failure to provide requisite quality of life improvement and gains for the people have resulted in unrest, an uncommon phenomena in the Middle Kingdom. Labor disruptions at Foxconn, the largest *Apple Iphone* component supplier, has brought undue attention and disruption to the supply chain while increasing manufacture costs. [133] Concern for the environment and public health has motivated protests in Shanghai as well. [134] Business analysts attribute recent labor force shifts away from China to other countries to the unrest. [135] Populations seem willing to sacrifice to establish fiscal momentum but not to sustain it. Sustainable manufacturing growth should account for laborer quality of life.

There are other cases for achieving labor stability in the face of growth. After World War II, the Japanese recovery and subsequent economic miracle showed a markedly different approach to labor during its industrial explosion.

From the 1950s until the 1990s, the Japanese economy saw 9-10% annual growth due to its manufacturing and technology. [136] Contrary to China and Mexico, Japanese firms and their laborers benefit from mutual loyalty. In exchanged for their effort, laborers can count on a living wage, medical disability, survivor benefits, and life long employment. [137, 138] Some companies also subsidise housing and dormitories. [138] The Japanese system is not perfect. Its laborers and managers are often at risk from over-work. [139] However, insurgency stemming from unrest is unheard of since the growth began.

Japan's industrial investments have been largely domestic resulting in slower growth when compared to China or Mexico. Perhaps the fealty and concern for access to basic services shown within Japanese industry has its roots there. It can provide an azimuth for the maquiladora system. The high employment in Mexico has not broken the poverty cycle as it did in Japan. Low pay has held the interest of investors but has not served the population as a whole. [128]

In an effort to analyze the situation for maquiladora immigrants and underpin social stability, researchers at the Center for Connected Learning and Computer-Based Modeling (CCL-CBM), Northwestern University, modelled the socio-economic environment in Tijuana. Their Tijuana Bordertowns model serves as a platform to examine the impact of policy decisions and migration rates on border region immigrants. [28] The CCL-CBM Tijuana model demonstrates the impact of industrial growth, regulation, and capacity development via a variety of statistical measures.

5.3 Problem Statement

As a hedge against instability and to ensure the long term viability of maquiladora investment, Mexican national and municipal leaders, corporations, and other stakeholders wish to minimize the fraction of immigrant labor population excluded from reasonable shelter, potable water, and reliable electricity in Tijuana, Mexico. MMRDP will recommend robust policy solutions based on an ABM that models a subregion of Tijuana.

Given I control factors and J exogenous factors, recommend control factor settings that minimize the response, S, providing outcomes that are locally optimal within the response space and somewhat insensitive to system variability. In this study, the response the long term Tijuana immigrant basic services exclusion rate.

$$\text{Minimize } S \sim F(X, Z) \tag{5.1}$$

subject to

$$x_i \leq x_{U(i)} \tag{5.2}$$

$$x_i \geq x_{L(i)} \tag{5.3}$$

$$\bigvee i \in I, I = 1..6 \tag{5.4}$$

$$Factor_j \sim G_j \tag{5.5}$$

$$\bigvee j \in J, J = 1..4 \tag{5.6}$$

$X_L, X_U -$ control variable upper and lower bounds.

$G -$ exogenous variable general distribution

5.4 Methodology

A systems approach to the problem consists of application of MMRDP centered on Tijuana's socio-economic, immigrant, and labor environments. The model measures, via simulation, the steady state percentage of the immigrant population without access to potable water. It employs experimental design, robust response surface methodology via dual response surface optimization and an optima filter for factor discrimination and best setting selection. It provides a solution strategy to analyze the response space and underpin infrastructure development policy in Tijuana, Figure 5.3.

5.4.1 Tijuana Agent-Based Model

In Tijuana, immigration rates and settlement patterns can be sporadic and unpredictable while infrastructure growth is sluggish, expensive, and reactive. The NetLogo ABM simulates elements of the complex environment and the socio-economic features of immigrant settlements in Tijuana, Mexico. [19]

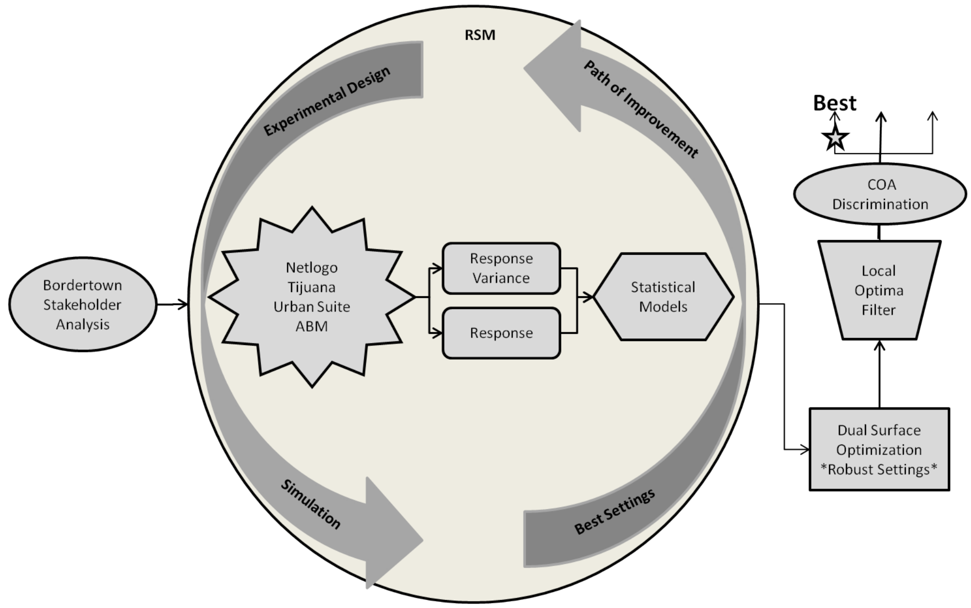


Fig. 5.3: The meta-model process incorporates stakeholder analysis, ABM simulation, and dual surface RSM for distinct robust solutions.

Complex settings, such as Tijuana's environment, often yield counterintuitive and emergent behavior that would not otherwise observers or considered. [140]

The Tijuana Bordertowns project features agents moving from other Mexican states to work in maquiladoras in Tijuana City. [28, 141] Agents in the simulation represent individuals whose attributes and traits are sampled from a demographic model. [141] The agent-based model simulates behavior for thousands of immigrants over a two year period for a given infrastructure policy in a small section of Tijuana, Figure 5.4. On a quad core system, the iteration rate is 2-4 per minute.

Immigrants arrive randomly $\sim exp(\lambda)$ seeking work, housing, and access

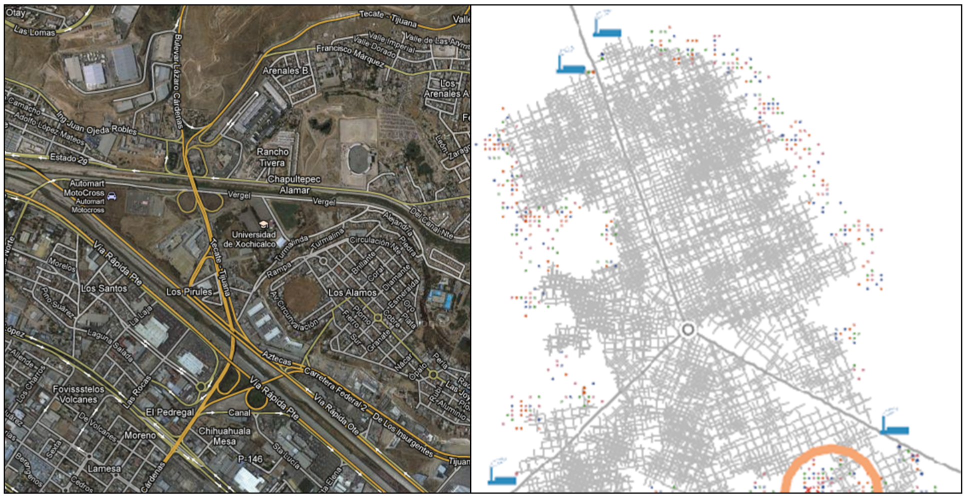


Fig. 5.4: Agent-Based Model of Tijuana City in the vicinity of Chihuahuala La Mesa.

to water and electricity. If immigrants arrive with savings, their prospects for basic services are better. Other, more veteran, laborers use their wages to pay bills, food and rent, and save the rest. An agent can stay put, move to better climes within Tijuana, or seek employment and residential status across the border in the United States. Generally, laborers seek better living conditions throughout a simulation run. The rate of upward mobility, a random feature of the model and a function of agents' savings rates and job availability, is positively correlated with the availability of regulated, acceptable housing and basic services within the city.

Within the model, agents (maquiladora laborers) arrive from other Mexican states, settle, and work. Initially, they live in any available location, typically the outskirts, and have little access to city utilities and services. The cycle of growth causes land values to increase and populations to move.

Agents are assessed costs for for food, transportation, housing, while trying to save their wages. After the simulation runs for two years, statistical stability is achieved in terms of the fraction of the population with out access to potable water and electricity. It is this rate of exclusion that the process aims to minimize, Figure 5.5.

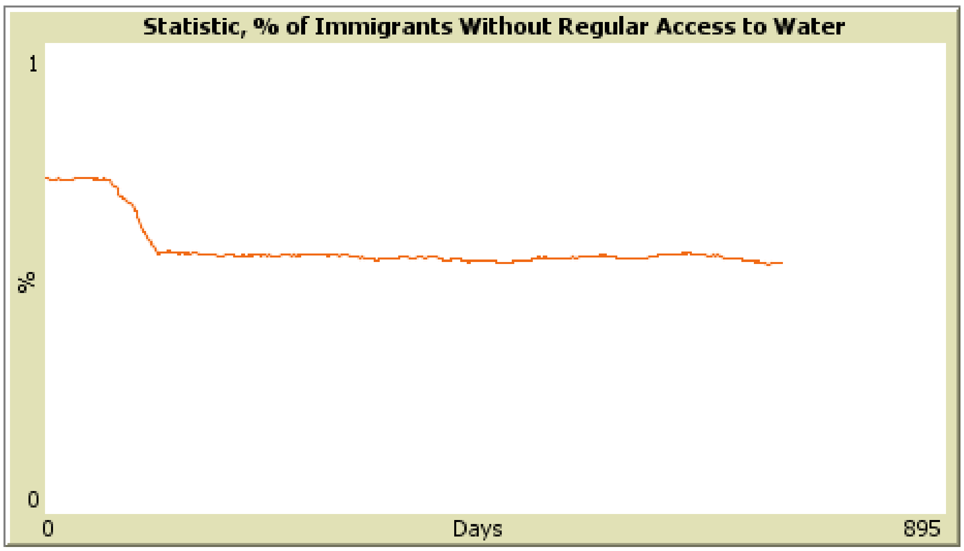


Fig. 5.5: The simulation statistic of interest stabilizes after two years of simulation run time.

The city and industry play a role in the simulation. The city can decide to periodically assess the state of the labor population. It then has several options to address service needs. It can rearrange neighborhood boundaries to improve service point access. It can more strictly regulate service delivery systems. It can build housing and service capacity at varied rates. For instance, the city can rapidly build to try and meet demand or it can build infrastructure slower but provide greater carrying capacity, capacity that exceeds current needs. Additionally, more factories can be built.

Among many agent attributes, the simulation tracks how many of the immigrant agents have regular access to potable water over time. The potable water access is a proxy for access to regulated electricity, reliable transportation, and adequate housing. The simulation accounts for preferred access by charging laborers rent via a graduated scale. As the city responds to growth and developed land becomes scarce, cost of living increases. This cycle continues with each wave of arrivals during a simulation replication. It important to note that the model does not calculate the cost to the municipality of development, rather it allows the use to vary policy and observe outcomes without economic constraints.

Maquiladora density, population density, city population growth, employment levels, and employee quality of life vary with respect to transportation, potable water access, and electricity access. [141] Access to low quality water sources increases disease rates, malnutrition, and negatively impacts long-term quality of life. [111] Electricity and transportation serve to improve quality of live and also ensure the workers have ready access to job opportunity and information. A satisfied, mobile workforce is important to the maquiladora investors for production, retention, and growth.

Along with other borders towns, the Mexican Trade Commission, and employers seek to sustain the pool of labor by providing an attractive alternative to other sources of income in Mexico. This analysis requires a sophisticated understanding of all relevant stakeholders and the socio-economic environment. It is optimized using MMRDP.

5.4.2 Evaluation of Infrastructure Portfolios

The Tijuana ABM has six control factors and four exogenous factors with various levels, Tables 5.2 and 5.3. Control factors are those settings one can change while exogenous factors are those that are measurable but unmanageable. [142]

Tab. 5.2: Tijuana ABM Control Factors are a mixture of integer and continuous variables.

	City Growth	Maquiladoras	Service Centers	Required Capital	Colonia Size	Carrying Capacity
Definition	Decision to include service growth as a regular part of policy and budget	Tariff-free manufacturing centers in ABM	Service providers of potable water and electricity	Average initial savings of new immigrants	Relative colonia size compared to current sizes	Service capacity of city infrastructure compared to current needs
Units	Binary	Factories	Centers	USD	Multiplier	Integer Steps
Levels	2	8	3	110	9	6
Min	On	1	1	150	0	1
Max	Off	8	3	1200	2	6

Tab. 5.3: The Tijuana ABM Exogenous Factors are continuous variables.

	Migration	Crossing Ticks	Initial Density	Building Ticks
Definition	Inter-arrival time between Tijuana immigration arrivals	Inter-arrival time between immigration waves to US	Population Density of Colonias (> 1 is undesirable)	Time between municipal assessments for Colonia improvements
Units	Days	Days	Fraction	Months
Mean	3	10	0.75	3.25
Variance	9	100	0.2	1

The decision space is much larger than the Jalalabad problem, three more total factors (one exogenous and two control) each with more levels. Full enumeration of the control space yields 285,120 design points. The global optimal point and its variance are unknown.

RSM can begin at any point in the control space. The study used three initial points. Each is a random selection from within the control factors; City Growth, Number of Maquiladoras, Number of Service Centers, Required Initial Capital for New Laborers, Colonia Size, and Infrastructure Carrying Capacity, Table 5.4. Each yielded a locally optimal and robust control setting recommendation via MMRDP.

Tab. 5.4: Each initial point is a random point in the control field.

Initial Point	**City Growth**	**Maquiladoras**	**Service Centers**	**Required Capital**	**Colonia Size**	**Carrying Capacity**	**Response Average**	**Response Variance**
1	1	3	2	800	0.75	6	0.411	0.457
2	1	2	2	820	0.25	4	0.446	0.00702
3	0	2	1	920	1.75	2	0.590	0.00725

Within the RSM, the Tijuana agent-based model was run iteratively using 2^6 full-factorial designs centered on each initial point and subsequent improving points. This *resolution-IV* experimental design, derived from a foldover *resolution-III* design creates orthogonal, noncolinear, experimental sets free from colinearity. [31] The designs are crossed with instances of exogenous variables, yielding replications due to address the stochastic nature of the model, *Migration Rate, Emigration Rate, Initial Density, Building Ticks.* Each experimental design run required 96 hours on a dedicated host server. The total computation time for the initial experiments was over 12 days. This was true for for successive experiments as well. The R Statistical Package, version 2.15, was used estimate the response surface and paths of greatest improvement leading to further experiments and three locally optimal points. [62] The average response value for each design point was fit using using a GLM regression model due to the mix of continuous and integral factors. The appropriate fit was determined by measuring deviance of first order, second order, and interaction models. Second order models with interactions produced the best fit at improving and locally optimal points.

Via a crossed array design, each design point incorporated 10 replications. The factors were fit via GLMs with an identity link to the design point estimated response values to allow for gaussian error estimates.

After reaching classic RSM local optimality, the response variable, *infrastructure exclusion rate*, was minimzed using dual response surface optimization. To facilitate this step, satisficing behavior was determined to be response values greater than or equal to the median, ≤ 0.43. In the neighborhood surrounding each locally optimal point, the factors were fit to variance data. In this case, LTS regression models were used as the response was distributed according to a χ^2 distribution. [22] The best fit for the variance models was found via second order models with interactions. Regression models for the response mean and variance were combined with the control factor limits within an optimization problem. The objective function minimized the *infrastructure exclusion rate* response subject to reduced variance levels and the constraints. The resultant nonlinear, mixed integer problem was solved using GAMS software, see Appendix F. [77] In each of the three instances, it yielded a robust optimal control setting recommendation. Finally, the local optima filter was used to discriminate between the robust course of action recommendations to provide a ranking in terms of likelihood of achieving suitable response behavior.

5.5 Results

From three random start points, MMRDP yielded three robust optimal recommendations, Table 5.5. MMRDP sought to minimize the value of the response however, the locally optimal robust recommendations increase bias to achieve lower variance. This is evident in the progress of the response values from initial point to classic RSM local optima to dual surface optima, Figure 5.6. Two of the three robust optima are statistically the same due to their overlapping 95% confidence intervals, Table 5.6 and Figure 5.7.

Tab. 5.5: MMRDP finds local and robust optima by experimenting with all initial points.

Local Optima	**City Growth**	**Maquiladoras**	**Service Centers**	**Required Capital**	**Colonia Size**	**Carrying Capacity**	**Response Average**	**Response Variance**
1	1	6	1	150	2	1	0.631	0.01640
2	1	2	1	840	0.5	3	0.394	0.00172
3	1	1	2	940	1.75	3	0.371	0.00243

Local Optima Filter Results

To address the statistical similarity of robust optima two and three, MMRDP employs the local optima filter. For robust optimal points, it estimates

Optima 1	City Growth	Maquiladoras	Service Centers	Required Capital	Colonia Size	Carrying Capacity	Response Average	Response Variance
Start Point	1	3	2	800	0.75	6	0.411	0.45712
Classic RSM	1	2	3	820	0.75	4	0.292	0.32423
Dual Surface	1	6	1	150	2	1	0.631	0.01640

Optima 2	City Growth	Maquiladoras	Service Centers	Required Capital	Colonia Size	Carrying Capacity	Response Average	Response Variance
Start Point	1	2	2	820	0.25	4	0.446	0.00702
Classic RSM	0	1	3	800	0.75	4	0.297	0.00230
Dual Surface	1	2	1	840	0.5	3	0.394	0.00172

Optima 3	City Growth	Maquiladoras	Service Centers	Required Capital	Colonia Size	Carrying Capacity	Response Average	Response Variance
Start Point	0	2	1	920	1.75	2	0.590	0.00725
Classic RSM	1	1	3	920	1.5	1	0.342	0.00257
Dual Surface	1	1	2	940	1.75	3	0.371	0.00243

Fig. 5.6: Three local optima are robust in the presence of exogenous factors, as shown by the reduction in variance from the corresponding classic RSM local optima.

Tab. 5.6: The 95% confidence intervals for optima 2 and 3 overlap providing no distinct best choice.

Local Optima	Response Estimate	Response Variance	Half Width	Upper Limit	Lower Limit
1	0.63	0.02	0.02	0.65	0.61
2	0.39	0.00	0.03	0.43	0.36
3	0.37	0.00	0.04	0.41	0.33

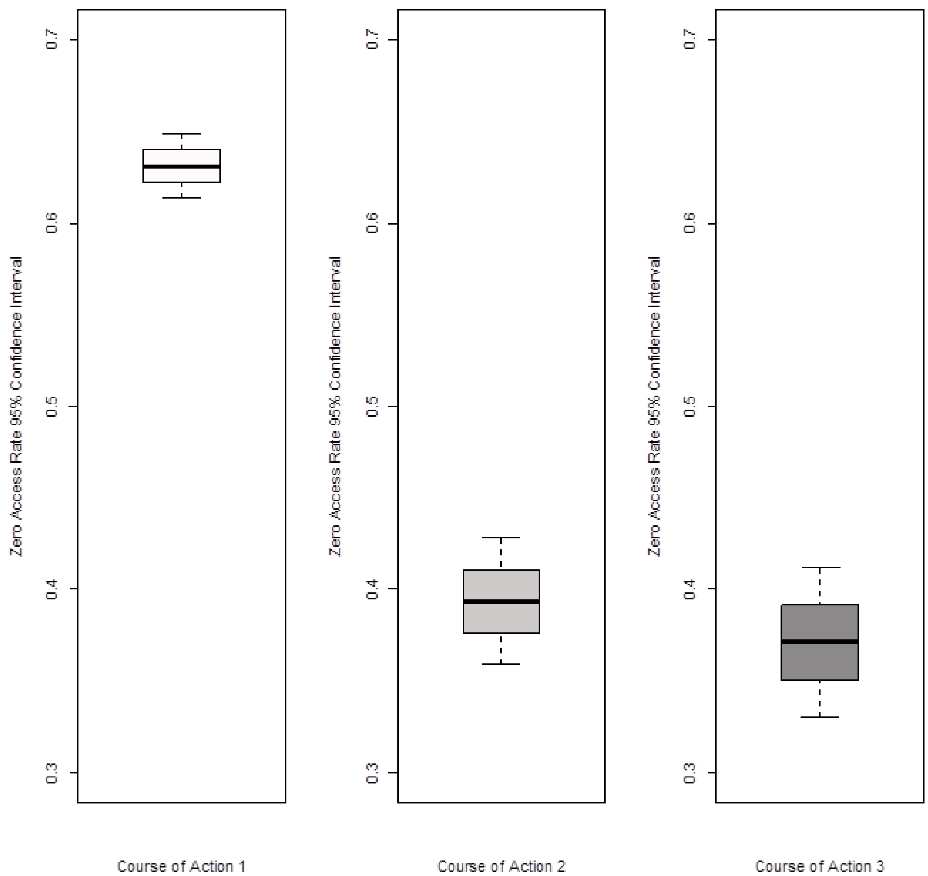


Fig. 5.7: The second and third robust optimal points show no statistically significant difference but they both outperform the first robust optimal point in terms of minimal response value.

the associated multi-dimensional surface area defined by the variability from the exogenous factors. Within the set being considered, the optima with the largest surface area has the largest associated probability of achieving satisficing behavior. In this study, the ellipsoid radii are derived by the length of the exogenous factor vector from the optima to the satisficing bound using fixed control settings. The satisficing bound is defined as long term immigrant infrastructure exclusion rates $\leq .43$, the median value of the response. Robust optima 1 fails consideration given the satisficing bound. Experimentation about optima 2 and 3 were used define PGDs from each optima to the satisficing bound. The corresponding PGDs were used define the two radii vector for optima 2 and 3 and create surface area estimates.

Tab. 5.7: Course of action 3 has a larger relative probability of achieving satisficing behavior than course of action 2 using the ellipsoid surface area estimates.

Exogenous Factor Satisficing Radii					
Robust Optima	**Migration**	**Crossing**	**Initial Density**	**Building**	**Surface Area**
2	8	85	0.7	0.5	5315.44
3	5	93	0.8	2.5	10959.18

Within course of action 3, it is possible to outline the effect of the exogenous factors further. The estimated relationship between the exogenous factor is second order and includes interactions with the control factors. However, a simple view of the relationship will suffice to indicate which exogenous factor has the most influence, Figure 5.8. The two most critical exogenous factors for response variability are population density and migration rate.

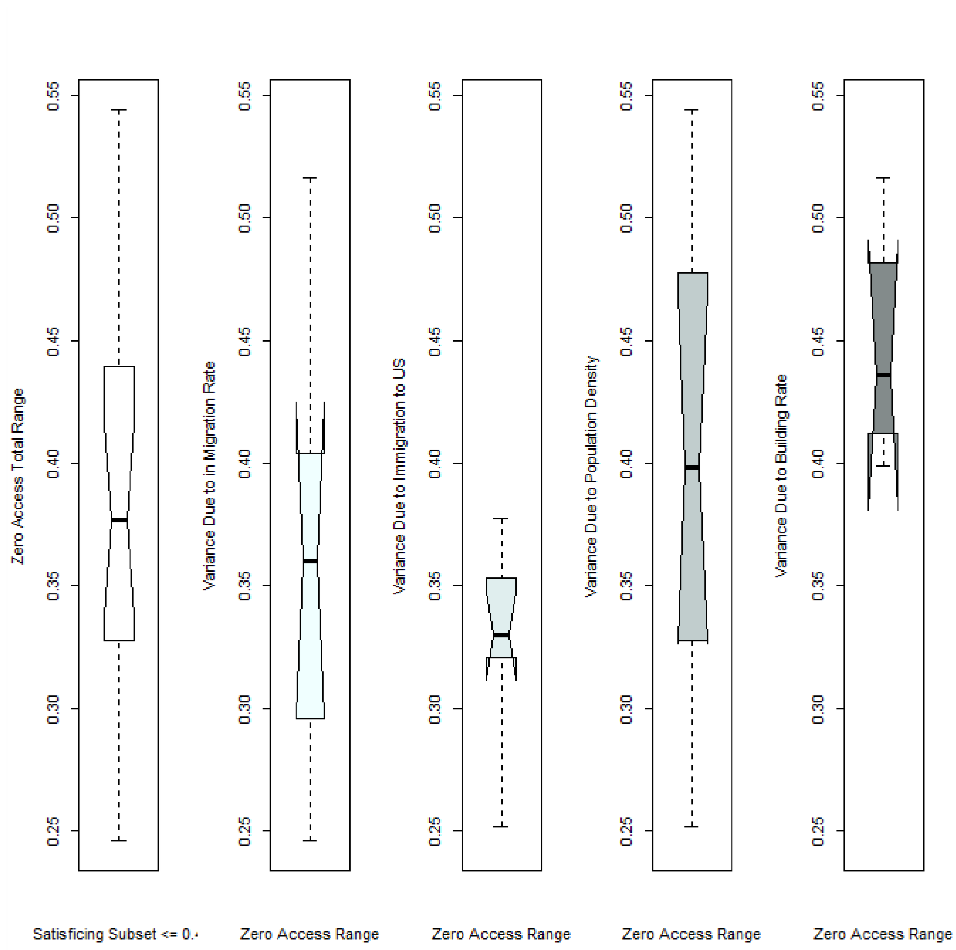


Fig. 5.8: The exogenous factors population density and migration rate induce the greatest variability in the response.

Pareto Efficiency

In the study, the economic effects and costs are not explicitly measured in order to consider a large range of public policies in Tijuana. The final step in MMRDP is an examination the courses of action for Pareto optimality. This is not possible in the traditional sense due to the lack of fiscal data. However, one may consider the number of maquiladoras as a proxy for cost. More maquiladoras is a more costly investment. This is an extremely simplified view as more maquildoras means more revenue for the state and factories as well. The evidence of this relationship is the near exponential manufacturing growth in Mexico since the 1980's.

The efficiency curve illustrate the tradeoff between bias and variance. Robustness can invite deliberate suboptimal response performance in order to reduce variability. The naive (nonrobust) approach generates a curve that outperforms the robust approach, Figure 5.9. The robust method curve recommends optima and course of action 3. The local optima filter provides additional evidence that optima 3 is the best course of action for the study, Table 5.8.

Tab. 5.8: Optima 3 is the MMRDP recommended setting.

Local Optima	**City Growth**	**Maquiladoras**	**Service Centers**	**Required Capital**	**Colonia Size**	**Carrying Capacity**	**Response Average**	**Response Variance**
3	1	1	2	940	1.75	3	0.371	0.00243

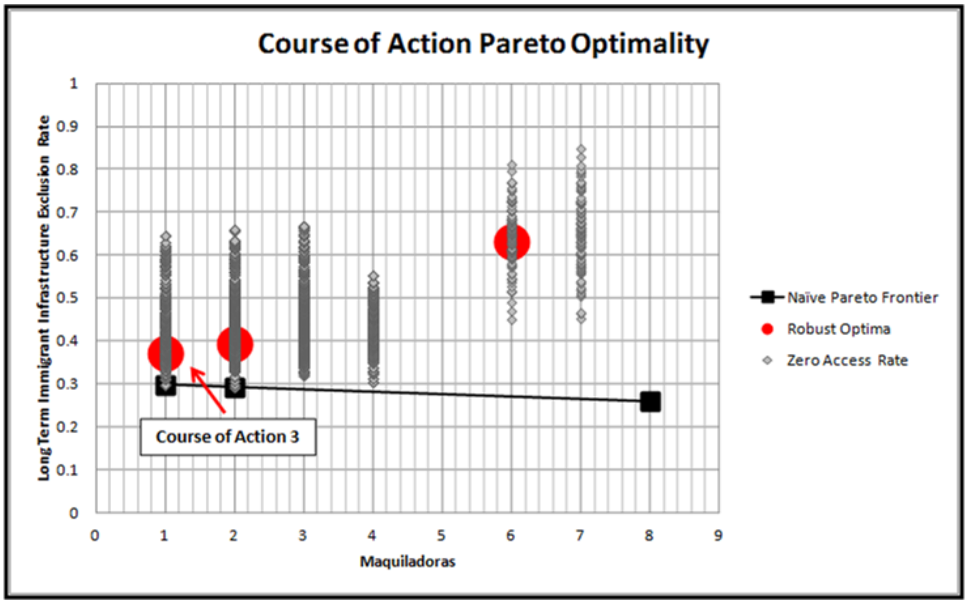


Fig. 5.9: The robust optima appear suboptimal but has the advantage of reduced sensitivity to exogenous factors.

5.6 Conclusion

Course of action 3 is a logical selection. It represents a very conservative approach to industrialization, likely due to the resources necessary to minimize the response. Without explicitly accounting for fiscal constraints or revenue, it is the most conservative course of action for the city and the businesses. The simulation recommends only 1 maquiladora and 2 services centers in the area of interest, but reality is different. In actuality there are at least five maquiladoras but only 1 service center in the same area.

Course of action 3 recommends that laborers arrive with a high level of savings. This is unrealistic the lack of employment opportunity elsewhere in Mexico for unskilled labor. However, it gives stakeholders an estimate of a per immigrant arrival cost for base infrastructure access.

The conservative approach of course of action 3 is also somewhat proactive. It recommends infrastructure capacity at three times the available rate. This may be an expense that is unfeasible given the reduced levels of revenue that course of action 3 may invite compared to current levels. However, it implies that at a minimum, more regular assessment of base infrastructure access must be measured to ensure appropriate infrastructure growth. More frequent and detailed assessment within the city may provide more impetus for policy change and point to more appropriate carrying capacity and colonia sizes in future planning endeavors.

In the representative area, the meta-model infrastructure portfolio selection process can require significant overhead to scope the problem and it also introduces bias. Despite that and the ABM construction effort to estimate relationships, the meta-model demonstrates flexibility and robustness in the face of complexity and varied random behavior. The process offers users the ability to trade speed and computational frugality for fidelity and detailed exploration of the response surface. Run multiple times, it can provide many feasible locally optimal settings to improve recommendations that could not otherwise be supported while still meeting minimum capacity screening criteria. It provides policy makers and stakeholders a means to justify enormous infrastructure investment in the name of stability by improving laborer quality of life.

In contrast to the Jalalabad model, the Tijuana model does not account for cost of infrastructure investment and policy changes. It is naive to assume that economic considerations will be disregarded. However, the long term cost of stability is real and its impact on maquiladora investment warrant careful consideration of municipal infrastructure investment to improve the laborers quality of life in ensure long-term stability. As an approach to improve the model, it may be used without economic features to eliminate unfeasible courses of action that do not support stable practices. Future versions of the model can then apply cost date to the smaller feature space to highlight efficient and acceptable solutions.

Mexico's maquiladora system is the result of ready labor, infrastructure, locale, and favorable business policy. However, leaders have failed to understand the tenuous position of the system. It is not robust to labor disruptions nor is it able to indefinitely absorb laborer living conditions. The desire to build upon past success is understandable, however the capacity gaps will likely worsen and reach a tipping point. At this point, continuing to ignore laborer quality of life is just as unprofitable as undertaking a plan to maximize immigrant infrastructure access. A moderate approach to stem the lack of infrastructure access is a necessary start but more elusive due to complexity. MMRDP can guide policy by offering better solutions within the morass of the current maquiladora environment in each city.

6. CONCLUSION

”We talk too much about the money and not enough about the benefits.”

∼ Thomas J. Donohue, President, U.S. Chamber of Commerce, on infrastructure development

6.1 Contributions

This research effort consists of a codified robust optimization methodology for complex adaptive systems and presents two major applications for instituting system of infrastructure systems solutions that address social stability. MMRDP demonstrates academic rigor and promise in application. Previous research validates the use of a systems approach that uses an ABM to represent interactions within and between social groups, a geographic locale, and infrastructure policy. The research also substantiates the use of generalized linear models and ordinary least squares or robust regression techniques within repeated application of response surface methodology to discover multiple locally optimal response values and corresponding setting sets.

MMRDP uses dual surface optimization to reduce solution sensitivity to exogenous behaviors and induce robust design. However within robust design, exogenous behavior can foul solution discrimination efforts. Variance is ubiquitous and confidence intervals may overlap making it difficult to rank order recommended, locally optimal, control settings. Estimating P-space satisficing response behavior hyper area addresses these confidence interval overlap problems. Comparing courses of action via the optima filter is an improvement on confidence intervals and associated methods. It presents users a better picture of performance in terms of the satisficing bound, it does not assume gaussian error, and it can represent high order relationships as necessary. Thompson's work in hyper-ellipsoid estimation on conjunction with steepest path approaches from RSM provides a rapid, accurate method that minimizes computational burden and takes advantage of simulation data.

MMRDP deliberately increases bias when comparing the robust solutions to basic RSM solutions. However, the variance reduction in the robust solution is consistent in effect but not value. The feasibility constraints can greatly limit the effectiveness of the variance reduction. Individual problem characteristics determine the ease of finding locally optimal points, regression models, variance characteristics, and robust solution selection.

Rather than seeking peak optimal points, MMRDP searches for satisficing behavior in high dimension spaces. In these nebulous environments this is an achievable goal in that it allows for the presence of control and exogenous factors. It also increases the probability of meeting stakeholder desires.

The application of P-space ellipsoid surface area to estimate hyper-surface area associated with desired response behavior to discriminate between courses of action is an advancement in the science. It provides a relative probability of success measure between candidate solutions for stakeholders.

MMRDP is well suited to analyze and optimize settings in the Jalalabad and Tijuana problems. While they are distinctly and dramatically different situations and environments, they are both related. They seek to improve the quality of life of the population in order to achieve stability measured via a proxy statistic. Both problems are complex, involve human behavior and infrastructure, and demand intense effort for proper stakeholder analysis and problem definition. They use simple rule sets to model interactions between agents, the environment, and infrastructure settings. They provide low variance estimates of satisficing behavior while bridging capacity gaps using noneconomic measures.

Application of MMRDP extends beyond development, disaster relief, and regional growth. Due to its ability to model facets of complexity, MMRDP can address system effects along other lines. For instance, it could be used to address outcomes of policy decisions within a number of complex systems. MMRDP would be useful on the study of recidivism within inhabitants of the US Criminal Justice System. Though a wildly different topic, MMRDP would be applicable in the study of the US energy system implementation of renewable energy policies. Given MMRDP's flexibility, it could also be used for optimizing multi-attribute training for individuals or achieving an outcome within a system of education. MMRDP can be a tool for a variety of decision makers.

6.2 Future Work

MMRDP can improve. Given a class of problems, computational alternatives within MMRDP should be considered. For instance, within RSM and Robust RSM regression techniques abound. GLM's need not be the only regression model construct. GAMs, support vector machines, MARS, and myriad choices can be selected based on statistical merit. A broader selection of application can improve MMRDP subcomponent selection recommendations to minimize MSE via bias and variance reduction.

MMRDP is the benefactor of data from the ABM and at the same time all findings hinge on the ABM's veracity. Validation is therefore paramount. Further study is required to achieve it as a consistent milestone within MMRDP. A highlight of validation comes from previous work on the Jalalabad problem. Preliminary modeling showed how marginal returns from hospitals and water investment were only available when paired with security. [91] This supports anecdotal evidence provided by experienced military veterans and reflects the recommendations of current US military doctrine. [3]

Validation is unlikely to be a guarantee however. One cannot assume to validate a model of a complex system. A complex system by its nature has emergent behavior. Emergent behavior is unforeseen and unanticipated, making validation impossible. Another method to improve veracity may be to investigate the use of Bayesian models with MMRDP. As stated earlier the inherent imprecision of the overall effort can be improved via input from experts. Bayesian models address uncertainty in a different way by updating a prior, initial, density function. It is improved by examining the likelihood that the observed data is well explained by the prior density function. They allow the practitioner to inject expert opinion into the model. [50] However, the computational complexity of Bayesian models makes their inclusion in MMRDP a significant challenge to overcome.

MMRDP is not a dynamic tool in that is assumes constant environmental characteristics when it generates models from observed data. To preform as a viable large scale, long term tool, MMRDP would benefit from a change detection function. This would require the inclusion of time series analysis at a minimum. Most time series analysis requires homoscedasticity and invariant model parameters. [143] These assumptions defeat the purpose of time series analysis in MMRDP. However, more current methods address time-varying parameters in series. [144] The computational burden of these models is high however they could be used to find seams in time and in space where parameters are fixed. This segmentation can improve the prospects of large models be defining boundaries for sub elements of a larger regional model. It can also identify where different infrastructure portfolios would be more effective for the population at different points in time. This would allow for a phased plan, an important feature for the time lines associated with infrastructure development.

MMRDP accounts for exogenous behavior but currently does not address measurement error. Measurement error research grew from automobile and other manufacturing processes. Gage reliability and reproducibility (Gage R&R) is a systematic process that limits measurement variability by analyzing operator variance. [145] It provides the practitioner, through random sampling, an understanding of variablity associated with measuring versus that associated with the value being recorded. While not an area focus for this research, it would have great value. It would ultimately provide practitioners with smaller levels of measurement error thereby making models and MMRDP filtering of recommendations better reflections of the true system behavior.

The robust design methods only add a degree of insensitivity to the exogenous factors. Within current RSM application, consideration has not been given to the accuracy of the input control variables. Work has been done in terms of risk analysis and RSM, Chen et al.(2003), and Al-Omar(2002). But this amounts to accounting for possible types of outcomes and assigning probabilities, as in classic application of risk analysis. However, an intuitive and useful notion of response error associated with RSM needs greater attention.

Variation and measurement error are base motivations for probability theory and statistics. [146] However, they are accounted for selectively by analysts. While a key component of models in physics, other models assume them away. In fact, Avkiran and Thoraneenitiyan (2010) state that productivity and service models, which can well represent relevant infrastructure models, rarely employ measurement error at all assuming that the variance of inputs is fixed and/or zero. [147] At a bare minimum, this practice yields an incorrect levels of estimated mean and estimated variance and can falsely assume probability distribution models for the same.

MMRDP would serve improve its scale. In its current form, the stakeholder analysis and lack of validation confidence limit MMRDP use to small regions and cities. Stakeholders may benefit from nation-level application. A proposed strategy would include stitching together small sets of stakeholder analysis efforts in a hierarchy where the local information would inform data about the larger area of interest. This multi-level approach has been used with success in social modeling. [148] The effort would require a preprocessing step to divide the larger regions an an initial step to segment the work. It would also require a more sophisticated understanding of infrastructure capacity within and between subregions. Optimality would be achieve both locally in each region and meta-locally within the entire gross area. Given the proper motivation to undertake the analysis and build models of each subregion, this would increase the viability of MMRDP as a tool on a much larger scale than demonstrated with this study.

MMRDP could be slightly altered to give stakeholders more choices within the factor settings via the optimal filter. Instead of fixing locally optimal control setting levels discovered via single and dual surface RSM and exploring the variance of the response around each due to exogenous factors, one could do the opposite. Fix the exogenous factors at their mean and explore which control settings illicit satisficing behavior or better. This method would give stakeholders more freedom to tradeoff performance and cost within the set surrounding each optima. On it own, this method would not be as conservative as the primary application of the optima filter. In order to maintain the conservative stance, portfolios that generate satisficing behavior would still require exploration of their behavior via PGD and corresponding hyper-ellipsoid surface areas for comparison and ranking.

Within the class of complex problems, MMRDP robust optimality is an obtainable and desirable objective. Governments and other organizations with finite resources that want to generate specific outcomes in complex dynamic environments have no other tool available with a multidisciplinary and codified methodology. MMRDP can serve as a general algorithm for infrastructure allocation evaluation in humanitarian relief missions or regional military operations. This research provides a necessary and useful extensions of RSM and a systems methodology with which to employ them.

MMRDP provides practitioners with a codified systems engineering process that generates tuned optimal system settings in the face of complex and dynamic environments. It combines critical aspects of stakeholder analysis, systems thinking, modeling, experimental design, computer aided simulation, response surface methodology extensions, robustness, optimization, and measurement error.

This tailored combination of methods is a unique in its construction and its output. Its use of an ABM to represent the environment and incorporate effects of geography on agents provides practitioners with ability to use simple rule sets and implement models for any location to gain unique results. The process estimates stochastic, high dimension, and nonlinear relationships with a univariate response. It generates a response surface that incorporates robust regression, robust design, and recommendations to stakeholders reflecting their understanding of control variables.

MMRDP is exclusive in its inputs, subcomponents, and analysis. It is best applied in nebulous and poorly understood environments. It provides solution recommendations in environments where it it very unlikely that one may select acceptable control measures on their own, particularly when desiring exogenous factor insensitivity. There exists no other process like MMRDP because typically one does not consider noneconomic evaluation of infrastructure systems. Within the robust application of RSM, stakeholder expectations can be misplaced if areas of improved response performance are not mapped and quantified. Robust filtering does exactly that. It is a critical step given the poorly understood environments that are suitable for MMRDP, such as second order effects of infrastructure investment.

MMRDP has shown its value via its properties and its application. It is a viable method to serve global and national agency ends by underwriting policy and strategy for long term population stability through improved quality of life. The desire to limit insurgency is real but the means are elusive. Insurgencies grow within marginalized people in ignored geographic regions. Once leaders determine an insurgency exists, it is too late to defeat it via conventional means. MMRDP may benefit from future study and refinement, it is ready to support decision makers and stakeholders in these difficult environments.

APPENDIX

A. TEST FUNCTIONS

A.1 DeJong's First Function in n Dimensions

$$f(x) = \sum_{i=1}^{n} x_i^2$$

Constraints

$$-5.12 \leq x_i \leq 5.12$$

Global Minimum

$$f(x) = 0$$

$$x_i = 0$$

$$i = 1..n$$

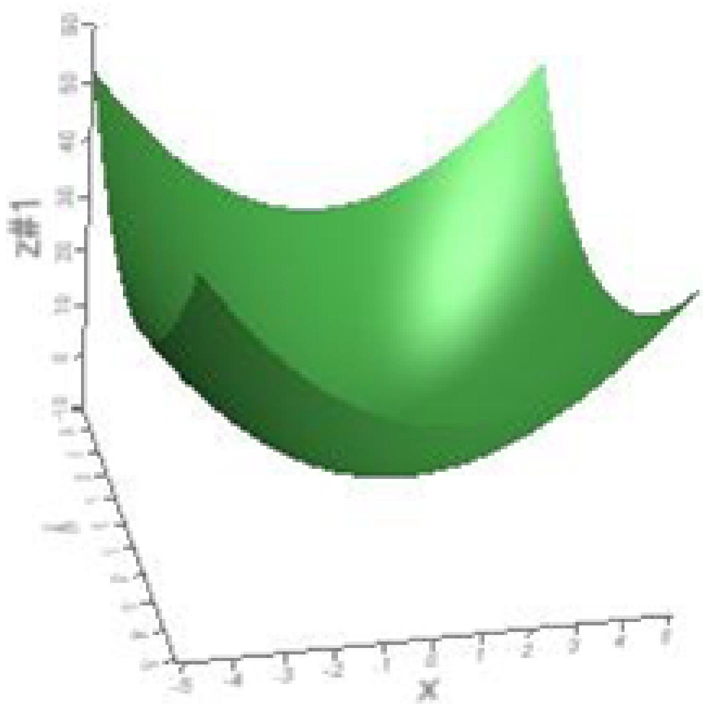


Fig. A.1: Dejong's First Function in 3 dimensions

A.2 Rosenbrock's Valley in n Dimensions

$$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$$

Constraints

$$-2.048 \leq x_i \leq 2.048$$

Global Minimum

$$f(x) = 0$$

$$x_i = 0$$

$$i = 1..n$$

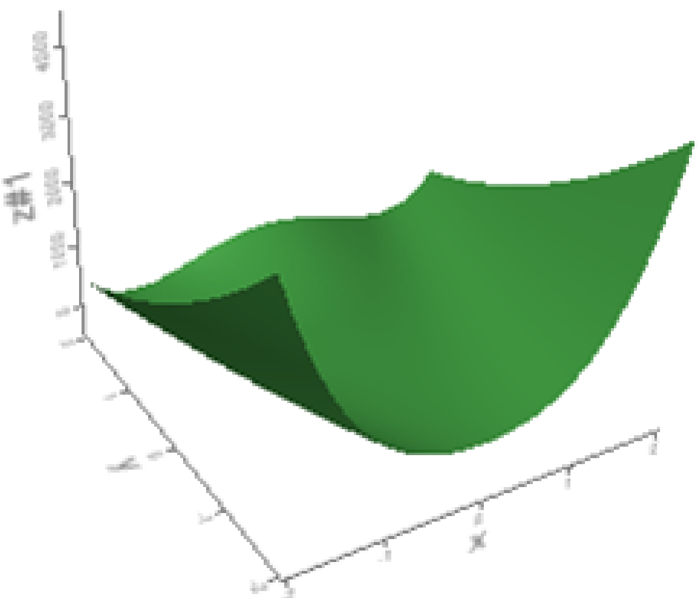


Fig. A.2: Rosenbrock's Valley in 3 dimensions

A.3 Rastrigin's Function in n Dimensions

$$f(x) = 10n + \sum_{i=1}^{n}[x_i^2 - 10\cos(2\pi x_i)]$$

Constraints

$$-5.12 \leq x_i \leq 5.12$$

Global Minimum

$$f(x) = 0$$

$$x_i = 0$$

$$i = 1..n$$

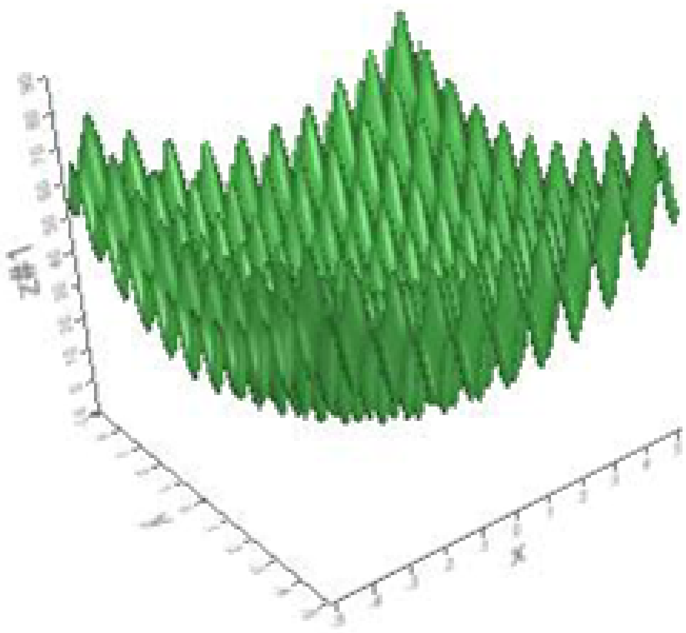


Fig. A.3: Rastrigin's Function in 3 dimensions

A.4 Schwefel's Function in n Dimensions

$$f(x) = \sum_{i=1}^{n} [-x_i \sin(\sqrt{|x_i|})]$$

Constraints

$$-500 \le x_i \le 500$$

Global Minimum

$$f(x) = -418.9829$$

$$x_i = 420.9687$$

$$i = 1..n$$

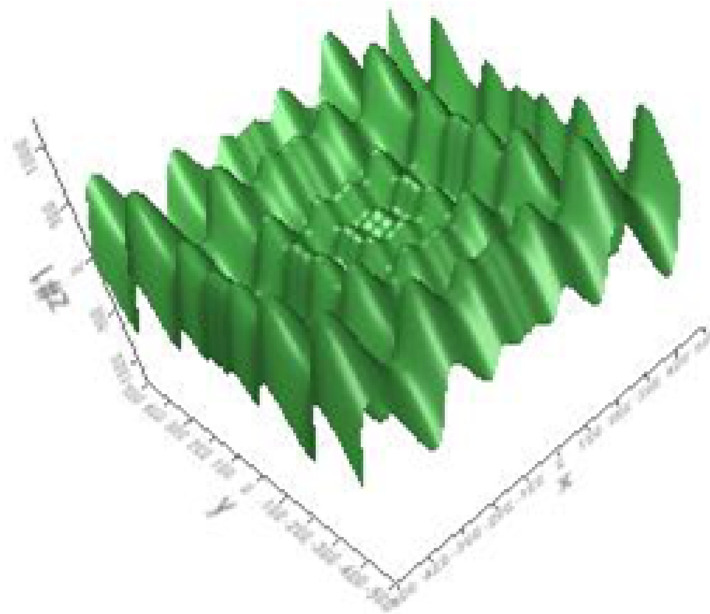


Fig. A.4: Schwefel's Function in 3 dimensions

B. TEST FUNCTION STOCHASTIC BEHAVIOR

B.1 Gaussian Random Variable Simulation Output

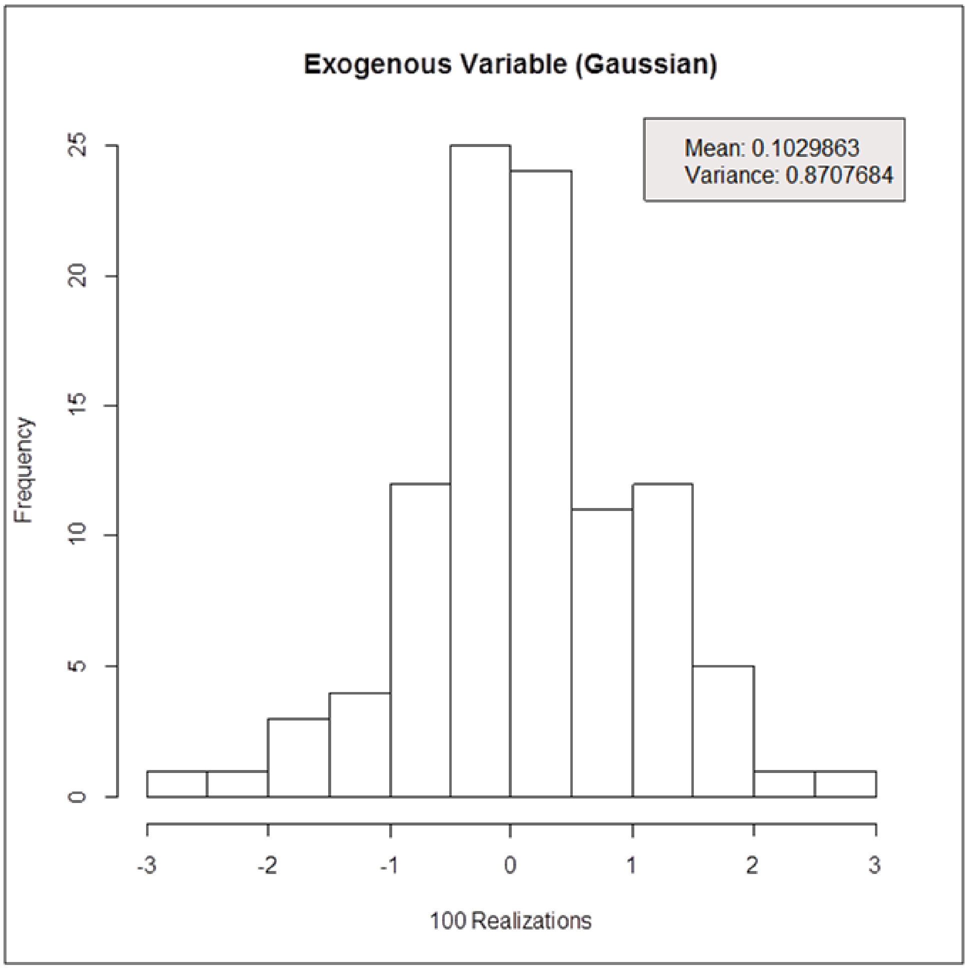


Fig. B.1: The simulated functions include realizations of standard normal behavior.

B.2 Uniform Random Variable Simulation Output

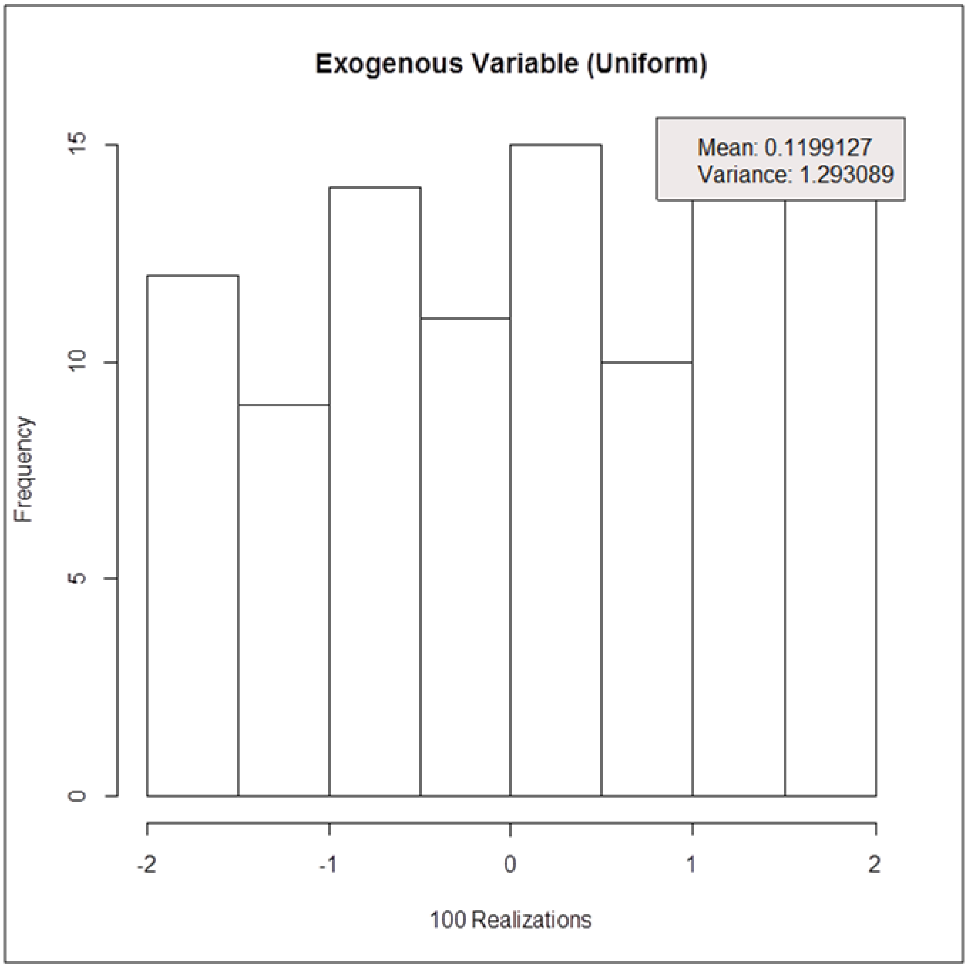


Fig. B.2: The simulated functions include realizations of uniform behavior centered on 0.

B.3 Exponential Random Variable Simulation Output

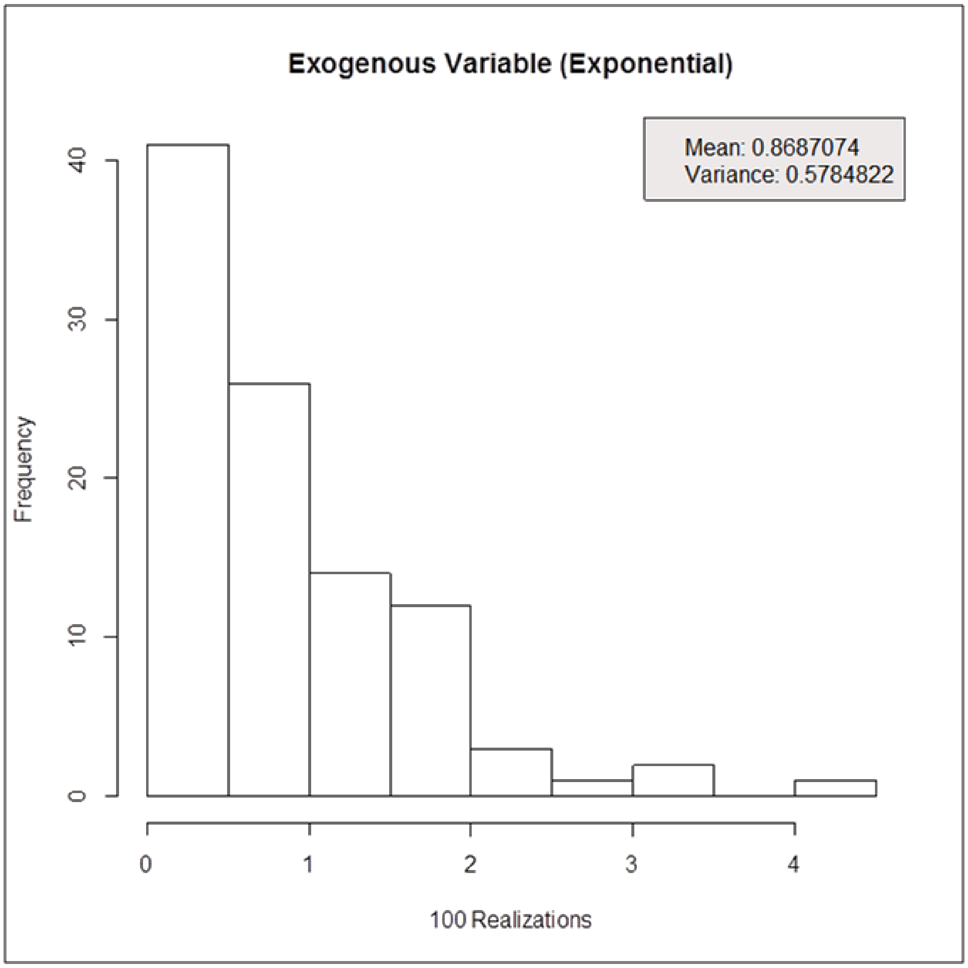


Fig. B.3: The simulated functions include realizations of exponential behavior with a mean of 1.

B.4 Sum of Gaussian and Exponential Random Variables Simulation Output

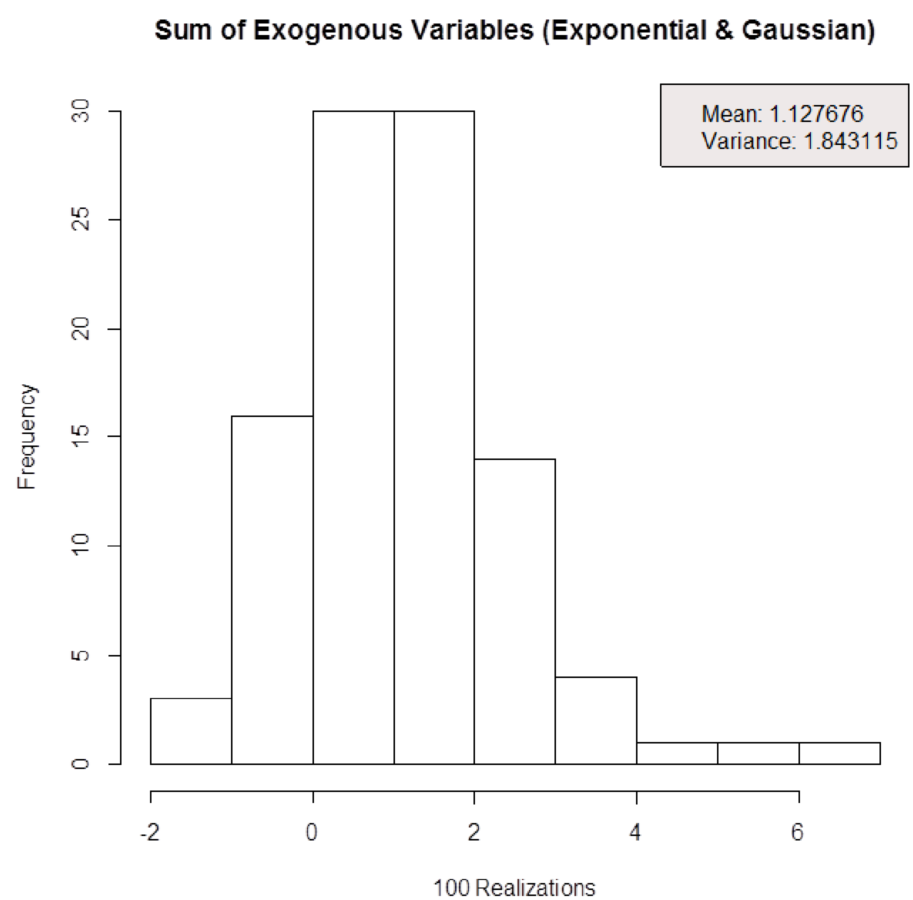


Fig. B.4: The simulated functions include realizations of the sum of exponential and gaussian behavior.

B.5 Sum of Gaussian and Uniform Random Variables

Simulation Output

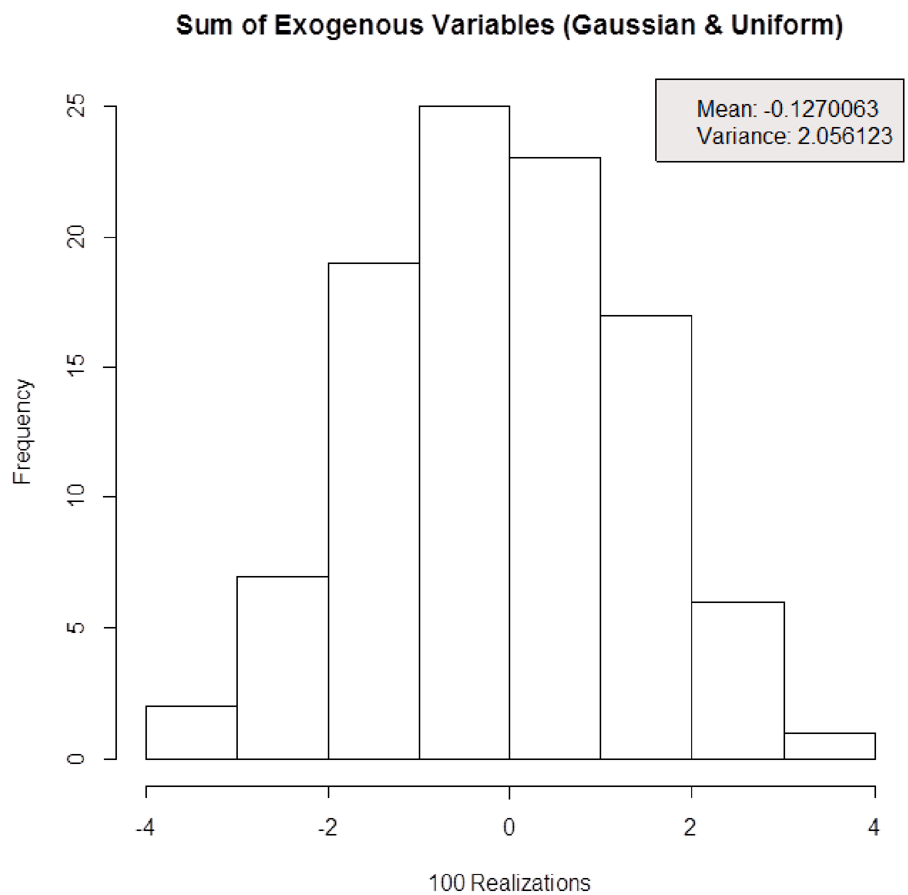


Fig. B.5: The simulated functions include realizations of the sum of uniform and gaussian behavior.

B.6 Sum of Exponential and Uniform Random Variables

Simulation Output

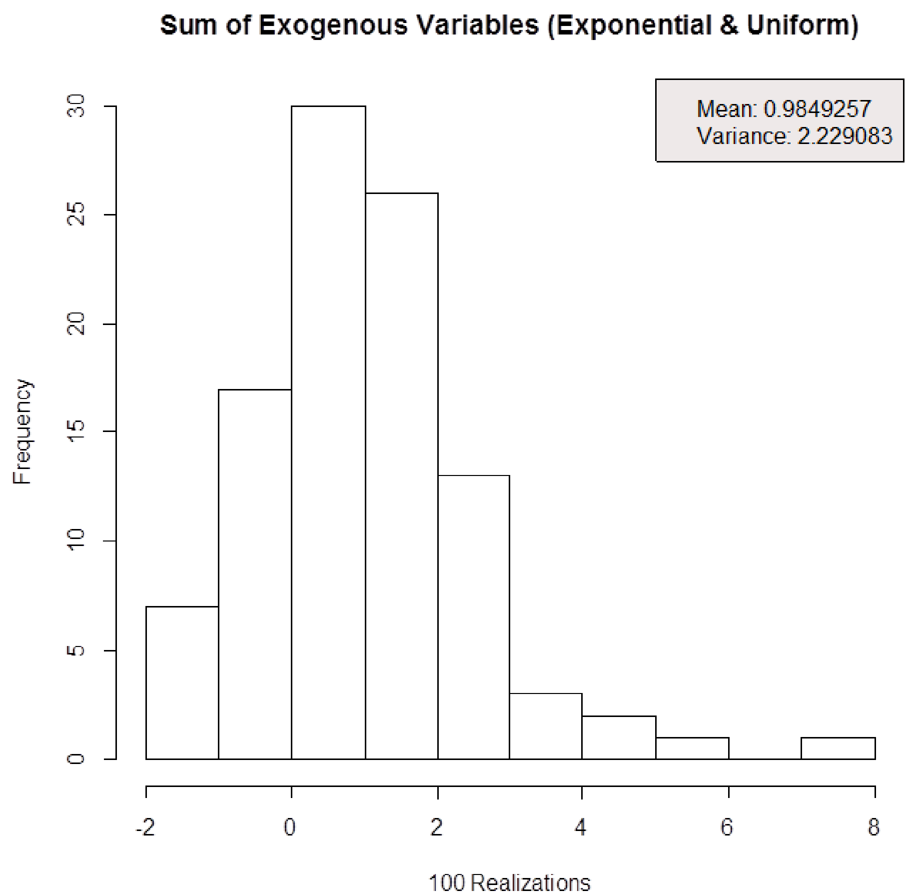


Fig. B.6: The simulated functions include realizations of the sum of uniform and exponential behavior.

B.7 Sum of Gaussian, Exponential and Uniform Random Variables Simulation Output

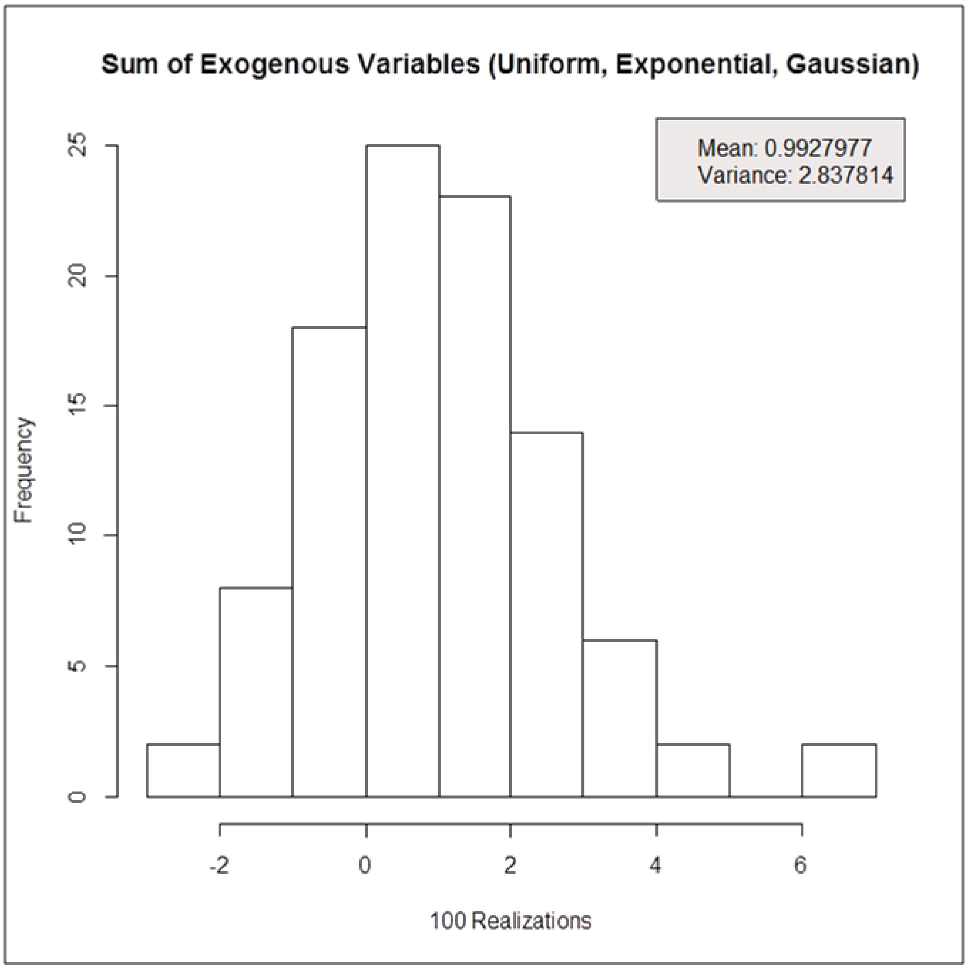


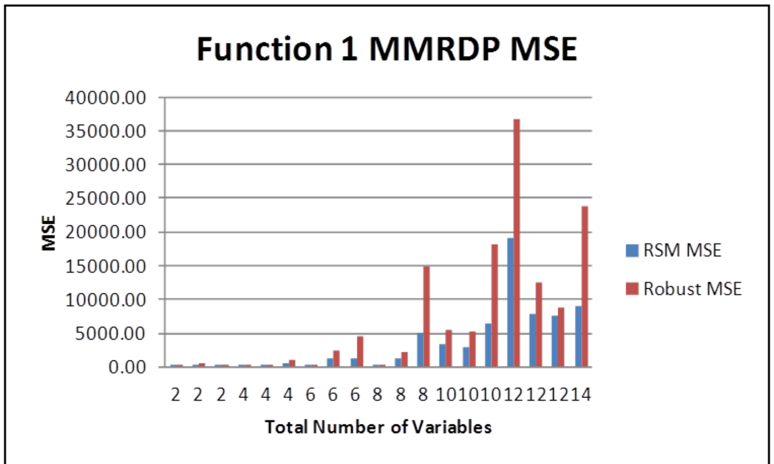
Fig. B.7: The simulated functions include realizations of the sum of uniform and exponential and gaussian behavior.

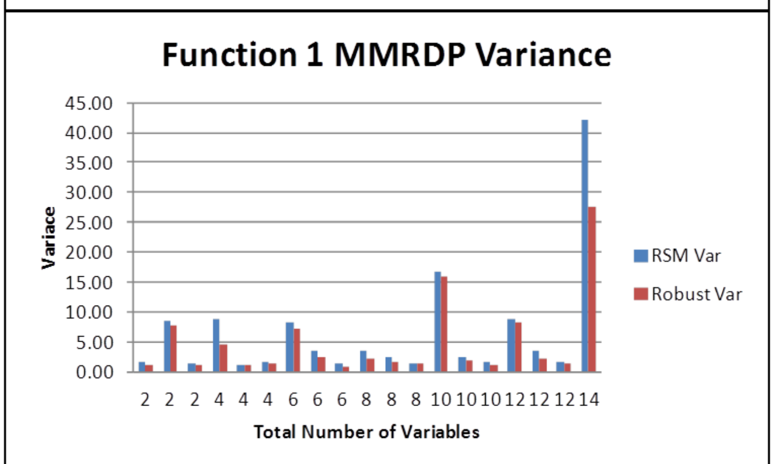
C. TEST FUNCTION MMRDP RESULTS

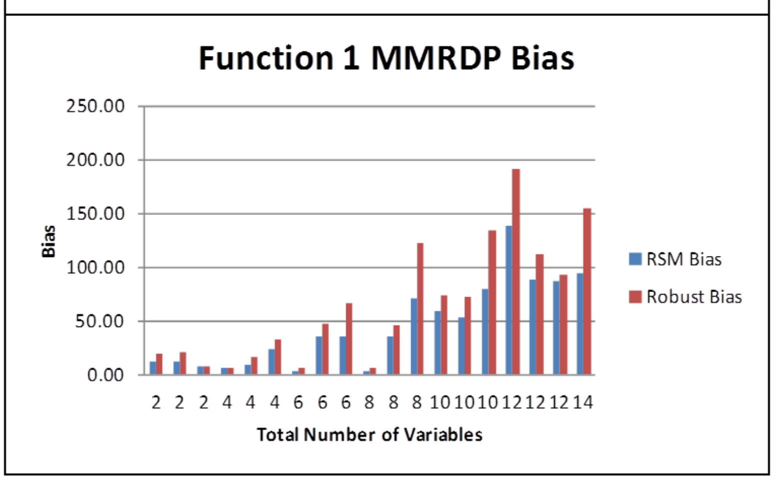
C.1 DeJong's First Function in N Dimensions with Stochastic Behavior

Func #	Exo Vars	Control Vars	Total Vars	Reps per DP	1-α Conf Lvl	Error Type	Known Opt Point	RSM Opt Point	RSM Var	RSM Half Width	RSM Opt Covers Known?	RSM Bias	RSM MSE	Robust Opt Point	Robust Var	Robust Half Width	RSM Opt Covers Known?	Robust Bias	Robust MSE
1	1	1	2	30	0.95	Gaussian	0.00	12.72	1.64	3.09	NO	12.72	163.53	18.97	1.23	2.67	NO	18.97	361.05
1	1	3	4	30	0.95	Gaussian	0.00	9.55	1.19	2.63	NO	9.55	92.41	15.97	1.02	2.43	NO	15.97	255.92
1	1	5	6	30	0.95	Gaussian	0.00	35.86	3.40	4.45	NO	35.86	1289.34	48.03	2.42	3.75	NO	48.03	2309.28
1	1	7	8	30	0.95	Gaussian	0.00	36.21	2.55	3.85	NO	36.21	1313.89	46.51	1.60	3.05	NO	46.51	2164.56
1	1	9	10	30	0.95	Gaussian	0.00	53.06	2.55	3.85	NO	53.06	2817.91	71.91	1.85	3.28	NO	71.91	5172.90
1	1	11	12	30	0.95	Gaussian	0.00	89.01	3.54	4.54	NO	89.01	7926.32	112.18	2.30	3.66	NO	112.18	12586.20
1	1	1	2	30	0.95	Non-Gaussian	1.00	13.89	8.42	7.00	NO	12.89	174.51	21.55	7.86	6.76	NO	20.55	430.33
1	1	1	2	30	0.95	Non-Gaussian	0.00	7.71	1.46	2.92	NO	7.71	60.88	8.12	1.07	2.50	NO	8.12	67.07
1	1	3	4	30	0.95	Non-Gaussian	0.00	24.29	1.58	3.03	NO	24.29	591.73	32.16	1.32	2.77	NO	32.16	1035.30
1	1	5	6	30	0.95	Non-Gaussian	0.00	35.90	1.42	2.87	NO	35.90	1289.99	66.77	0.83	2.20	NO	66.77	4458.46
1	1	7	8	30	0.95	Non-Gaussian	0.00	70.92	1.48	2.93	NO	70.92	5031.74	121.93	1.45	2.91	NO	121.93	14867.61
1	1	9	10	30	0.95	Non-Gaussian	0.00	79.52	1.54	2.99	NO	79.52	6325.50	134.63	1.15	2.58	NO	134.63	18127.54
1	1	11	12	30	0.95	Non-Gaussian	0.00	87.56	1.60	3.05	NO	87.56	7668.94	93.17	1.46	2.91	NO	93.17	8682.89
1	3	1	4	30	0.95	Non-Gaussian	1.00	6.62	8.87	7.18	YES	5.62	40.46	6.91	4.49	5.11	NO	5.91	39.43
1	3	3	6	30	0.95	Non-Gaussian	1.00	3.84	8.15	6.89	YES	2.84	16.22	6.76	7.20	6.47	YES	5.76	40.39
1	3	5	8	30	0.95	Non-Gaussian	1.00	4.96	3.46	4.48	YES	3.96	19.16	7.17	2.13	3.52	NO	6.17	40.23
1	3	7	10	30	0.95	Non-Gaussian	1.00	59.65	16.86	9.90	NO	58.65	3456.68	75.55	15.93	9.63	NO	74.55	5574.10
1	3	9	12	30	0.95	Non-Gaussian	1.00	139.11	8.72	7.12	NO	138.11	19083.09	192.84	8.18	6.90	NO	191.84	36809.33
1	3	11	14	30	0.95	Non-Gaussian	1.00	95.47	42.13	15.65	NO	94.47	8965.76	155.55	27.65	12.68	NO	154.55	23914.13

Fig. C.1: The results of MMRDP using Dejong's first function show the tradeoff between bias and variance when arriving at the robust solution.


Function 1 MMRDP MSE
40000.00
35000.00
30000.00
25000.00
20000.00
15000.00
10000.00
5000.00
0.00
MSE
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10 12 12 12 14
Total Number of Variables
RSM MSE
Robust MSE

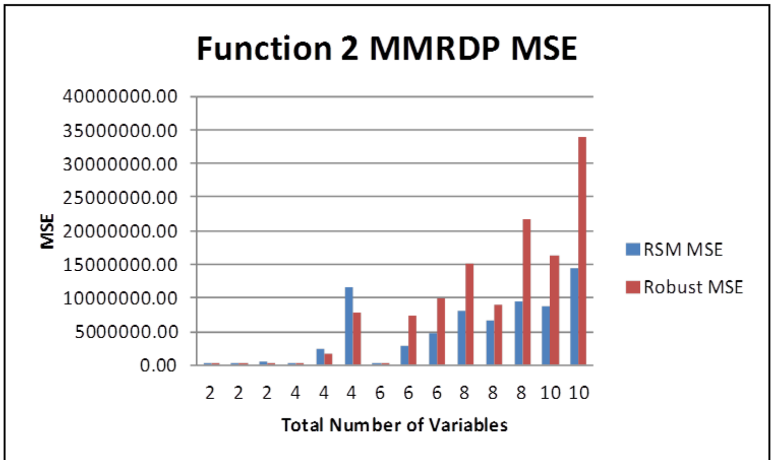

Function 1 MMRDP Variance
45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
Variace
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10 12 12 12 14
Total Number of Variables
RSM Var
Robust Var

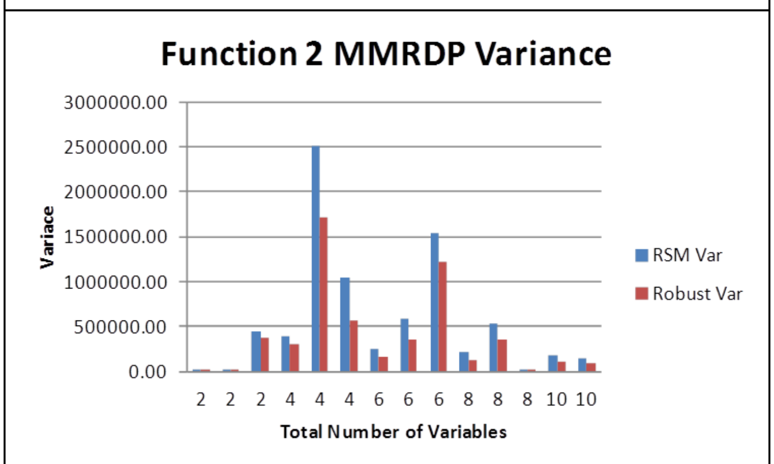

Function 1 MMRDP Bias
250.00
200.00
150.00
100.00
50.00
0.00
Bias
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10 12 12 12 14
Total Number of Variables
RSM Bias
Robust Bias

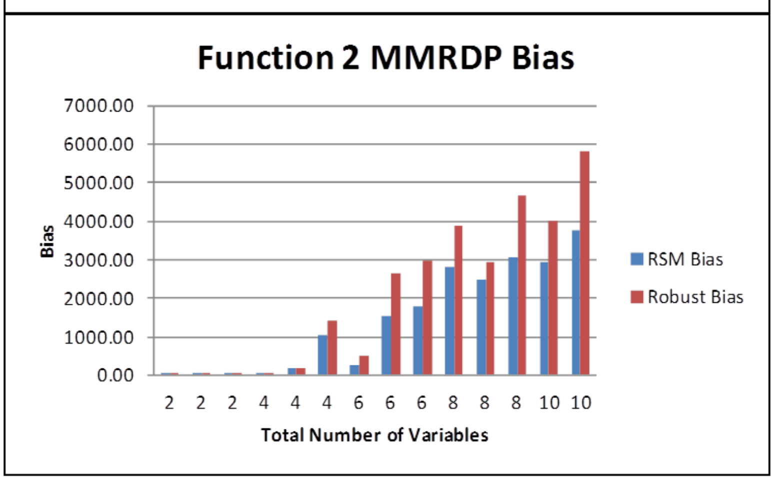
C.2 Rosenbrock's Valley in N Dimensions with Stochastic Behavior

Func #	Exo Vars	Control Vars	Total Vars	Reps per DP	1-α Conf Lvl	Error Type	Known Opt Point	RSM Opt Point	RSM Var	RSM Half Width	RSM HW Covers Known?	RSM Bias	RSM MSE	Robust Opt Point	Robust Var	Robust Half Width	Robust Hw Covers Known?	Robust Bias	Robust MSE
2	1	1	2	30	0.95	Gaussian	0.00	22.73	10303.00	244.76	YES	22.73	10819.65	42.62	7694.12	211.51	YES	42.62	9510.66
2	1	1	2	30	0.95	Non-Gaussian	1.00	15.39	13301.00	278.10	YES	14.39	13508.07	26.05	9623.40	236.55	YES	25.05	10250.74
2	1	1	2	30	0.95	Non-Gaussian	0.00	24.98	450350.00	1618.19	YES	24.98	450974.00	44.55	377772.89	1482.08	YES	44.55	379757.67
2	3	1	4	30	0.95	Non-Gaussian	1.00	17.04	394983.40	1515.46	YES	16.04	395240.75	19.35	298682.62	1317.83	YES	18.35	299019.33
2	1	3	4	30	0.95	Gaussian	0.00	177.98	2504318.17	3815.93	YES	177.98	2535993.29	185.12	1716430.61	3159.14	YES	185.12	1750701.36
2	3	3	6	30	0.95	Non-Gaussian	1.00	1025.10	1053173.52	7825.38	YES	1024.10	11580508.81	1429.24	575459.17	5784.46	YES	1428.24	7794473.08
2	1	5	6	30	0.95	Gaussian	0.00	273.91	255238.44	1218.23	YES	273.91	330263.61	500.76	163206.57	974.15	YES	500.76	413964.38
2	3	5	8	30	0.95	Non-Gaussian	1.00	1531.98	591828.23	1855.04	YES	1530.98	2935730.34	2635.70	353091.74	1432.85	NO	2634.70	7294714.85
2	1	7	8	30	0.95	Gaussian	0.00	1765.53	1543035.27	2995.32	YES	1765.53	4660125.92	2967.22	1213128.18	2655.88	NO	2967.22	10017500.73
2	3	7	10	30	0.95	Non-Gaussian	1.00	2792.91	220017.10	1131.05	NO	2791.91	8014750.63	3860.48	135321.66	887.03	NO	3859.48	15030889.89
2	1	9	10	30	0.95	Gaussian	0.00	2472.25	538159.17	1768.93	NO	2472.25	6650162.75	2935.94	351930.14	1430.49	NO	2935.94	8971683.38
2	3	9	12	30	0.95	Non-Gaussian	1.00	3073.41	3073.41	133.68	NO	3072.41	9442788.91	4660.02	1892.77	104.91	NO	4659.02	21708339.03
2	1	11	12	30	0.95	Gaussian	0.00	2926.06	181229.60	1026.53	NO	2926.06	8743060.63	4022.48	103037.43	774.02	NO	4022.48	16283416.57
2	3	11	14	30	0.95	Non-Gaussian	1.00	3772.88	153455.90	944.60	NO	3771.88	14380557.27	5826.01	100840.84	765.73	NO	5825.01	34031630.89

Fig. C.2: The results of MMRDP using Rosenbrock's valley function show the tradeoff between bias and variance when arriving ar the robust solution.


Function 2 MMRDP MSE
40000000.00
35000000.00
30000000.00
25000000.00
20000000.00
15000000.00
10000000.00
5000000.00
0.00
MSE
2 2 2 4 4 4 6 6 6 8 8 8 10 10
Total Number of Variables
RSM MSE
Robust MSE

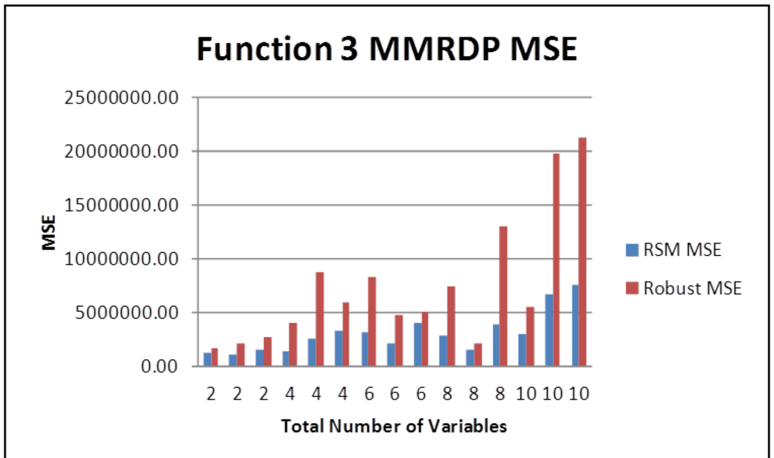

Function 2 MMRDP Variance
3000000.00
2500000.00
2000000.00
1500000.00
1000000.00
500000.00
0.00
Variace
2 2 2 4 4 4 6 6 6 8 8 8 10 10
Total Number of Variables
RSM Var
Robust Var

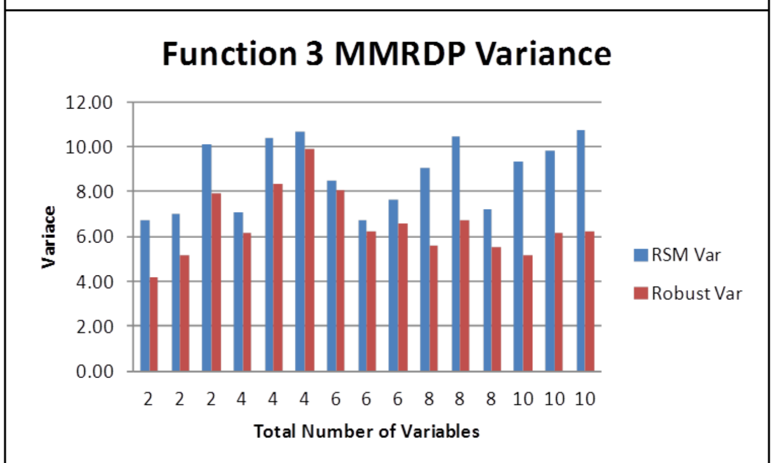

Function 2 MMRDP Bias
7000.00
6000.00
5000.00
4000.00
3000.00
2000.00
1000.00
0.00
Bias
2 2 2 4 4 4 6 6 6 8 8 8 10 10
Total Number of Variables
RSM Bias
Robust Bias

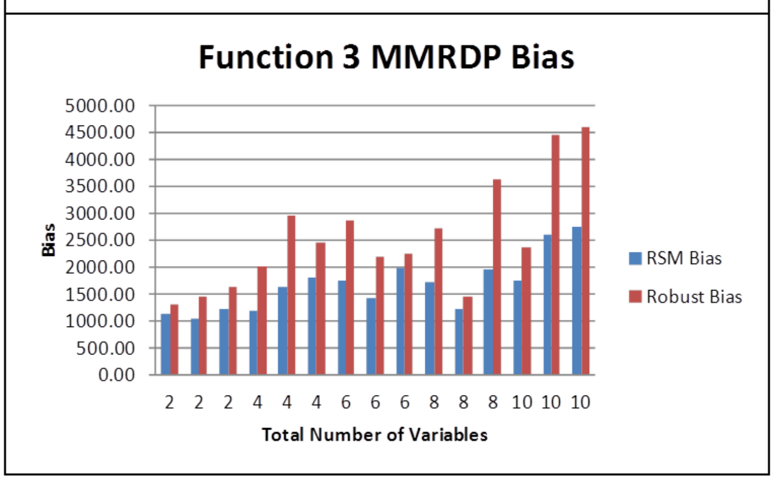
C.3 Rastrigin's Function in N Dimensions with Stochastic Behavior

Func #	Exo Vars	Control Vars	Total Vars	Reps per DP	1-α Conf Lvl	Error Type	Known Opt Point	RSM Opt Point	RSM Var	RSM Half Width	RSM HW Covers Known?	RSM Bias	RSM MSE	Robust Opt Point	Robust Var	Robust Half Width	Robust Hw Covers Known?	Robust Bias	Robust MSE
3	1	1	2	30	0.95	Gaussian	0.00	1120.09	6.69	6.24	NO	1120.09	1254611.93	1314.09	4.15	4.91	NO	1314.09	1726845.43
3	1	1	2	30	0.95	Non-Gaussian	1.00	1026.52	6.99	6.38	NO	1025.52	1051701.86	1457.40	5.16	5.48	NO	1456.40	2121097.93
3	1	1	2	30	0.95	Non-Gaussian	0.00	1212.18	10.09	7.66	NO	1212.18	1469383.27	1634.33	7.95	6.80	NO	1634.33	2671027.22
3	3	1	4	30	0.95	Non-Gaussian	1.00	1191.95	7.04	6.40	NO	1190.95	1418363.36	2021.11	6.17	5.99	NO	2020.11	4080857.33
3	1	3	4	30	0.95	Gaussian	0.00	1613.57	10.42	7.78	NO	1613.57	2603618.52	2943.28	8.32	6.96	NO	2943.28	8662894.32
3	3	3	6	30	0.95	Non-Gaussian	1.00	1794.36	10.65	7.87	NO	1793.36	3216145.95	2435.41	9.92	7.59	NO	2434.41	5926385.55
3	1	5	6	30	0.95	Gaussian	0.00	1754.05	8.47	7.02	NO	1754.05	3076714.52	2864.59	8.04	6.84	NO	2864.59	8205899.45
3	3	5	8	30	0.95	Non-Gaussian	1.00	1431.02	6.73	6.26	NO	1430.02	2044971.91	2182.78	6.22	6.01	NO	2181.78	4760149.14
3	1	7	8	30	0.95	Gaussian	0.00	1989.99	7.61	6.65	NO	1989.99	3960070.84	2241.59	6.55	6.17	NO	2241.59	5024716.64
3	3	7	10	30	0.95	Non-Gaussian	1.00	1701.69	9.04	7.25	NO	1700.69	2892356.84	2712.55	5.59	5.70	NO	2711.55	7352514.79
3	0	9	9	30	0.95	None	0.00	1201.86	10.47	7.80	NO	1201.86	1444473.21	1445.04	6.72	6.25	NO	1445.04	2088157.16
3	1	9	10	30	0.95	Gaussian	0.00	1957.56	7.22	6.48	NO	1957.56	3832036.32	3608.93	5.54	5.68	NO	3608.93	13024404.64
3	3	9	12	30	0.95	Non-Gaussian	1.00	1730.01	9.35	7.37	NO	1729.01	2989469.30	2356.49	5.19	5.50	NO	2355.49	5548348.07
3	1	11	12	30	0.95	Gaussian	0.00	2586.23	9.79	7.54	NO	2586.23	6688599.33	4446.49	6.17	5.99	NO	4446.49	19771254.53
3	3	11	14	30	0.95	Non-Gaussian	1.00	2747.81	10.72	7.90	NO	2746.81	7544995.85	4606.73	6.23	6.02	NO	4605.73	21212753.95

Fig. C.3: The results of MMRDP using Rastringen's function show the tradeoff between bias and variance when arriving ar the robust solution.


Function 3 MMRDP MSE
25000000.00
20000000.00
15000000.00
10000000.00
5000000.00
0.00
MSE
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10
Total Number of Variables
RSM MSE
Robust MSE

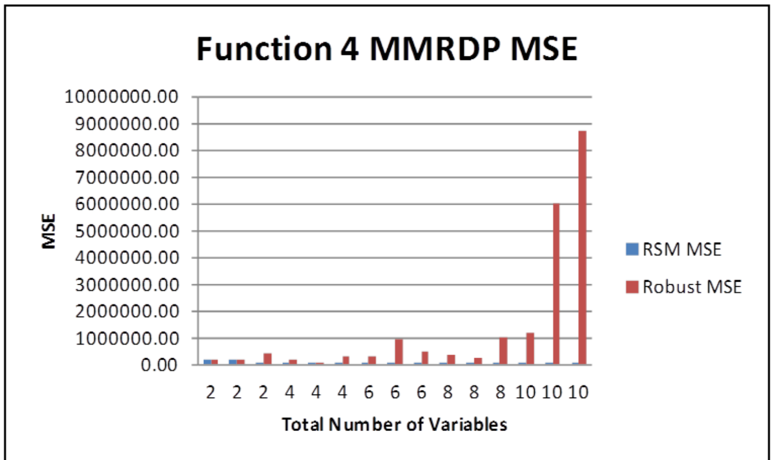

Function 3 MMRDP Variance
12.00
10.00
8.00
6.00
4.00
2.00
0.00
Variace
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10
Total Number of Variables
RSM Var
Robust Var

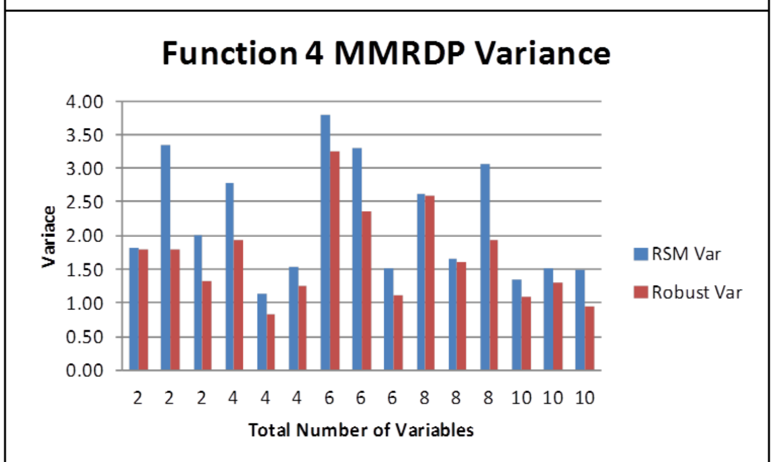

Function 3 MMRDP Bias
5000.00
4500.00
4000.00
3500.00
3000.00
2500.00
2000.00
1500.00
1000.00
500.00
0.00
Bias
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10
Total Number of Variables
RSM Bias
Robust Bias

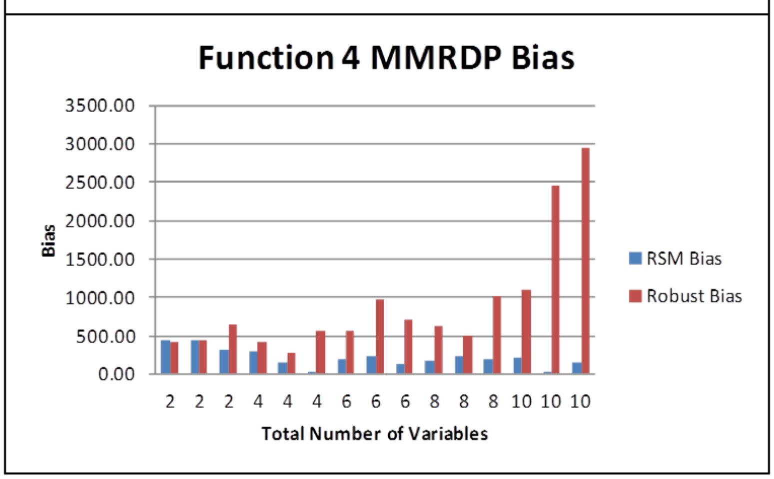
C.4 Schwefel's Function in N Dimensions with Stochastic Behavior

Func #	Exo Vars	Control Vars	Total Vars	Reps per DP	1-α Conf Lvl	Error Type	Known Opt Point	RSM Opt Point	RSM Var	RSM Half Width	RSM HW Covers Known?	RSM Bias	RSM MSE	Robust Opt Point	Robust Var	Robust Half Width	Robust Hw Covers Known?	Robust Bias	Robust MSE
4	1	1	2	30	0.95	Gaussian	-418.98	10.80	1.81	3.25	NO	429.78	184715.16	-37.45	1.79	3.23	NO	410.47	168483.76
4	1	1	2	30	0.95	Non-Gaussian	-418.98	18.43	3.35	4.41	NO	437.41	191333.39	-39.72	1.79	3.23	NO	431.65	186319.95
4	1	1	2	30	0.95	Non-Gaussian	-418.98	-100.54	2.00	3.41	NO	318.44	101407.88	-24.15	1.32	2.77	NO	646.50	417967.10
4	3	1	4	30	0.95	Non-Gaussian	-418.98	-132.25	2.78	4.02	NO	286.73	82217.96	404.47	1.93	3.35	NO	419.22	175751.22
4	1	3	4	30	0.95	Gaussian	-418.98	-275.28	1.13	2.56	NO	143.70	20651.94	658.88	0.84	2.21	NO	278.49	77558.97
4	3	3	6	30	0.95	Non-Gaussian	-418.98	-429.13	1.54	2.99	NO	10.15	104.57	132.23	1.26	2.70	NO	568.98	323744.88
4	1	5	6	30	0.95	Gaussian	-418.98	-220.64	3.79	4.70	NO	198.34	39343.70	121.53	3.25	4.35	NO	560.65	314328.17
4	3	5	8	30	0.95	Non-Gaussian	-418.98	-178.35	3.30	4.38	NO	240.64	57909.90	605.63	2.36	3.70	NO	972.96	946659.81
4	1	7	8	30	0.95	Gaussian	-418.98	-279.23	1.52	2.97	NO	139.75	19532.39	285.62	1.11	2.55	NO	709.26	503051.22
4	3	7	10	30	0.95	Non-Gaussian	-418.98	-250.23	2.61	3.90	NO	168.75	28478.81	471.76	2.60	3.89	NO	616.30	379829.33
4	0	9	9	30	0.95	None	-418.98	-194.32	1.64	3.09	NO	224.66	50473.27	277.23	1.60	3.05	NO	501.87	251878.86
4	1	9	10	30	0.95	Gaussian	-418.98	-231.36	3.06	4.22	NO	187.63	35206.92	-18.76	1.93	3.35	NO	1004.43	1008884.94
4	3	9	12	30	0.95	Non-Gaussian	-418.98	-210.70	1.35	2.80	NO	208.28	43382.18	636.17	1.08	2.50	NO	1104.69	1220344.56
4	1	11	12	30	0.95	Gaussian	-418.98	-385.38	1.51	2.97	NO	33.61	1130.94	992.74	1.29	2.74	NO	2449.72	6001130.85
4	3	11	14	30	0.95	Non-Gaussian	-418.98	-578.40	1.48	2.94	NO	159.42	25416.25	1951.49	0.94	2.34	NO	2955.91	8737378.04

Fig. C.4: The results of MMRDP using Schwefel's function show the tradeoff between bias and variance when arriving ar the robust solution.


Function 4 MMRDP MSE
10000000.00
9000000.00
8000000.00
7000000.00
6000000.00
5000000.00
4000000.00
3000000.00
2000000.00
1000000.00
0.00
MSE
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10
Total Number of Variables
RSM MSE
Robust MSE


Function 4 MMRDP Variance
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Variace
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10
Total Number of Variables
RSM Var
Robust Var


Function 4 MMRDP Bias
3500.00
3000.00
2500.00
2000.00
1500.00
1000.00
500.00
0.00
Bias
2 2 2 4 4 4 6 6 6 8 8 8 10 10 10
Total Number of Variables
RSM Bias
Robust Bias

D. INFRASTRUCTURE IMPACT ASSESSMENT

Compiled by: Derrick Hui, Tom Warner, Christopher Marsh, Lawrence Bowling, and Caleb Erikson, Department of Systems & Information Engineering, University of Virginia. [95]

D.1 Infrastructure Impact Analysis: Transportation

"The direct effect of transport infrastructure investment is to improve travel conditions for its users. Users' behavior will thus change, with wider impacts on the network. There could be further impacts, including accessibility, level and location of employment and increased efficiency, that will contribute to the regeneration of a region. The externalities generated by the investment in transport infrastructure also need to be recognized. [149] "

Direct User Benefits:

- Decreased travel time
- Regarded as the largest direct economic benefit of transportation infrastructure
- Decreased vehicle operating costs
- Increased vehicle safety

Wider Effects:

- Improves reliability and quality of transport services
- Induces trade

Socioeconomic Effects:

Accessibility

Cause:

Reduced travel time, Increased potential to travel

Metric:

Quantity of economic or social activity reachable by transport system

Effect:

Increase in the market size for manufacturing and/or labor, Increased competition and/or centralization

Employment

Cause:

Jobs created for construction, operation, and maintenance of the transportation system

Metric:

Direct employment for system, Indirect employment on complimentary projects

Efficiency

Cause:

Time and cost savings for industry or business

Improved reliability

Metric:

Profitability of affected business

Effect:

Productivity gains due to increased production and distribution, Economic growth

Afghanistan Impact

> "Now my children can walk safely and easily to school. The cars move faster and the drive is smoother. Now it's much easier for me to take my fruit and vegetables to the market. This paved road is very good"

- Sultan Mohammad. [150]

Previous Afghan Effects:

U.S. Agency of International Development

The construction of 1,000 kilometers of rural road:

Increased access to

- Clinics
- Hospitals
- Schools
- Markets

- Greater opportunity for farmers to move their products to market
- Employed over 2,000 Afghan construction personnel

Insurgency Impact:

> "In the northwest of the country, which is not an epicenter of the insurgency, bandits have seized control of the roads and routinely rob travelers. In Badghis province, despite millions of dollars in development, the roads remain unfinished and the Taliban have successfully used opposition to them as propaganda.". [151]

The insurgency have been shown to use roads in the following ways:

- Lay siege to villages
- Capture entire districts
- Attack road construction crews
- Execute complex ambushes against coalition forces

D.2 Infrastructure Impact Analysis: Water

Direct Effects:

- Reliable and easily accessible water supply
- Fewer waterbourne disease
- Metrics: life expectancy, disease rate, cost

Wider effects:

- System must be maintained
- Lower groundwater table and possible contamination at treatment plant site
- Soil erosion and compaction near site

Socioeconomic impact:

Increased economic activity in areas with water supplies

Cause: increased worker productivity and increased trade to region

Effect: increased standard of living

Metric: annual salary, GDP per capita of area, unemployment rate

Jobs created by workers to maintain and operate system

Afghanistan Impact:

17% of rural population have access to improved water source.

34% of urban population have access to improved water source.

Much of Afghanistan is without water sanitation, however the Afghani's must use the water source for it to be an effective source of preventing disease and stimulating the economic growth. Many Afghani's may not be able to pay for the use of water to their private homes. The government may need to consider subsidizing the installments of water spouts in populated areas. [152–154]

D.3 Infrastructure Impact Analysis: Power

Direct Effects:

- Provide electricity
- Improved irrigation systems
- Better education
- Computers, lighting, etc
- Advances for medical centers
- Refrigeration for drugs, sophisticated medical technology

Wider Effects:

- Reinforces Afghan governance capability
- Sense of ownership by local population
- Encourages capital flow with neighboring countries (Uzbekistan, Tajikistan, Turkmenistan)
- Export or import of power, power purchase agreements

Socioeconomic Effects:

- Improved Productivity
- Increased with ability to work at night
- Increased crop yields due to larger acreage
- Increased economic output with reliable power supply
- Employment
- Civil work
- Mechanical work
- Electrical work
- Change in quality of life
- Better care at medical centers
- More advanced care
- Better access to food

Adverse health effects:

- Introduction of certain types of power plants (e.g. coal) produce air-borne pollution
- Privatization

- Opportunity for entrepreneurs

Power in Afghanistan

> "Afghanistan, due to the terrain and widely scattered nature of the rural population, presents huge challenges to standard grid based electrification outside of the major cities. By emphasizing distributed, local power generation Afghanistan can potentially provide a model of power supply development in which distributed power generation on the periphery rapidly meets the immediate power needs of the population before the full grid expands from the country center (Kabul) out.". [155]

The Afghan National Development Strategy deems power the top priority for the country. [156]

D.4 Infrastructure Impact Analysis: Sanitation

> "Evidence suggests that exogenous problems can have a substantial impact on human health. Unsafe water supply, sanitation, and hygiene are responsible for 3% of all deaths and 4.4% of all years of life lost (YLL). But the poorest developing countries are the worst affected; 9% of these deaths occur in nonOECD countries and 90% of those dying are children." [157]

Direct User Benefits:

- Removal of waste water
- Reducing risk of floods
- Diverting excess runoff from precipitation or adverse weather
- Purifying water for reintroduction into system
- Removal of standing water
- Reduced risk of groundwater contamination

Wider Effects:

- System maintenance
- Decreased pollution of water sources

- Improves level of governance by introducing increased government oversight (regulations and stipulations)

Socioeconomic Effects:

Employment

Cause: Construction and upkeep of waste treatment plants, necessary government oversight

Effect: Increase in number employed

Metric: Employment rate

Increased Quality of Life

Cause: Effective transport of waste away from populated areas and clean water sources, provision of latrines

Effect: Reduced occurrence of waste-borne illnesses

Metric: Number of people that have contracted a waste-borne pathogen per 1000 people, occurrence of disease-causing microorganisms in water supply

Benefit to Agriculture

Cause: Treatment of animal waste

Effect: Decomposition of waste to fertilizer

Metric: Tons of fertilizer produced domestically

Risk of contamination of current water sources

Cause: Effective transport and treatment of waste

Effect: Reduced risk of waste-water contaminating wells and aquifers

Metric: Particulate matter in well water (parts per million)

Previous Afghan Effects: UNICEF launched a relief program in Afghanistan in 2000 with financial assistance from USAID and the government of Denmark. To prevent epidemics of cholera and diarrhea:

- Constructed and rehabilitated wells and latrines
- Installed handpumps
- Trucked water to villages

By the end of 2000, UNICEF had provided water and sanitation services to 300,000 people in over 500 villages.

Insurgency Impact:

Improvements to sanitation by the Afghani government and by foreign aid are at risk of being compromised, as the Taliban and other insurgent forces are concurrently providing them. This can serve to delegitimize the Afghan government, and allows the insurgency to curtail services to vulnerable groups. [153]

D.5 Infrastructure Impact Analysis: Social Services

Agriculture:

Direct User Benefits:

Increased, consistent availability of food

Wider Effects:

- More taxable goods
- Potential for overfarming and depriving soil of nutrients.
- Irrigation with unclean water poses health risks

Socioeconomic Effects:

Increased economic activity due to inter-country trade

Cause: Improved farming techniques, larger diversity of crops

Effect: Increased crop output

Metric: Increase in agricultural domestic product

Education

Direct User Benefits:

- Improved literacy rate
- Greater chance at landing a high-paying job

- Good education increases employability, either domestically or with a foreign company

Wider Effects:

- Larger job market (both for applicants and type of jobs)
- Higher education can lead to growth in white-collar service sector
- More people qualified for technical jobs
- Proxy for health education

Socioeconomic Effects:

- Increased intellectual capital
- Improved health

Direct User Benefits:

- Provision of emergency care
- Disease treatment and prevention of epidemics

Wider Effects:

Health education

Socioeconomic Effects:

Improved quality of life

Cause: Increased access to health practitioners, improved awareness of diseases

Effect: Lower incidence of disease

Metric: Infant mortality rate, immunization rate, life expectancy

Limitations:Inability to provide consistent emergency care at locations other than hospitals

Clinics and offices may not be able to perform the needed level of care.

Range of health system coverage limited by transportation of a particular region. [153]

E. R STATISTICAL PACKAGE CODE

E.1 Basic RSM Script with Robust Regression Features

f

```
#RSM_basic_Dual_Response_testfunctions_#v_loop_centered_mean.R
#25FEb13
#TeagueEB

#basic RSM with optimization for Dual Response
# calculates random start points and allows for function selection
#looks at data as a data frame and uses 4 models for fit
# it also includes and automatic DOE generator
#1st order
#1st order with 2 way interactions
#2d order
#2d order with 2 way interactions

#looks at values
#Tests for normalcy in responses and residuals
#determine if robust modeling is necessary
#robust regression (assume nongaussian behavior)
  #Perform better than OLS when the OLS assumptions are not satisfied.
#uses cost constraint
#max cost is 13775
#uses factor constraints
#uses f test (anova with main effects model vs interaction or
#2d order model) for curve test
#test for curve using anova (f test)


# includes simulation
# 1) Select a starting point
# 2) Explore I/0 behavior in the neighborhood of SP
  #use a 1srt order polynomial (Intercept and main effects only)
  #Res IV design
#3) Choose steepest descent/ascent path based off of the values for Beta
```

```
#4) Run tests on path until path does not improve
#5) test for curvature
  # if curvature exists, step 6 else step 1
  # consider 3d order as nessary to avoid assessing an inflection
#point as a min or max
#6)In neighborhood of optimum desgin run and fit LSE of 2d order design
  #Use star design


#Use function from excel sheet for the simulation and as the unk


##############
#library calls
library(MASS) # for glm and confint and tests for normality
library (boot)
library("arm")
library(lme4)
library(stats) # for anova (f test) and shapiro test
library(car) # for QQ & residual plots & curvature test (residCurvTest)
library (survey)
library(aod) # for the wald test
library(micEcon) # for RSquared fcn
library(ppls) # for normalize.vector
library(alr3) # for residual plots and curvature test
library(robustbase) # for M estimation lmrob regression
library(memisc) # for cases function
library(quantmod) # for getSymbol
library(FrF2) # for DOE matrices
library(lmtest) # for likelihood ratio test of nested models
library(rsm) # used for psa PSI
library(combinat) # for combinations and permutation
library(earth)#for mars
library(scatterplot3d) # for 3d plots
library(gplots) # for filled contour plots
```

```
library(akima) # for interpolation


#clear screen cntl  l

#point to dir

setwd("C:/Users/Administrator/Documents/UVA JUL10-JUN13/courses/Dissertation/Defense/R Code/Test Fu
#setwd("~/UVA JUL10-JUN13/courses/Dissertation/Defense/R Code/Test Functions")

work_dir <- "C:/Users/Administrator/Documents/UVA JUL10-JUN13/courses/Dissertation/Defense/R Code/T

func_dir <- paste(work_dir, "/R_Call_Functions/", sep = "")

optim_work_dir <- paste(work_dir, "/R_Call_OPTIM_FILES/", sep = "")

####################################################################################################
#source Functions
####################################################################################################

######## reads in levels for factors
source(paste(func_dir, "CCD_FRAME.r", sep = ""))

#########################F test to check for curvature
source(paste(func_dir, "CURVE_CHECK_f.r", sep = ""))

####reads in the DOE frame from a premade text file
source(paste(func_dir, "DOE_FRAME.r", sep = ""))

######## reads in levels for factors
source(paste(func_dir, "DOE_LEVELS_f.r", sep = ""))

######applies approprate step sizes to each column in the DOE
source(paste(func_dir, "DOE_STEP_f.r", sep = ""))
```

```
##################Check a point to see if it's within limits an then correct it
source(paste(func_dir, "FACTOR_CHECKPOINT.r", sep = ""))

##################Check a point to see if it's within limits an then correct it
source(paste(func_dir, "FACTOR_CHECKDF_f.r", sep = ""))

##################Create Replicates of the function ( this is done instead of a simulation)
source(paste(func_dir, "GEN_REPLICATES_f.r", sep = ""))

#####applies start points to DOE Frame
source(paste(func_dir, "NEW_START.r", sep = ""))

#########test residuals for normality and plots them
source(paste(func_dir, "NORM_TEST.r", sep = ""))

#########################create a matrix where the columns are normalized
source(paste(func_dir, "NORMMATRIX.R", sep = ""))

#########################create a matrix where the columns are normalized based on another matrix
source(paste(func_dir, "NORMMATRIX_f.R", sep = ""))

#######undoes normalization of experimental values
source(paste(func_dir, "UN_NORMAL.r", sep = ""))

#######undoes normalization of experimental values that have been centered
source(paste(func_dir, "UN_NORMAL_and_center.r", sep = ""))

####### creates normalization of experimental values after centering them using the same matrix as
source(paste(func_dir, "NORMMATRIX_and_center.r", sep = ""))

####### creates normalization of experimental values after centering them using another matrix as a
source(paste(func_dir, "NORMMATRIX_f_and_center.r", sep = ""))

#######Generates an expression based on selection and produces the expression characteristics
source(paste(func_dir, "GEN_EXPRESSION.r", sep = ""))
```

```
#######Generates an expression in n dimensions and reports to GEN_EXPRESSION.r
source(paste(func_dir, "EXPRESSION_DIM.r", sep = ""))

#######Generates a value as a part of a point
source(paste(func_dir, "GEN_START_POINT.r", sep = ""))

#######Generates a start point in n space point
source(paste(func_dir, "GEN_START_VECT.r", sep = ""))

#######Generates the CCD piece of the DOE
source(paste(func_dir, "GEN_CCD_f.r", sep = ""))

#######Generates the main piece of the DOE
source(paste(func_dir, "GEN_DOE_f.r", sep = ""))

#######Generates values for environmental factors
source(paste(func_dir, "GEN_ENV_f.r", sep = ""))

#######Generates information about stationary points
source(paste(func_dir, "EIGENVALS_FROM_B.r", sep = ""))

#######writes expression into a gms file for optimization
source(paste(func_dir, "GEN_GAMS_SCRIPTv4_generic.r", sep = ""))

#######determines Hyper-Ellisoid surface area via estimate
source(paste(func_dir, "H_E_APPROX_f.r", sep = ""))

############################## for H_E_APPROX_f.r######
source(paste(func_dir, "D_FACT_f.r", sep = ""))



#******************************************************************************
#begin scripting
```

```
##**************************************************************************

#Step 1
#Function,Dimension,Replication, & Min or Max & Full Factorial Main Effects DOE with CCD Generation
##########################################################################################################
##########################################################################################################

#select a function from a list of functions
func_number = 1  # there are 7
# 1 - Dejong's first function in n dimensions
# 2 - Rosenbrock's valley in n dimensions
# 3 - Ratringin's function in n dimensions
# 4 - Schwefel's function in n dimension
# 5 - parabola with a min of zero
# 6 - parabola with a max of zero
# 7 - linear function


#select the corrsct info file on the variables for the function
char_info = paste("varchar_", func_number, ".txt", sep = "")

var_source  <- char_info #gives info on all variables and steps for functions


tot_var <- 4 #total number of variables (dimensions)
tot_evars <- 3 # number of environmental variables
#tot_cvars <- # of control variables

#number of replications per design point
reps <- 30 #must be > 2

#Initialize loop counter
PSI_loop <- 1

# max number of steps taken
```

```
p_max = 50 #This can be adjusted should it be dynamic?

#create a list with info on the environmental variables
#1 is norm, 2 is exp, 3 is uniform
e_chars<- matrix(nrow = 3, ncol = tot_evars, byrow = FALSE)
e_chars[,1] <- c(1, 0, 2) #c(2, 50, 1) # #c(3, -2, 2) #
# normal random variable (type, mean, sd) mean of 0
e_chars[,2] <- c(2, 50, 1) #c(3, -2, 2) #
# exp random variable (type, rate, ignore)small mean
e_chars[,3] <- c(3, -2, 2)
# uniform random variable (type, min, max)mean of zero

#create a vector of the means for all RV
e_mean <- seq(0, length.out = tot_evars)
for (c in 1:tot_evars) {
  if (e_chars [1,c] != 3)
  {e_mean[c] <- e_chars [2,c] } else
  {e_mean[c] <- (e_chars [3,c]-e_chars [2,c])/2}
}

e_sd <- seq(0, length.out = tot_evars)
for (c in 1:tot_evars) {
  if (e_chars [1,c] != 3)
  {e_sd[c] <- e_chars [1,c] } else
  {e_mean[c] <- (e_chars [2,c]-e_chars [1,c])/2}
}

#create initial values for the evironmental variables
env_vals <- GEN_ENV_f.r(tot_var, tot_evars, e_chars)

# 1 - Dejong's first function in n dimensions
# 2 - Rosenbrock's valley in n dimensions
# 3 - Ratringin's function in n dimensions
# 4 - Schwefel's function in n dimension
```

```
#problem type minimze = true is minimize
minimize = TRUE
# could just flip response rather than a bunch of if/than statements

#RSM Search improving boolean holder variable
improving = TRUE

#linearity boolean
boo_linear = TRUE

#level of conf
alpha <- .001

#how many iterations of experimentation have you done
#an iteration is either running a DOE with a model
#output or creating a route the getting those values
loop_count <- 1

#create a full factorial main effects DOE of appropriate size
basic.DOE <- GEN_DOE_f.r(tot_var)

#modify CCD for with alpha
a<- 2^(1/2)

#load factor min, max, and steps
factor.levels <- DOE_LEVELS_f.r(var_source)
#varsource may have holders beyond the current number of dimensions
factor.levels <- factor.levels[,1:tot_var]
# get factor.levels to the right number of columns

#generate a first order full factorial design CCD
CCD.DOE <- GEN_CCD_f.r(tot_var)

#col names in CCD.DOE = col names in basic.DOE
```

```
dimnames(CCD.DOE)[[2]] <- dimnames(basic.DOE)[[2]]

#combine CCD and DOE for 2d order models
basic.DOE <- rbind(basic.DOE, a*CCD.DOE)
# ignore the warning (not meaningful)
# just use this DOE from now on its simpler

#modify DOE step to reflect step for each factor
stepped.DOE <- DOE_STEP_f.r(var_source, basic.DOE)
# produces frame that is stepped correctly

#generate a start point
start_point <- GEN_START_VECT.r(var_source, tot_var, func_dir)

#updated design matrix
updated_design <- NEW_START.r(start_point, stepped.DOE)
#use only tot_var cols

#clean the experimental matrix to avoid NA and -inf
updated_design[is.na(updated_design)] <- 0
#updated_design[is.infinite(as.data.frame(updated_design))] <- 0

#add center point
updated_design <- rbind(updated_design, start_point)

#check that the factors are within bounds
updated_design <- FACTOR_CHECKDF_f.r(factor.levels, updated_design, tot_evars)

#initiate responses vector for simulation
responses <- matrix(data = NA, nrow = ((2*tot_var) +
(2^tot_var)+1), ncol = 1, byrow = FALSE)

#generate a set of env_var and replications per DP
checky = 0
# checky is a counter that helps run the loop if reps only = 1
```

```
for(checky in 1:reps) {
# begin create replications
#install the random variables for the e vars in the updated design
env_vals <- NULL
env_vals <- GEN_ENV_f.r(tot_var, tot_evars, e_chars)
#put in the random variables for e_vars
#generates 2^tot_var + 2* tot_var + 1
#rows per column for full fact doe, ccd, and start point
updated_design[,1:tot_evars] <- env_vals
# take the values for the env var
if(checky > 1){responses <-
cbind(responses, GEN_REPLICATES_f.r(as.matrix(updated_design), 1, func_dir, func_number))}
else { responses <- GEN_REPLICATES_f.r(as.matrix(updated_design), 1, func_dir, func_number)}
checky = checky+1} # end create replications of the responses

#clean the responses to avoid NA and INF
responses[is.na(responses)] <- 0
#responses[is.infinite(responses)] <- 0

#initialize the history of best values
Best_V_Track <- cbind(updated_design, responses[,1])
#intialize the tracker and add a columns for the response
Best_V_Track[Best_V_Track < 1000000000] <- 25000
# intialize the tracker value to 25000 for all spots in dataframe

#fix col names  this is a way to change names
hold <- colnames(Best_V_Track)
hold[(tot_var+1)] <- "Response"
colnames(Best_V_Track) <- hold

# get intial point on the tracker
B_V_Index <- 0 # intialize index in track with best value
Best_V_Track[1,1:tot_var] <- (start_point)
Best_V_Track[1,(tot_var+1)]<-
GEN_SIM_f.r(Best_V_Track[1,1:tot_var],1, func_dir, func_number) # intial response
```

```
B_V_Index <- as.numeric(dim(Best_V_Track)[1])
#current row of best value and best set of factors

#make a column of the best responses by row and generate ave response per row
#what to do if non normal response, do you weight it?
  top_responses <- matrix(responses[,1], ncol = 1)
#initialze the column as an nx1
  top_ave_resp <- top_responses
# use to choose best using top average value rather than top value
  resp_var <- top_ave_resp
#holder for the variance of each set of responses
  for(row_cnt in 1:dim(responses)[1]){#loop on rows
    if(minimize == TRUE){ #begin look for min
      top_responses[row_cnt,1] <- min(responses[row_cnt,])
      top_ave_resp[row_cnt,1] <- ave(responses[row_cnt,])[1]
      resp_var[row_cnt,1] <- var(responses[row_cnt,])[1]
      } else {#begin look for max
        top_responses[row_cnt,1] <- max(responses[row_cnt,])
        top_ave_resp[row_cnt,1] <- ave(responses[row_cnt,])[1]
        resp_var[row_cnt,1] <- var(responses[row_cnt,])[1]}}
#end else loop on rows


#find first true values for the tracker depending on if its a min or max problem
if (minimize == TRUE)
{Best <- which(top_ave_resp == min(top_ave_resp), arr.ind = TRUE)} else
  #gets address of best response row for a min
{Best <- which(top_ave_resp == max(top_ave_resp), arr.ind = TRUE)}
#gets address of best response row

#use Best to locate and select the first true value and the corresponding n space location
Best_V_Track[2,1:tot_var] <- (updated_design[Best[1],])
Best_V_Track[2,tot_var + 1]<- (min(responses[Best[1],1:reps]))
Best_V_Track<- Best_V_Track[1:2,]#shrink B_V_Track to 2 row to start
B_V_Index <- as.numeric(dim(Best_V_Track)[1])
```

```
#update current row of best value and best set of factors

#update the matrix optimal value and its address
 Best_val <-Best_V_Track[B_V_Index,tot_var+1]
 Best_factors <-  Best_V_Track[B_V_Index,1:tot_var]
# returns values of factors that gen the current best value



#initialize the overall surface archive
#surf_archive is all of the points visited and
#the reponses generated by DOE and the path of improvement
#columns tot_var+1 to tot_var+reps are the reponses

 surf_archive <- cbind(updated_design, responses, top_responses,
                        top_ave_resp, resp_var )
# in the future do this then rbind to surf_archive

#use best point from Best_V_Track as a new start point to begin the PSI
 next_point <- Best_factors



#Step 2
#Testing for Stochasitc then Gaussian Random Behavior in response then Conduct Regression (within v
###########################################################################################################
###########################################################################################################
  while (improving == TRUE) { #begin first order model looping for steps 2,3,4,5,6

    print(paste("step 2, loop- ", PSI_loop))

NORM_tru = NULL # initialize marker for gaussian behavior of the response



#test for guassian behavior
#replicate some responses from the same set
    #(best value b/c that is the next starting point)
```

```
#generate a set of replications per DP
  test_dummy <- seq(0, length.out = 100) # response holder

  #create a matrix to hold the repeated start point control variables
  control_holder <- NULL
  control_holder <- matrix(data = as.matrix(next_point[(tot_evars+1)
  :tot_var]), nrow = 100, ncol = (tot_var-tot_evars), byrow = TRUE)

  #create instances of the env variables
  env_vals <- NULL #initialize
  env_vals <- GEN_ENV_f.r(10, tot_evars, e_chars)#put in the random variables for e_vars
  evar_test <- env_vals[1:100,] # take 100 rows of the env vars

  #bind it all togther then replicate
  evar_test <- cbind(evar_test, control_holder) # take 100 rows of the env vars
  test_dummy <- GEN_REPLICATES_f.r(as.matrix(evar_test), 1, func_dir, func_number)
   # generate 100 replicates

 #test for stochastic behavior
   if (var(as.numeric(responses[dim(responses)[1],]) > 0 ))
   {print("Stochastic Response Behavior is Detected")}

  NORM_tru <- NORM_TEST.r(responses[dim(responses)[1],],alpha)
   # if true - gaussian, if false - nongaussian

#Create a normalized vector of factors using function NORM matrix on the cols
factors_norm <- NORMMATRIX_and_center.r(updated_design)
factors_norm_reps <- factors_norm
   # initialize to help with regression  this turn the replications into long columns
responses_reps <- as.matrix(responses[,1])
   # intialize a new response column on the 1st col of responses

response_var_holder <-  responses_reps #initialize a holder for the variance
response_var_holder   <- as.matrix(ncol = 1, apply(responses, 1,var))
   # assign each row the variance value of the replication at a point
```

```
#create a large set of factors and 1 response to enable regression for allreplications
for (index_reps in 2:reps) { # begin loop
  factors_norm_reps <- rbind(factors_norm, factors_norm_reps)
  #length the normalized factor rows by 1 replication
  responses_reps <- rbind(responses_reps,as.matrix(responses[,index_reps]))
  # lengthen the response by 1 replication
} # end loop

#change matrix to dataframe
factors_norm_reps_DF <- data.frame(factors_norm_reps)
#ignore the warning


if (NORM_tru == TRUE) { # using OLS
#OLS linear regression only model
#main effects model
print("Using OLS Regression")

model.1 <- lm(formula = responses_reps ~ ., data = factors_norm_reps_DF)
#uses normalized factor values # first order model
#model with just interactions
model.1int <- lm(formula = responses_reps ~ .^2,  data = factors_norm_reps_DF)
#uses normalized factor values # first order model with 2 way interactions
#model with 2d order relationship
model.2 <- lm(formula = responses_reps ~  I(factors_norm_reps_DF^2),
              data = factors_norm_reps_DF) #2d order model
#model with interactions and squared values
model.2int<- lm(formula = responses_reps ~ .^2  + I(factors_norm_reps_DF^2),
                data = factors_norm_reps_DF)
#uses normalized factor values # 2d order model with 2 way interactions
} else { # use WLS for robust regressions #lmrob not working, using lmrob instead
print("Using Robust Regression Techniques")
#main effects model
model.1 <- lmrob(formula = responses_reps ~ .,
```

```
                 data = factors_norm_reps_DF)
#uses normalized factor values # first order model
#model with just interactions
model.1int <- lmrob(formula = responses_reps ~ .^2,
                    data = factors_norm_reps_DF)
#uses normalized factor values # first order model with 2 way interactions
#model with 2d order relationship
model.2 <- lmrob(formula = responses_reps ~  I(factors_norm_reps_DF^2),
                 data = factors_norm_reps_DF) #2d order model
#model with interactions and squared values
model.2int<- lmrob(formula = responses_reps ~ .^2  + I(factors_norm_reps_DF^2),
                   data = factors_norm_reps_DF)
#uses normalized factor values # 2d order model with 2 way interactions
}# end else for types of regresssion (OLS vs WLS)
  #} #end while loop to ensure regressions are completed

#Create a list (reg_model)for 1, 1int, 2, 2int
   #it will hold models for OLS or WLS depending on norm_tru
reg_model <- list(one_ = model.1, one_int = model.1int,
                  two_ =  model.2, two_int = model.2int)

#to get var cov matrix
#vcov(lm.object)


#Step 3
#Statistical Model Selection  (within while loop)
##################################################################################################
##################################################################################################
   print(paste("step 3, loop- ", PSI_loop))


boo_linear = NULL #intialize

#test for better fit ME vs ME and interactions model
```

```
boo_linear <- CURVE_CHECK_f.r(reg_model$one_,
                              reg_model$two_int, alpha, NORM_tru)
#  linear versus 2d order with interactions
#true means that there is little differnce between the models
   #so use the simpler one, IE the main effects model
#if false -  fit 2d order model and optimize
#if true - stick with linear model and find a path of steepest
# ascent/descent and continue

#force a linear loop #boo_linear == TRUE

if (boo_linear == FALSE) { # begin IF - linear model is inappropriate
  #use reg_model$two_int for prediction
  print("test confirms a better fit from the 2d order model")
  print("procede to optimization")
  improving  = FALSE
  #kicks out of the loop  Do you have to put a break here?
  break # stops while improving loop

}  else{ #begin if linear model IS appropriate (boo_linear is true)
  #use reg_model$one for prediction



#Step 4
#Path of Steepest improvement (PSI) (within while loop)
##########################################################################################################################
##########################################################################################################################
print(paste("step 4, loop- ", PSI_loop))


#reset normalized route
Route <- NULL
#reset responses
responses = NULL
est_responses = NULL
```

```
#normalize the vector after you drop the intercept
PSI <- NULL
PSI <- reg_model$one_int$coefficients
PSI_ <- PSI[-1] # take away the first column (the intercept)

#modify the PSI so the steps are only in the direction of the control variables
#includes zero out steps in the env dir
PSI_[1:tot_evars]<- seq(0,0, length = tot_evars)
#vector of 0s of length tot_envars

############################
#Normalize the PSI
norm_PSI <- normalize.vector(PSI_[1:tot_var])
#this is your step size along and you only do it for the main effects (tot_var)

#your should procede with larger one first from PSI
step_PSI <- norm_PSI/max(abs(norm_PSI))
# this is your adjusted step size  it's normalized so take the largest step in the dir of greatest


#determine direction of gradient from step_PSI using current best point and step_PSI
#you don't need to use the ave_responses here
Best = NULL
if (minimize == TRUE)
{Best <- which(responses_reps == min(responses_reps), arr.ind = TRUE)} else
  #gets address of best response row for a min
{Best <- which(responses_reps == max(responses_reps), arr.ind = TRUE)}
#gets address of best response row
n_pts <- predict(reg_model$one_, as.data.frame(rbind(factors_norm_reps_DF[Best[1],],
                 factors_norm_reps_DF[Best[1],]+step_PSI))) #find first 2 points
change <- n_pts[2] - n_pts[1] #this is the difference between the start and next step


#ensure that you step in the right direction if its a min or max problem
```

```
if (minimize == TRUE) {#min problem so go in a negative dir
  if (change > 0) {step_PSI = -step_PSI}} #end if

if (minimize == FALSE) {#max problem so go in a positive dir
    if (change < 0) {step_PSI = -step_PSI}}  #end if



#Create the path to the new spot
Route <- factors_norm_reps_DF[1:p_max,]
# use regression basis data frame with correct number of rows
Route[1,] <- factors_norm_reps_DF[Best[1],]
#normalized start point is the best point thus far
dummy_holder <- Route[1,] # for the loop to build the route

# the points along the step will be for estimates
#of the response theses are normalized
for(p_count in 1:p_max-1)  {#begin path creation loop
     Route[p_count+1,] <- dummy_holder + (p_count*step_PSI)
     } # end path creation loop

#decode the normalized route back to input regular
#value levels and check bounds using updated_design
  Route_regular <- Route # set size and shape
  Route_regular[Route_regular > -10000000000] <- NA
# reset values

#use updated_design to recode route into route_regular
  Route_regular <- UN_NORMAL_and_center.r(as.matrix(Route),
                       as.matrix(updated_design), tot_var)
  Route_regular <- FACTOR_CHECKDF_f.r(factor.levels,
                       Route_regular, tot_evars)
#check values for feasibility and corrects them
  Route_regular <- as.data.frame(Route_regular)

#Generate observed  and estimated responses for the path (route)
```

```
  rt_cnt = 0 #  rt_cnt is a counter that helps run the loop if reps only = 1
  for( rt_cnt in 1:reps) {# begin create replications loop
    #install the random variables for the e vars in the updated design
    env_vals <- NULL
    env_vals <- GEN_ENV_f.r(tot_var, tot_evars, e_chars)
    #put in the random variables for e_vars

    #CORRECT THE LENGTH OF THE ENV-VALS ROWS
    Route_regular[,1:tot_evars] <- env_vals [1:dim(Route_regular)[1],]
    # take the values for the env var

    #normalize the Regular route using last updated_design values
    Route <- NORMMATRIX_f_and_center.r(Route_regular, as.matrix(updated_design))
    Route <- rbind(factors_norm_reps_DF,as.matrix(Route))
    #use this form of data to make predict happy below
    Route <- Route[(dim(factors_norm_reps_DF)[1]+1):
                    (dim(factors_norm_reps_DF)[1]+p_max),]
    #route is now correct rows and format IOT predict

    #create an observation and an estimate
    if( rt_cnt > 1){responses <- cbind(responses,
    GEN_REPLICATES_f.r(as.matrix(Route_regular), 1, func_dir,
                      func_number)) #observation
    est_responses <- cbind(est_responses, predict(reg_model$one_,
                                                  Route))    #estimation
                  } else
    {responses <- GEN_REPLICATES_f.r(as.matrix(Route_regular),
                                     1, func_dir, func_number)
     # the else is to build the first column
    est_responses <- predict(reg_model$one_, Route) }  #estimation
              rt_cnt =  rt_cnt+1}
# end create replications of the responses and estimated resposes loop

  #make a column of the best responses by row  and a column of average
#for observations and estimated responses
```

```
  #if not normal do i use a weighted average?  not sure
  top_responses <- matrix(responses[,1], ncol = 1)
#initialze the column as an nx1
  top_e_responses <- matrix(est_responses[,1], ncol = 1) #same
  top_ave_resp <- top_responses #same
  top_e_ave_resp <- top_e_responses #same
  resp_e_var<- top_e_ave_resp #same
  resp_var<- resp_e_var #same

  for(row_cnt in 1:dim(responses)[1]){#loop on rows
    if(minimize == TRUE){ #begin look for min
      top_responses[row_cnt,1] <- min(responses[row_cnt,])
      top_e_responses[row_cnt,1] <- min(est_responses[row_cnt,])
      top_ave_resp[row_cnt,1] <- ave(responses[row_cnt,])[1]
      #average observation values
      top_e_ave_resp[row_cnt,1] <- ave(est_responses[row_cnt,])[1]
      #average estimate values
      resp_var[row_cnt,1] <- var(responses[row_cnt,])[1]
      #find variance of the response
    }else {top_responses[row_cnt,1] <- max(responses[row_cnt,])
          #begin look for max
      top_e_responses[row_cnt,1] <- max(est_responses[row_cnt,])
      top_ave_resp[row_cnt,1] <- ave(responses[row_cnt,])[1]
          #average observation values
      top_e_ave_resp[row_cnt,1] <- ave(est_responses[row_cnt,])[1]
          #average estimate values
      resp_e_var[row_cnt,1] <- var(est_responses[row_cnt,])[1]}}
#find the variance of the response
   #end else loop on rows


  #Connect the replications with the route in regular coordinates
  Route_regular <- cbind(Route_regular, responses, top_responses,
        top_ave_resp, resp_var) # observations
  Route <- cbind(Route, est_responses, top_e_responses,
```

```
                top_e_ave_resp, resp_e_var )#estimations

# more analysis between the two sets of outcomes can be done here
# such as mean and variance comparisons

#update the surface archive using observations only
  surf_archive <- rbind(surf_archive, Route_regular)

#show a true plot of the path of improvement
  plot(c(1:dim(Route_regular)[1]), Route_regular$top_ave_responses,
       pch="*", xlab='Path of Improvement Steps', ylab='Response Value',
       col = 20, main = " Basic RSM with Univariate Response : Path From Last Best Point")
  lines(Route$top_e_responses, pch=19, col = "red")
  #sd_e = sd(est_responses)/dim(Route)[1]
# calculations of standard error of the estimate  # could do this per observation row as well and n
  #sd_e = by row to do
  #sd_e = sd(responses)/dim(Route_regular)[1]  # calculations of standard error of the observation
  #Route_error<-seq(sd_e,dim(Route)[1]*sd_e,by=sd_e) # ignore the warning
  #lines(Route$top_e_responses + sqrt(Route_error)) #plot lower and upper standard error of the est
  #lines(Route$top_e_responses - sqrt(Route_error))



#Step 5
#Next Point Selection/First Order Model Stopping Criteria (within while loop)
##############################################################################################################################################################
##############################################################################################################################################################
 print(paste("step 5, loop- ", PSI_loop))
#use an if to differnentiate between a minimzation or a maximization problem

 if (minimize == TRUE) { # if it is a minimzation problem

    if (min(Route_regular$top_ave_resp) >=  Best_val) {
      #begin its a min and not improving
      improving = FALSE# a way to get out of the
      #1st order business go back and do a 2d order from last min
```

```
    print(paste("After", PSI_loop, "paths,
                first order model yields no improvement on current min"))
    print("Generating 2d order model")
    #script runs a DOE with CCD around the best point
    break
    # ends the while improving loop and moves on to 2d order model

  }else #its a min and improving
    {minimum <- which(Route_regular ==
    min(Route_regular$top_ave_resp), arr.ind = TRUE)
     #gets address of lowest average value
    Row_best <- minimum[1]
    next_point <- Route_regular[Row_best, 1:tot_var]
     # returns row of best
    Best_val <- Route_regular$top_ave_resp[Row_best]
     #update new best value

     print(paste("After", PSI_loop,
      "paths, first order model yields improvement"))

    #update the history of best values

      adder <- NULL
      adder <- Route_regular[Row_best,
        c(1:tot_var,tot_var+reps+2)]
      names(adder) <- names(Best_V_Track)
      Best_V_Track <- rbind(Best_V_Track,adder)
    }# end else for max and improving

}else { # begin else its a maximization problem
   #check to see if there is improvement along the path and where is it
   if (max(Route_regular$top_ave_resp) <=
         Best_val) { #if not improving for a max
    improving = FALSE
# a way to get out of the 1st order business go back and do a 2d order from last min
```

```
      print(paste("After", PSI_loop,
"paths, first order model yields no improvementon current max"))
      print("Generating 2d order model")
      #script runs a DOE with CCD around the best point
      break
  # ends the while improving loop and moves on to 2d order model
    }else
      {maximum <- which(Route_regular ==
    max(Route_regular$top_ave_resp), arr.ind = TRUE)
       #gets address of lowest value in col 3
      Row_best <- maximum[1]
      next_point <- Route[Row_best, 1:tot_var]
       # returns row of best
      Best_val <- Route_regular$top_ave_resp[Row_best]
       #update new best value

      #update the history of best values
       Best_V_Track <- rbind(Best_V_Track,
      Route_regular[Row_best,c(1:tot_var,tot_var+reps+2)])
    }# end else for max and improving
  }# end else for maximization problem

    Best_V_Track <- as.data.frame(Best_V_Track)
# coerce into a dataframe

    cat(sprintf("first order model complete\n") )
    cat(sprintf("first order local optimal value is %f\n ",
    Best_V_Track$Response[dim(Best_V_Track)[1]]))

    #print the best point
    # we have to be cuareful here about the  environmental factors
    o_count  <- 0
    for(o_count in 1:(dim(Best_V_Track)[2])) { #begin print loop
    cat(sprintf(" Best average response %s level is %f \n",
  colnames(Best_V_Track)[o_count],
```

```
  Best_V_Track[dim(Best_V_Track)[1], o_count]))
    cat(sprintf("\n"))
    } # end print loop

    } #end else - boo_linear is TRUE (linear model IS appropriate)


#Step 6
#New Experiment and start point for linear modeling (within while loop)
##########################################################################################################
##########################################################################################################
   print(paste("step 6, loop- ", PSI_loop))

   # reset for new values
  updated_design <- NULL

#updated design matrix
  updated_design <- NEW_START.r(next_point,
  stepped.DOE) #use only tot_var cols

#clean the experimental matrix to avoid NA and -inf
  updated_design[is.na(updated_design)] <- 0
  #updated_design[is.infinite(as.data.frame(updated_design))] <- 0

#add center point
  updated_design <- rbind(updated_design, next_point)

#check that the factors are within bounds
  updated_design <-
    FACTOR_CHECKDF_f.r(factor.levels, updated_design, tot_evars)

#clean responses vector for simulation
  responses <- matrix(data = NA, nrow =
  ((2*tot_var)+(2^tot_var)+1), ncol = 1, byrow = FALSE)
   #one column b/c of the upcoming cbind
```

```
#generate a set of env_var and replications per DP
  checky = 0 # checky is a counter that helps run the loop if reps only = 1
  for(checky in 1:reps) {# begin create replications
    #install the random variables for the e vars in the updated design
      env_vals <- NULL
      env_vals <- GEN_ENV_f.r(tot_var, tot_evars, e_chars)
      #put in the random variables for e_vars
    #generates 2^tot_var + 2* tot_var +
      #1 rows per column for full fact doe, ccd, and start point
      updated_design[,1:tot_evars] <- env_vals
      # take the values for the env var
      if(checky > 1){responses <- cbind(responses,
      GEN_REPLICATES_f.r(as.matrix(updated_design), 1,
      func_dir, func_number))}
      else { responses <- GEN_REPLICATES_f.r(as.matrix(updated_design), 1, func_dir, func_number)}
      checky = checky+1}
   # end create replications of the responses


  if (PSI_loop > 1000) { # too many iterations
    improving = FALSE
    # a way to get out of the 1st order business go back and do a 2d order using CCD
    break
      cat(sprintf("terminating on loop effort %i \n ", PSI_loop))
      cat(sprintf("Local Optimal Value Reached is %f \n ", Best_val))
      cat(sprintf(" \n   ") )

    } # too many iterations
# UPdate Counter for number of times PSI is created and followed
   PSI_loop <- PSI_loop+1

  } # end while improving loop
```

```
if(improving==FALSE){print ("no more improvements to be had")
                    } else {print("still improving but bailed")}


#Step 7
#2d order exploration
################################################################################################################
################################################################################################################
#examine first order design vs first orderuse second order model from last regression


##########################
#USE LAST REGRESSION RESULTS
boo_linear = NULL #reintialize
#test for better fit ME vs ME and interactions model
boo_linear <- CURVE_CHECK_f.r(reg_model$one_int,
reg_model$two_int, alpha, NORM_tru)
#true means that there is little differnce between
#the models so use the simpler one, IE the main effects model


#Step 8
#BIAS and VARIANCE COMPARISON
################################################################################################################
################################################################################################################

cat(sprintf("Known Local Optimal Value Reached is %f \n ",
Best_V_Track[dim(Best_V_Track)[1], dim(Best_V_Track)[2]]))
cat(sprintf("Known Local Optimal Point is %f \n ",
Best_V_Track[dim(Best_V_Track)[1], 1:(dim(Best_V_Track)[2]-1)]))
cat(sprintf(" \n   ") )


if (boo_linear == TRUE) # use the simpler model
  {final_model = reg_model$one_int
```

```
    cat(sprintf("Use First Order Interaction Model"))  } else
     { # use the more complex model
     cat(sprintf("Use Second Order Model with Interactions"))
     final_model = reg_model$two_int
     }


  NMSEcrpred <- mean((final_model$residuals)^2)/
  mean((mean(responses_reps)- responses_reps)^2)
  print("NMSE")
  print(NMSEcrpred)

  plot(c(1:dim(Best_V_Track)[1]),
       Best_V_Track$Response, pch="*", xlab='Points',
       ylab='Response Value', col = 20,
       main = "Basic RSM: Improvement by Iteration")
  plot(responses_reps,final_model$residuals,
       pch="*", xlab='True Values', ylab='Residuals',
       col = 20, main = "RSM 2d Order Model Residuals vs Known")

  qqPlot(final_model$residuals,
         main="Basic RSM First Order Interaction Model Residual Q-Q PLot")
## drawing the QQplot

  summary(final_model)

#Variance estimate
est_var = 0
for (c in 1:tot_var){est_var = est_var + as.numeric(final_model$cov[c,c])}
if (length(est_var) == 0) {est_var = 0}
if (est_var == 0){est_var = (summary(final_model)$sigma)^2}

#examine bias
Base_Model= GEN_EXPRESSION.r(func_number, tot_var, func_dir)#get the known model min
BIAS <- Base_Model$opt - Best_val
```

```
print("Variance")
print(var((responses)[1,]))
print("Estimate variance")
print(est_var)
print("Best Response")
print(Best_V_Track$Response[length(Best_V_Track$Response)])
print("Bias")
print(BIAS)
print("NMSE")
print(NMSEcrpred)



#Step 9
#create the second surface (variance)
##########################################################################################################
##########################################################################################################
#check for behavior type and recommend a type of regression
NORM_tru = NULL # initialize marker for gaussian behavior of the response

DOE_length <- dim(factors_norm_reps_DF)[1]/reps #how many reps to make IOT regress the variance

#test for guassian behavior
#replicate some responses from the same set (best value b/c that is the next starting point)
#generate a set of replications per DP
#test_dummy <- seq(0, length.out = DOE_length) # response holder

#create a matrix to hold the repeated start point control variables
control_holder <- NULL
control_holder <- matrix(data = as.matrix(Best_factors[(tot_evars+1):tot_var]),
nrow = DOE_length, ncol = (tot_var-tot_evars), byrow = TRUE)

#create instances of the env variables
env_vals <- NULL #initialize
env_vals <- GEN_ENV_f.r(round(sqrt(DOE_length)),
tot_evars, e_chars)#put in the random variables for e_vars
```

```
evar_test <- env_vals[1:DOE_length,]  # take correct number of rows of the env vars

#bind it all togther then replicate
evar_test <- cbind(evar_test, control_holder)
NORM_tru <- NORM_TEST.r(response_var_holder[dim(responses)[1],],
            alpha) # if true - gaussian, if false - nongaussian

#create a regression based on the variance for the final model
nec_fac_rows <- factors_norm_reps_DF[1:DOE_length,]
#the necessary rows used of the total of factors_norm_reps_DF
#holder for the variance of the model outcome at the above levels

NORM_tru = TRUE #for now


if (NORM_tru == TRUE)  { # using OLS
  #OLS linear regression only model
  #main effects model
  print("Using OLS Regression")



  #model with just interactions
  model.1int <- lm(formula = response_var_holder ~ .^2,
                  data = nec_fac_rows)
  #uses normalized factor values # first order model with 2 way interactions
  #model with interactions and squared values
  model.2int<- lm(formula = response_var_holder ~ .^2
                 + I(nec_fac_rows^2), data = nec_fac_rows)
  #uses normalized factor values # 2d order model with 2 way interactions
} else { # use WLS for robust regressions #lmrob not working, using lmrob instead
  print("Using Robust Regression Techniques")
  #model with just interactions
  model.1int <- lmrob(formula = response_var_holder ~ .^2,
                     data = nec_fac_rows)
```

```
  #uses normalized factor values # first order model with 2 way interactions
  #model with interactions and squared values
  model.2int<- lmrob(formula = response_var_holder ~ .^2
                   + I(nec_fac_rows^2), data = nec_fac_rows)
  #uses normalized factor values # 2d order model with 2 way interactions
}# end else for types of regresssion (OLS vs WLS)
#} #end while loop to ensure regressions are completed

#Create a list (v_reg_model)for 1, 1int, 2, 2int
#it will hold models for OLS or WLS depending on norm_tru
v_reg_model <- list(one_int = model.1int,
                   two_int = model.2int)

#test for better fit
boo_linear <- CURVE_CHECK_f.r(v_reg_model$one_int,
v_reg_model$two_int, alpha, NORM_tru)
#  linear versus 2d order with interactions
#true means that there is little differnce
#between the models so use the simpler one, IE the main effects model
#if false -  fit 2d order model and optimize
#if true - stick with linear model and find a path of steepest ascent/descent and continue

if (boo_linear == FALSE) { # begin IF - linear model is inappropriate
  #use reg_model$two_int for prediction
  print("test confirms a better fit from the 2d order variance model")
  var_model <- v_reg_model$two_int

}  else{#use reg_model$two_int for prediction
  print("test confirms a better fit from the 1st order variance model")
  var_model <- v_reg_model$one_int } #end boo_linear loop for variance model



#Step 10
#Export the selected model to GAMS
```

```
##################################################################################################
##################################################################################################
##################################################################################################

##report optimal position
##translate from centered norm
Regular_Recommendation_point <- Best_factors
Regular_Recommendation_point_Norm <-
  rbind(updated_design,Best_factors)
Regular_Recommendation_point_Norm <-
  NORMMATRIX_and_center.r(Regular_Recommendation_point_Norm)
# normalized best point from RSM

  #ID the normalized means of the environmental variables
  Eval_Norm_mean <- seq(0, length.out = tot_evars) # intialize a holder
  for (cou in 1:tot_evars) {
  Eval_Norm_mean[cou] =
    colMeans(Regular_Recommendation_point_Norm[cou],
             na.rm = FALSE, dims = 1) }#take normalized mean of env vars }

Regular_Recommendation_point_Norm <-
  Regular_Recommendation_point_Norm[dim(Regular_Recommendation_point_Norm)[1],]
# this is the normed point


#0 for max 1 for min
GEN_GAMS_SCRIPTv4_generic.r
(1,paste("Properties_optim_",tot_var,"_",tot_evars  ,sep = ""),
 optim_work_dir, final_model, tot_var, tot_evars,
 Eval_Norm_mean, e_sd, var_model, sd_var)
#sink()


#robust recommendation point
Robust_Recommendation_point  <-
```

```
  as.data.frame( UN_NORMAL_and_center.r(t(as.matrix(c(Eval_Norm_mean,
  0))), as.matrix(updated_design), tot_var))


#report points
print("RSM Point")
print(Regular_Recommendation_point)
colnames(Robust_Recommendation_point) <-
  colnames(Regular_Recommendation_point)
print("Dual RSM Point")
print(Robust_Recommendation_point )

Dual_RSM_VALUE = 13.260 #from GAMS


#step 11
#look at impact of recommeded position above to determine the limits used for ellipse
########################################################################
###############################################################################
# run a simulation with the fixed control values and the range of environmental values
ellipse_sims <- 10000 #simulations to run (rows)
exp_DF <- as.data.frame(matrix(data = -10000000,
        nrow = ellipse_sims, ncol = tot_var))
colnames(exp_DF) <- colnames(factors_norm_reps_DF)
exp_DF[,(tot_evars+1):tot_var] = Robust_Recommendation_point[1,
(tot_evars+1):tot_var]

#create a set of environmental variables and place them in exp_DF columns
env_vals_data <- GEN_ENV_f.r(tot_var, tot_evars, e_chars)
for(trip in 1:((ellipse_sims/dim(env_vals_data)[2]))) {
    env_vals_data <- rbind(env_vals_data,GEN_ENV_f.r(tot_var,
    tot_evars, e_chars))}

exp_DF[,1:tot_evars]  = env_vals_data[1:dim(exp_DF)[1],]
#allow range of env variables and fixed control var
```

```
#gen replications
exp_DF_responses <- GEN_REPLICATES_f.r(as.matrix(exp_DF),
                             1, func_dir, func_number)

exp_DF <- cbind(exp_DF, exp_DF_responses)
colnames(exp_DF)[tot_var+1] <- "EX_response"



#Find the edges we want the min and max of each
#environmental variable that just meet the performance criteria
#gotta revisit this with a max and to change acceptable behavior
Satis_Beh_level <- 13 #anything less than this is okay

exp_DF_Satis_address <- which(exp_DF[,tot_var+1] < Satis_Beh_level)

Good_exp_DF <- exp_DF[exp_DF_Satis_address,] # DF with all good value

#find min and max of all columns
Add_Good_exp_DF <- Good_exp_DF[1:2,] # 1 is min 2 is max
  for (count in 1:tot_evars){
    Add_Good_exp_DF[1,count] = min(Good_exp_DF[,count])
    Add_Good_exp_DF[2,count] = max(Good_exp_DF[,count])
  }

Add_Good_exp_DF <- Add_Good_exp_DF[,1:tot_var]
#min amd max of all variables that meet satisficing behavior

#you could then vary the control variables with
#fixed env variables to see how policy matters
#brown doesn't like it

#step 12
#geometry measurement
##########################################################################
```

```
###############################################################################
## get confidence interval it's on excel sheet


##get ellipse SA
x_vec <- Add_Good_exp_DF[1,1:tot_evars] #initialize
x_vec <- abs(Add_Good_exp_DF[2,1:tot_evars]- Add_Good_exp_DF[1,1:tot_evars])/2

# you should normalize this in the real deal to eliminate the influence of scale
Satis_Area <- H_E_APPROX_f.r(x_vec, func_dir)

print("env var radii")
print(x_vec)

print("Hyper Ellipsoid Approximate Area")
print(Satis_Area)


###############################
#for fun
x_vec = c(10,20,30, 40)
Satis_Area <- H_E_APPROX_f.r(x_vec, func_dir)
print(Satis_Area)
#END
```

F. GENERAL ALGEBRAIC MODELING SYSTEM (GAMS) CODE

F.1 General Dual Response Surface Optimization Code

```
$ontext
solves generic problem for robust parameters
requires a nonlinear constraint now
$offtext

Variables
x_1 input environmental var1
x_2 input environmental var2
x_3 input environmental var3
x_4 input control var4
Outcome_of_Interest Objective Function ;

Equations
OBJ define objective function
con_E_1 ENV linear constraint 1
con_E_2 ENV linear constraint 2
con_E_3 ENV linear constraint 3
con_L_4 Control linear constraint 4
con_U_4 Control linear constraint 4
qcon1 quadratic constraint 1
qcon2 quadratic contraint 2 ;



OBJ.. Outcome_of_Interest =e= 12.3594083966838 + 1.16654866247444 * x_1 +
 3.37908676860181 * x_2 + 0.302698129144805 * x_3 + 2.14011408367422 * x_4
- 3.04224053644248 * x_1 * x_2 - 0.423047375812866 * x_1 * x_3
- 0.553524071931917 * x_1 * x_4 - 1.50330824578895 * x_2 * x_3
- 2.51929167231507 * x_2 * x_4 + 0.254340744296701 * x_3 * x_4 ;

$ontext
Subject To
$offtext
```

```
con_E_1 .. x_1 =e= 0.554636676355028 ;
con_E_2 .. x_2 =e= 0.254956103947439 ;
con_E_3 .. x_3 =e= 0.562918564456854 ;
con_L_4 .. x_4 =g= 0 ;
con_U_4 .. x_4 =l= 1 ;
qcon1 .. 12.3594083966838 + 1.16654866247444 * x_1
+ 3.37908676860181 * x_2 + 0.302698129144805 * x_3
+ 2.14011408367422 * x_4 - 3.04224053644248 * x_1 * x_2
- 0.423047375812866 * x_1 * x_3
- 0.553524071931917 * x_1 * x_4 - 1.50330824578895 * x_2 * x_3 -
2.51929167231507 * x_2 * x_4 + 0.254340744296701 * x_3 * x_4 =g= 0  ;

qcon2 .. 12.3594083966838 + 1.16654866247444 * x_1 +
3.37908676860181 * x_2 + 0.302698129144805 * x_3 +
2.14011408367422 * x_4 - 3.04224053644248 * x_1 * x_2 -
 0.423047375812866 * x_1 * x_3 - 0.553524071931917 * x_1 * x_4 -
1.50330824578895 * x_2 * x_3 - 2.51929167231507 * x_2 * x_4 +
0.254340744296701 * x_3 * x_4 =l= 5 ;

$ontext
RUN OPTIMIZATION MODEL
$offtext

Model Properties_optim_4_3 /all/ ;

Solve Properties_optim_4_3 using MIQCP minimizing Outcome_of_Interest ;

$ontext
consider all mixed quadratic solvers
MINLP
RMINLP *
MIQCP
RMIQCP *
MPEC
```

```
RMPEC *
other solvers MIP, RMIP, NLP, LP, MCP, CNS, DNLP, QCP
$offtext

Display x_1.l, x_2.l, x_3.l, x_4.l, Outcome_of_Interest.l ;
```

F.2 Jalalabad Dual Response Surface Optimization Code

```
$ontext
solves jbad problem for robust parameters
requires a nonlinear constraint now and I want to fix the
environmental variable values
$offtext

Variables
Wells input control var 1
Hospitals input control var 2
Security input control var 3
ANSF input control var 4
Wealth input environmental var 1
Ethnicity input environmental var 2
Deaths input environmental var3
Opinion Objective Function ;

Integer Variables Wells, Hospitals, Security, ANSF ;
Positive Variables Wealth, Ethnicity, Deaths ;

Equations
top define objective function (linear)
con0 linear constraint 0
con1 linear constraint 1
con2 linear constraint 2
con3 linear constraint 3
con4 linear constraint 4
con5 linear constraint 5
con6 linear constraint 6
con7 linear constraint 7
con8 linear constraint 8
con9 linear constraint 9
con10 linear constraint 10
qcon1 quadratic constraint 1
```

```
qcon2 quadratic contraint 2 ;



top.. Opinion =e= - 16636.0531117491 + 595.939093042811 * Wells -
546.51585629555 * Hospitals + 465.056607271189 * Security -
649.100453410244 * ANSF + 35.0994901600187 * Wealth +
386.322624758043 * Ethnicity + 37.6663470290266 * Deaths -
19.8706710909187 * Wells * Wells -
21.5234251244179 * Hospitals * Hospitals +
0 * Security * Security + 0 * ANSF * ANSF -
0.0341823638580988 * Wealth * Wealth -
2.20337403829448 * Ethnicity * Ethnicity
 - 0.60291830031308 * Deaths * Deaths +
2.75053396535278 * Wells * Hospitals
+ 21.9794889667183 * Wells * Security -
1.2140544609311 * Wells * ANSF
+ 3.22382473014026 * Wells * Wealth -
7.70755594205386 * Wells * Ethnicity +
2.57664540592889 * Wells * Deaths
+ 45.50878897696 * Hospitals * Security
+ 16.875661130284 * Hospitals * ANSF
+ 1.18835599790886 * Hospitals * Wealth
+ 6.15660603991043 * Hospitals * Ethnicity
+ 1.62614990572716 * Hospitals * Deaths -
78.8433812084868 * Security * ANSF
- 6.86048284112075 * Security * Wealth -
2.65946195593054 * Security * Ethnicity
 - 1.79870818554652 * Security * Deaths -
0.388540372070931 * ANSF * Wealth
+ 8.25098038825402 * ANSF * Ethnicity -
1.88796608043232 * ANSF * Deaths
- 0.362381551301084 * Wealth * Ethnicity -
0.359284346016301 * Wealth * Deaths
- 0.211434108085828 * Ethnicity * Deaths ;
```

```
$ontext
Subject To
$offtext

con0 .. Wells =g= 0 ;
con1 .. Wells =l= 3 ;
con2 .. Hospitals =g= 1 ;
con3 .. Hospitals =l= 3 ;
con4 .. Security =g= 0 ;
con5 .. Security =l= 1 ;
con6 .. ANSF =g= 0 ;
con7 .. ANSF =l= 1 ;
con8 .. Wealth =e= 29.9512195121951 ;
con9 .. Ethnicity =e= 85.0177383592018 ;
con10 .. Deaths =e= 13.5676274944568 ;
qcon1 .. 14897.7897265209 − 2427.86876328487 ∗ Wells
− 2010.21790062714 ∗ Hospitals − 9475.03680132395 ∗ Security
+ 23344.8014464483 ∗ ANSF + 484.454780790978 ∗ Wealth
− 141.428870728064 ∗ Ethnicity − 1286.21050142143 ∗ Deaths
+ 765.879206991294 ∗ Wells ∗ Wells +
2257.39361648258 ∗ Hospitals ∗ Hospitals
+ 0 ∗ Security ∗ Security + 0 ∗ ANSF ∗ ANSF +
2.00393871941766 ∗ Wealth ∗ Wealth
+ 2.65556871801094 ∗ Ethnicity ∗ Ethnicity −
9.40585017846207 ∗ Deaths ∗ Deaths
− 78.9282264559726 ∗ Wells ∗ Hospitals −
931.568570188538 ∗ Wells ∗ Security
− 2788.23691451205 ∗ Wells ∗ ANSF +
7.08405425930839 ∗ Wells ∗ Wealth
+ 23.6799145121029 ∗ Wells ∗ Ethnicity −
8.29033817746078 ∗ Wells ∗ Deaths
− 430.465720657925 ∗ Hospitals ∗ Security +
1664.16469042741 ∗ Hospitals ∗ ANSF
+ 19.5411894156747 ∗ Hospitals ∗ Wealth
```

```
- 101.307624276003 * Hospitals * Ethnicity
+ 17.3471550584183 * Hospitals * Deaths -
3472.93781382067 * Security * ANSF
+ 38.7464728507852 * Security * Wealth -
41.2603491836068 * Security * Ethnicity
- 41.4449073743057 * Security * Deaths +
38.5797233130601 * ANSF * Wealth
- 216.138554353917 * ANSF * Ethnicity -
181.430080081984 * ANSF * Deaths
- 7.50528995267218 * Wealth * Ethnicity -
5.68339997654623 * Wealth * Deaths
+ 21.6194574736393 * Ethnicity * Deaths =g= 0   ;

qcon2 .. 14897.7897265209 - 2427.86876328487 * Wells
- 2010.21790062714 * Hospitals - 9475.03680132395 * Security
+ 23344.8014464483 * ANSF + 484.454780790978 * Wealth
- 141.428870728064 * Ethnicity - 1286.21050142143 * Deaths
+ 765.879206991294 * Wells * Wells +
2257.39361648258 * Hospitals * Hospitals
+ 0 * Security * Security + 0 * ANSF * ANSF +
2.00393871941766 * Wealth * Wealth
+ 2.65556871801094 * Ethnicity * Ethnicity -
9.40585017846207 * Deaths * Deaths
- 78.9282264559726 * Wells * Hospitals -
931.568570188538 * Wells * Security
- 2788.23691451205 * Wells * ANSF +
7.08405425930839 * Wells * Wealth
+ 23.6799145121029 * Wells * Ethnicity -
8.29033817746078 * Wells * Deaths
- 430.465720657925 * Hospitals * Security +
1664.16469042741 * Hospitals * ANSF
+ 19.5411894156747 * Hospitals * Wealth
- 101.307624276003 * Hospitals * Ethnicity
+ 17.3471550584183 * Hospitals * Deaths -
3472.93781382067 * Security * ANSF
```

```
+ 38.7464728507852 * Security * Wealth -
41.2603491836068 * Security * Ethnicity
- 41.4449073743057 * Security * Deaths +
38.5797233130601 * ANSF * Wealth
- 216.138554353917 * ANSF * Ethnicity -
181.430080081984 * ANSF * Deaths
- 7.50528995267218 * Wealth * Ethnicity -
5.68339997654623 * Wealth * Deaths
+ 21.6194574736393 * Ethnicity * Deaths =l= 500 ;

$ontext
RUN OPTIMIZATION MODEL
$offtext

Model JBAD_optim /all/ ;

Solve JBAD_optim using MIQCP maximizing Opinion ;

$ontext
consider all mixed quadratic solvers
MINLP
RMINLP *
MIQCP
RMIQCP *
MPEC
RMPEC *
other solvers MIP, RMIP, NLP, LP, MCP, CNS, DNLP, QCP
$offtext

Display Wells.l, Hospitals.l, Security.l, ANSF.l, Wealth.l,
Ethnicity.l, Deaths.l, Opinion.l ;
```

F.3 Tijuana Dual Response Surface Optimization Code

```
$ontext
solves tj problem for robust parameters

$offtext

Variables
colonia_size input control var 1
maquiladoras input control var 2
city_growth input control var 3
required_capital input control var 4
carrying_capacity input control var 5
service_centers input control var 6
migration_ticks input environmental var1
crossing_ticks input environmental var2
init_density input environmental var3
building_ticks input environmental var4
No_Water_Rate Objective Function ;

Binary Variables city_growth ;
Integer Variables maquiladoras, required_capital,
carrying_capacity, service_centers, migration_ticks,
crossing_ticks, building_ticks ;
Positive Variables  colonia_size, init_density ;

Equations
top define objective function (linear)
con0 linear constraint 0
con1 linear constraint 1
con2 linear constraint 2
con3 linear constraint 3
con4 linear constraint 4
con5 linear constraint 5
con6 linear constraint 6
```

```
con7 linear constraint 7
con8 linear constraint 8
con9 linear constraint 9
con10 linear constraint 10
con11 linear constraint 11
con12 linear constraint 12
con13 linear constraint 13
con14 linear constraint 14
con15 linear constraint 15
qcon1 quadratic constraint 1
qcon2 quadratic contraint 2 ;



top.. No_Water_Rate =e= 2.70565692882453 +
1.70097733199104 * colonia_size + 0.145733931719902 * maquiladoras -
0.0468463038237284 * city_growth -
0.00915491608211609 * required_capital +
0.173292638066306 * carrying_capacity +
0.0116071317484104 * service_centers +
0.0250345023259445 * migration_ticks -
0.00937563867938412 * crossing_ticks +
0.142780628924493 * init_density -
0.0133557028374103 * building_ticks -
0.429523792333567 * colonia_size * colonia_size -
0.00239118868154128 * maquiladoras * maquiladoras +
0 * city_growth * city_growth +
5.32192577219425e-06 * required_capital * required_capital -
0.00215591259675241 * carrying_capacity * carrying_capacity +
0.0181566425895818 * service_centers * service_centers +
2.65870591147775e-06 * migration_ticks * migration_ticks -
2.07033521120881e-05 * crossing_ticks * crossing_ticks -
0.0973340563923149 * init_density * init_density +
0.000348284691709261 * building_ticks * building_ticks +
0.000802481606297115 * colonia_size * maquiladoras +
```

```
0.0116619028626727 * colonia_size * city_growth +
4.81916880636648e-05 * colonia_size * required_capital +
0.0110773027296882 * colonia_size * carrying_capacity -
0.0616140582352136 * colonia_size * service_centers -
0.00214444603183015 * colonia_size * migration_ticks +
0.000385776556828555 * colonia_size * crossing_ticks -
0.0254140360867118 * colonia_size * init_density +
0.00583991336468383 * colonia_size * building_ticks +
0.0039515664315855 * maquiladoras * city_growth -
0.000124401358283802 * maquiladoras * required_capital -
0.00262598988530515 * maquiladoras * carrying_capacity +
0.00043131737488799 * maquiladoras * service_centers +
0.000388416866027953 * maquiladoras * migration_ticks +
0.000356614921082457 * maquiladoras * crossing_ticks -
0.0155941638533717 * maquiladoras * init_density +
0.000288492240722143 * maquiladoras * building_ticks -
6.29798748177082e-06 * city_growth * required_capital +
0.00207707797922167 * city_growth * carrying_capacity -
0.00172452992433469 * city_growth * service_centers -
0.000121624852736611 * city_growth * migration_ticks -
0.000530809099338623 * city_growth * crossing_ticks +
0.0355021222011358 * city_growth * init_density +
0.00055127557403271 * city_growth * building_ticks -
0.000202497467914457 * required_capital * carrying_capacity
- 4.04995651133513e-05 * required_capital * service_centers
- 2.06526560017132e-05 * required_capital * migration_ticks
+ 1.02214920894501e-05 * required_capital * crossing_ticks
+ 0.000167355245768302 * required_capital * init_density +
1.28320879381106e-05 * required_capital * building_ticks +
0.00326421785862024 * carrying_capacity * service_centers +
6.09299705696893e-05 * carrying_capacity * migration_ticks -
 0.000252052807034688 * carrying_capacity * crossing_ticks -
 0.0109963206738276 * carrying_capacity * init_density +
 0.00280077380318248 * carrying_capacity * building_ticks
- 0.00100405548340211 * service_centers * migration_ticks
```

```
− 8.92187748347011e−05 ∗ service_centers ∗ crossing_ticks
+ 0.0022380257125031 ∗ service_centers ∗ init_density
− 0.00104492022005652 ∗ service_centers ∗ building_ticks
+ 6.32935145256014e−06 ∗ migration_ticks ∗ crossing_ticks
− 0.000803822763073537 ∗ migration_ticks ∗ init_density −
0.000302551522867998 ∗ migration_ticks ∗ building_ticks −
4.13522374509258e−05 ∗ crossing_ticks ∗ init_density +
4.21416968011705e−05 ∗ crossing_ticks ∗ building_ticks
− 0.0146106379049282 ∗ init_density ∗ building_ticks  ;

$ontext
Subject To
$offtext

con0 .. colonia_size  =g= 0 ;
con1 .. colonia_size  =l= 2 ;
con2 .. maquiladoras  =g= 1 ;
con3 .. maquiladoras  =l= 8 ;
con4 .. city_growth  =g= 0 ;
con5 .. city_growth  =l= 1 ;
con6 .. required_capital  =g= 150 ;
con7 .. required_capital  =l= 1200 ;
con8 .. carrying_capacity  =g= 1 ;
con9 .. carrying_capacity  =l= 6 ;
con10 .. service_centers  =g= 1 ;
con11 .. service_centers  =l= 3 ;
con12 .. migration_ticks =e= 7.32297099937409 ;
con13 .. crossing_ticks =e= 9.51679532651784 ;
con14 .. init_density =e= 0.801752555810557 ;
con15 .. building_ticks =e= 3.25443354892552 ;
qcon1 .. 2.70565692882453 + 1.70097733199104 ∗ colonia_size
+ 0.145733931719902 ∗ maquiladoras −
0.0468463038237284 ∗ city_growth −
0.00915491608211609 ∗ required_capital +
0.173292638066306 ∗ carrying_capacity +
```

```
0.0116071317484104 * service_centers +
0.0250345023259445 * migration_ticks -
0.00937563867938412 * crossing_ticks +
0.142780628924493 * init_density -
0.0133557028374103 * building_ticks -
0.429523792333567 * colonia_size * colonia_size -
0.00239118868154128 * maquiladoras * maquiladoras +
0 * city_growth * city_growth +
5.32192577219425e-06 * required_capital * required_capital -
0.00215591259675241 * carrying_capacity * carrying_capacity +
 0.0181566425895818 * service_centers * service_centers +
2.65870591147775e-06 * migration_ticks * migration_ticks
- 2.07033521120881e-05 * crossing_ticks * crossing_ticks
- 0.0973340563923149 * init_density * init_density
+ 0.000348284691709261 * building_ticks * building_ticks
+ 0.000802481606297115 * colonia_size * maquiladoras
+ 0.0116619028626727 * colonia_size * city_growth
+ 4.81916880636648e-05 * colonia_size * required_capital
+ 0.0110773027296882 * colonia_size * carrying_capacity
- 0.0616140582352136 * colonia_size * service_centers
- 0.00214444603183015 * colonia_size * migration_ticks
+ 0.000385776556828555 * colonia_size * crossing_ticks
- 0.0254140360867118 * colonia_size * init_density
+ 0.00583991336468383 * colonia_size * building_ticks
+ 0.0039515664315855 * maquiladoras * city_growth
- 0.000124401358283802 * maquiladoras * required_capital
- 0.00262598988530515 * maquiladoras * carrying_capacity
+ 0.00043131737488799 * maquiladoras * service_centers
+ 0.000388416866027953 * maquiladoras * migration_ticks
+ 0.000356614921082457 * maquiladoras * crossing_ticks
- 0.0155941638533717 * maquiladoras * init_density
+ 0.000288492240722143 * maquiladoras * building_ticks
- 6.29798748177082e-06 * city_growth * required_capital
+ 0.00207707797922167 * city_growth * carrying_capacity
- 0.00172452992433469 * city_growth * service_centers
```

```
- 0.000121624852736611 * city_growth * migration_ticks
- 0.000530809099338623 * city_growth * crossing_ticks
+ 0.0355021222011358 * city_growth * init_density
+ 0.00055127557403271 * city_growth * building_ticks
- 0.000202497467914457 * required_capital * carrying_capacity
- 4.04995651133513e-05 * required_capital * service_centers
- 2.06526560017132e-05 * required_capital * migration_ticks
+ 1.02214920894501e-05 * required_capital * crossing_ticks
+ 0.000167355245768302 * required_capital * init_density
+ 1.28320879381106e-05 * required_capital * building_ticks
+ 0.00326421785862024 * carrying_capacity * service_centers
+ 6.09299705696893e-05 * carrying_capacity * migration_ticks
- 0.000252052807034688 * carrying_capacity * crossing_ticks
- 0.0109963206738276 * carrying_capacity * init_density
+ 0.00280077380318248 * carrying_capacity * building_ticks
- 0.00100405548340211 * service_centers * migration_ticks
- 8.92187748347011e-05 * service_centers * crossing_ticks
+ 0.0022380257125031 * service_centers * init_density
- 0.00104492022005652 * service_centers * building_ticks
+ 6.32935145256014e-06 * migration_ticks * crossing_ticks
- 0.000803822763073537 * migration_ticks * init_density
- 0.000302551522867998 * migration_ticks * building_ticks
- 4.13522374509258e-05 * crossing_ticks * init_density
+ 4.21416968011705e-05 * crossing_ticks * building_ticks
- 0.0146106379049282 * init_density * building_ticks =g= 0 ;

qcon2 .. 2.70565692882453 + 1.70097733199104 * colonia_size
+ 0.145733931719902 * maquiladoras -
0.0468463038237284 * city_growth -
0.00915491608211609 * required_capital
+ 0.173292638066306 * carrying_capacity
+ 0.0116071317484104 * service_centers
+ 0.0250345023259445 * migration_ticks
- 0.00937563867938412 * crossing_ticks
+ 0.142780628924493 * init_density
```

```
− 0.0133557028374103 ∗ building_ticks
− 0.429523792333567 ∗ colonia_size ∗ colonia_size
− 0.00239118868154128 ∗ maquiladoras ∗ maquiladoras
+ 0 ∗ city_growth ∗ city_growth +
5.32192577219425e−06 ∗ required_capital ∗ required_capital
− 0.00215591259675241 ∗ carrying_capacity ∗ carrying_capacity
+ 0.0181566425895818 ∗ service_centers ∗ service_centers
+ 2.65870591147775e−06 ∗ migration_ticks ∗ migration_ticks
− 2.07033521120881e−05 ∗ crossing_ticks ∗ crossing_ticks
− 0.0973340563923149 ∗ init_density ∗ init_density
+ 0.000348284691709261 ∗ building_ticks ∗ building_ticks
+ 0.000802481606297115 ∗ colonia_size ∗ maquiladoras
+ 0.0116619028626727 ∗ colonia_size ∗ city_growth
+ 4.81916880636648e−05 ∗ colonia_size ∗ required_capital
+ 0.0110773027296882 ∗ colonia_size ∗ carrying_capacity
− 0.0616140582352136 ∗ colonia_size ∗ service_centers
− 0.00214444603183015 ∗ colonia_size ∗ migration_ticks
+ 0.000385776556828555 ∗ colonia_size ∗ crossing_ticks
− 0.0254140360867118 ∗ colonia_size ∗ init_density
+ 0.00583991336468383 ∗ colonia_size ∗ building_ticks
+ 0.0039515664315855 ∗ maquiladoras ∗ city_growth
− 0.000124401358283802 ∗ maquiladoras ∗ required_capital
− 0.00262598988530515 ∗ maquiladoras ∗ carrying_capacity
+ 0.00043131737488799 ∗ maquiladoras ∗ service_centers
+ 0.000388416866027953 ∗ maquiladoras ∗ migration_ticks
+ 0.000356614921082457 ∗ maquiladoras ∗ crossing_ticks
− 0.0155941638533717 ∗ maquiladoras ∗ init_density
+ 0.000288492240722143 ∗ maquiladoras ∗ building_ticks
− 6.29798748177082e−06 ∗ city_growth ∗ required_capital
+ 0.00207707797922167 ∗ city_growth ∗ carrying_capacity
− 0.00172452992433469 ∗ city_growth ∗ service_centers
− 0.000121624852736611 ∗ city_growth ∗ migration_ticks
− 0.000530809099338623 ∗ city_growth ∗ crossing_ticks
+ 0.0355021222011358 ∗ city_growth ∗ init_density
+ 0.00055127557403271 ∗ city_growth ∗ building_ticks
```

```
- 0.000202497467914457 * required_capital * carrying_capacity
- 4.04995651133513e-05 * required_capital * service_centers
- 2.06526560017132e-05 * required_capital * migration_ticks
+ 1.02214920894501e-05 * required_capital * crossing_ticks
+ 0.000167355245768302 * required_capital * init_density
+ 1.28320879381106e-05 * required_capital * building_ticks
+ 0.00326421785862024 * carrying_capacity * service_centers
+ 6.09299705696893e-05 * carrying_capacity * migration_ticks
- 0.000252052807034688 * carrying_capacity * crossing_ticks
- 0.0109963206738276 * carrying_capacity * init_density
+ 0.00280077380318248 * carrying_capacity * building_ticks
- 0.00100405548340211 * service_centers * migration_ticks
- 8.92187748347011e-05 * service_centers * crossing_ticks
+ 0.0022380257125031 * service_centers * init_density
- 0.00104492022005652 * service_centers * building_ticks
+ 6.32935145256014e-06 * migration_ticks * crossing_ticks
- 0.000803822763073537 * migration_ticks * init_density
- 0.000302551522867998 * migration_ticks * building_ticks
- 4.13522374509258e-05 * crossing_ticks * init_density
+ 4.21416968011705e-05 * crossing_ticks * building_ticks
- 0.0146106379049282 * init_density * building_ticks =l= .002 ;

$ontext
RUN OPTIMIZATION MODEL
$offtext

Model TJ_optim_experiment_3 /all/ ;

Solve TJ_optim_experiment_3 using MIQCP minimizing No_Water_Rate ;

$ontext
consider all mixed quadratic solvers
MINLP
RMINLP *
MIQCP
```

```
RMIQCP *
MPEC
RMPEC *
other solvers MIP, RMIP, NLP, LP, MCP, CNS, DNLP, QCP
$offtext

Display colonia_size.l, maquiladoras.l, city_growth.l,
required_capital.l, carrying_capacity.l, service_centers.l,
migration_ticks.l, crossing_ticks.l, init_density.l,
building_ticks.l, No_Water_Rate.l ;
```

G. ACRONYMS AND ABBREVIATIONS

ABM - Agent-based model

AIC - Akaike Information Criterion

AGC - US Army Geospatial Center

ANN - Artificial Neural Network

ANSF - Afghan National Security Forces

ARB - Acceptable Response Boundary

ASCE - American Society of Civil Engineers

BIC - Bayesian information Criterion

BLUE - Best Least Unbiased Estimator

CCD - Central Composite Design

CCL-CBM - Center for Connected Learning and Computer-Based Modeling

CF - Coalition Forces

CLT - Central Limit Theorem

C/NRCD - Center for Nation Reconstruction and Capacity Development

COIN - Counter-insurgency

COA - Course of Action

C_p - Mallow's C_p Statistic

DOD - Department of Defense

DA - Department of the Army

DF - Direct Fire

DOE - Design of Experiment

FM - Field Manual

FOUO - For Official Use Only

GA - Genetic Algorithm

GAM - Generalized Additive Model

GAMS - General Algebraic Modeling System

GIRoA - Government of the Islamic Republic of Afghanistan

GLM - Generalized Linear Model

IED - Improvised Explosive Device

IID - Independent and Identically Distributed

IIMM - International Infrastructure Management Manual

ISAF - International Security Assistance Force

JIEDDO - Joint IED Defeat Organiztion

LOO - Line of Operation

MARS - Multivariate Adaptive Regression Splines

MMRDP - Meta-Model and Robust Design Process

MSE - Mean Squared Error

NAFTA - North American Free Trade Agreement

NATO - North Atlantic Treaty Organization

NGIA - National Geospatial Intelligence Agency

NMSE - Normalized Mean Squared Error

NN - Neural Networks

NOLHS - Nearly Orthogonal Latin Hyper-cube Sampling

NPS - Naval Postgraduate School

OLS - Ordinary Least Squares

PGD - Path of Greatest Degradation

PSA - Path of Steepest Ascent

PSD - Path of Steepest Descent

R&R - Reliability and Reproducibility

RSM - Response Surface Methodology

RMSE - Root Mean Squared Error

SA - Simulated Annealing

SNR - Signal to Noise Ratio

SSE - Sum of Squared Errors

UN - United Nations

USACE - United States Army Corps of Engineers

YLL - Years of Life Lost

H. THE YOUNG BRITISH SOLDIER

Rudyard Kipling, 1895

WHEN the 'arf-made recruity goes out to the East
'E acts like a babe an' 'e drinks like a beast,
An' 'e wonders because 'e is frequent deceased
Ere 'e's fit for to serve as a soldier.
Serve, serve, serve as a soldier,
Serve, serve, serve as a soldier,
Serve, serve, serve as a soldier,
So-oldier of the Queen!

Now all you recruities what's drafted to-day,
You shut up your rag-box an' 'ark to my lay,
An' I'll sing you a soldier as far as I may:
A soldier what's fit for a soldier.
Fit, fit, fit for a soldier . . .

First mind you steer clear o' the grog-sellers' huts,
For they sell you Fixed Bay'nets that rots out your guts -
Ay, drink that 'ud eat the live steel from your butts -
An' it's bad for the young British soldier.
Bad, bad, bad for the soldier . . .

When the cholera comes - as it will past a doubt -
Keep out of the wet and don't go on the shout,
For the sickness gets in as the liquor dies out,
An' it crumples the young British soldier.
Crum-, crum-, crumples the soldier . . .

But the worst o' your foes is the sun over'ead:
You must wear your 'elmet for all that is said:
If 'e finds you uncovered 'e'll knock you down dead,
An' you'll die like a fool of a soldier.
Fool, fool, fool of a soldier . . .

If you're cast for fatigue by a sergeant unkind,
Don't grouse like a woman nor crack on nor blind;
Be handy and civil, and then you will find
That it's beer for the young British soldier.
Beer, beer, beer for the soldier . . .

Now, if you must marry, take care she is old -
A troop-sergeant's widow's the nicest I'm told,
For beauty won't help if your rations is cold,
Nor love ain't enough for a soldier.
'Nough, 'nough, 'nough for a soldier . . .

If the wife should go wrong with a comrade, be loath
To shoot when you catch 'em - you'll swing, on my oath! -
Make 'im take 'er and keep 'er: that's Hell for them both,
An' you're shut o' the curse of a soldier.
Curse, curse, curse of a soldier . . .

When first under fire an' you're wishful to duck,
Don't look nor take 'eed at the man that is struck,
Be thankful you're livin', and trust to your luck
And march to your front like a soldier.
Front, front, front like a soldier . . .

When 'arf of your bullets fly wide in the ditch,
Don't call your Martini a cross-eyed old bitch;
She's human as you are - you treat her as sich,
An' she'll fight for the young British soldier.
Fight, fight, fight for the soldier . . .

When shakin' their bustles like ladies so fine,
The guns o' the enemy wheel into line,
Shoot low at the limbers an' don't mind the shine,
For noise never startles the soldier.
Start-, start-, startles the soldier . . .

If your officer's dead and the sergeants look white,
Remember it's ruin to run from a fight:
So take open order, lie down, and sit tight,
And wait for supports like a soldier.
Wait, wait, wait like a soldier . . .

When you're wounded and left on Afghanistan's plains,
And the women come out to cut up what remains,
Jest roll to your rifle and blow out your brains

An’ go to your Gawd like a soldier.
Go, go, go like a soldier,
Go, go, go like a soldier,
Go, go, go like a soldier,
So-oldier of the Queen!

I. REFERENCES

BIBLIOGRAPHY

[1] "Insurgent mexico." http://warriorpublications.wordpress.com/2011/12/22/insurgent-mexico/.

[2] S. G. Jones, *In the Graveyard of Empires: America's War in Afghanistan*. W. W. Norton & Company, reprint ed., Apr. 2010.

[3] U. S. D. of the Army., *The U.S. Army/Marine Corps counterinsurgency field manual : U.S. Army field manual no. 3-24 : Marine Corps warfighting publication no. 3-33.5 /*. Chicago :: University of Chicago Press,, university of chicago press ed. ed., 2007.

[4] E. Lopez, J. Salmon, T. Sisk, C. Torori, S. Campbell, and P. Morgan, "Capacity development in post-conflict." www.undp.org/content/undp/en/home/librarypage/capacity-building/capacity-development-in-post-conflict-countries/, 2010.

[5] H. Russ, "New york, new jersey put $71 billion price tag on sandy." http://www.usnews.nbcnews.com/_news/2012/11/26/15463835-new-york-new-jersey-put-71-billion-price-tag-on-sandy, Nov. 2012.

[6] D. o. D. DOD, "Progress toward security and stability in afghanistan." http://www.gpo.gov/, 2009.

[7] "CIA - the world factbook." https://www.cia.gov/library/publications/the-world-factbook/geos/af.html.

[8] A. Belasco, "The cost of iraq, afghanistan, and other global war on terror operations since 9/11." http://www.fas.org/sgp/crs/natsec/RL33110.pdf, 2007.

[9] A. Rashid, *Descent into Chaos: The U.S. and the Disaster in Pakistan, Afghanistan, and Central Asia*. Penguin (Non-Classics), revised ed., Apr. 2009.

[10] R. Mardirosian, "Infrastructure development in the shadow of conflict: Aligning incentives and attracting investment." www.crgp.stanford.edu/publications/, 2010.

[11] G. S. Parnell, P. J. Driscoll, and D. L. Henderson, *Decision Making in Systems Engineering and Management*. Wiley-Interscience, 1 ed., Feb. 2008.

[12] A. S. o. C. Engineers, *Report Card for America's Infrastructure 2009*. Amer Society of Civil Engineers, Jan. 2009.

[13] N. Z. A. M. Support, *International Infrastructure Management Manual 2006 Edition*. New Zealand Asset Management Support, Jan. 2006.

[14] "World bank infrastructure data," February 2013.

[15] A. Wagner, "Grundlegung der politischen oekonomie /." http://search.lib.virginia.edu/catalog/007958230, 1883.

[16] D. A. Aschauer, "Is public expenditure productive?," *Journal of monetary economics*, vol. 23, pp. 177–200, Mar. 1989.

[17] T. Lakshmanan, "The broader economic consequences of transport infrastructure investments," *Journal of Transport Geography*, vol. 19, no. 1, pp. 1 – 12, 2011.

[18] J. P. Cohen, "Economic benefits of investments in transport infrastructure,"

[19] S. Perz, A. Shenkin, G. Barnes, L. Cabrera, L. Carvalho, and J. Castillo, "Connectivity and resilience: A multidimensional analysis of infrastructure impacts in the southwestern amazon," *Social Indicators Research*, vol. 106, no. 2, pp. 259–285, 2012.

[20] E. Cavallo and C. Daude, "Public investment in developing countries: A blessing or a curse?," *Journal of Comparative Economics*, vol. 39, pp. 65–81, Mar. 2011.

[21] J. P. C. Kleijnen, S. M. Sanchez, T. W. Lucas, and T. M. Cioppa, "A user's guide to the brave new world of designing simulation experiments," *INFORMS Journal on Computing*, vol. 17, no. 3, pp. 263–289, 2005.

[22] A. Law and W. D. Kelton, *Simulation Modeling and Analysis*. McGraw-Hill Science/Engineering/Math, 3 ed., Dec. 1999.

[23] S. M. Sanchez, F. Moeeni, and P. J. Sanchez, "So many factors, so little time simulation experiments in the frequency domain," *International Journal of Production Economics*, vol. 103, pp. 149–165, Sept. 2006.

[24] T. M. Cioppa, T. W. Lucas, and S. M. Sanchez, "Military applications of agent-based simulations," in *Proceedings of the 36th conference on Winter simulation*, WSC '04, (Washington, D.C.), p. 171–180, Winter Simulation Conference, 2004. ACM ID: 1161774.

[25] "Project albert home." http://www.projectalbert.org/.

[26] D. Lovelace, V. Lalis, E. Lesnowicz, C. Michel, E. Teague, and S. Vandervleit, "Comparison of ground-based fire support capabilities of the marine expeditionary unit," *Scythe - Proceedings and Bulletin of the International Data Farming Community*, no. 2, pp. 32–33, 2007.

[27] S. Tisue and U. Wilensky, "Netlogo: A simple environment for modeling complexity," in *International Conference on Complex Systems*, pp. 16–21, 2004.

[28] U. Wilensky, "Netlogo." http://ccl.northwestern.edu/netlogo/, 1999.

[29] G. Young and J. Flacke, "Agent-based model of the growth of an informal settlement in dar es salaam, tanzania: An empirically informed concept," *International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environments Sake, Fifth Biennial Meeting, Ottawa, Canada*, 2010.

[30] G. E. P. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 13, no. 1, pp. pp. 1–45, 1951.

[31] J. P. Kleijnen, *Design and Analysis of Simulation Experiments*. Springer, softcover reprint of hardcover 1st ed. 2008 ed., Nov. 2010.

[32] A. I. Khuri and S. Mukhopadhyay, "Response surface methodology," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, pp. 128–149, Mar. 2010.

[33] A. Shaibu and B. Cho, "Another view of dual response surface modeling and optimization in robust parameter design," *The International Journal of Advanced Manufacturing Technology*, vol. 41, no. 7, pp. 631–641, 2009.

[34] G. E. P. Box, "Multi-Factor experimental designs for exploring response surfaces," *The Annals of Mathematical Statistics*, vol. 28, pp. 195–241, Mar. 1957. Mathematical Reviews number (MathSciNet): MR85679; Zentralblatt MATH identifier: 0080.35901.

[35] T. M. Cioppa and T. W. Lucas, "Efficient nearly orthogonal and Space-Filling latin hypercubes," *Technometrics*, vol. 49, pp. 45–55, Feb. 2007.

[36] J. Neter, M. Kutner, W. Wasserman, C. Nachtsheim, and J. Neter, *Applied Linear Statistical Models*. McGraw-Hill/Irwin, 4 ed., Feb. 1996.

[37] A. Agresti, *Categorical Data Analysis*. Wiley-Interscience, 2 ed., July 2002.

[38] E. W. Weisstein, "Likelihood ratio – from wolfram MathWorld." http://mathworld.wolfram.com/PowerMean.html.

[39] R. D. Reiss, *Statistical analysis of extreme values : from insurance, finance, hydrology and other fields /*. Basel ;: Birkhaäuser,.

[40] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001 ed., Mar. 2012.

[41] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, pp. pp. 370–384, 1972.

[42] C. E. McCulloch, *Generalized, linear, and mixed models /*. Hoboken, N.J. :: Wiley,, 2nd ed. ed.

[43] R. H. Myers, *Generalized linear models : with applications in engineering and the sciences /*. New York :: J. Wiley,.

[44] G. G. Moisen and T. S. Frescino, "Comparing five modelling techniques for predicting forest characteristics," *Ecological Modelling*, vol. 157, pp. 209–225, Nov. 2002.

[45] L. Amaral and J. Ottino, "Complex systems and networks: challenges and opportunities for chemical and biological engineers," *Chemical Engineering Science*, vol. 59, pp. 1653–1666, Apr. 2004.

[46] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer, 2nd ed. 2009. corr. 3rd printing 5th printing. ed., Feb. 2009.

[47] S. Crino and D. E. Brown, "Global optimization with multivariate adaptive regression splines," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, pp. 333–340, Apr. 2007.

[48] J. Friedman, "Multivariate adaptive regression splines," *Annals of Statistics*, vol. 19, pp. 1–67, 1991.

[49] J. Leathwick, J. Elith, and T. Hastie, "Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions," *Ecological Modelling*, vol. 199, pp. 188–196, Nov. 2006.

[50] K. B. Misra, ed., *Handbook of Performability Engineering.* Springer, 2008 ed., Aug. 2008.

[51] W. Klibi, A. Martel, and A. Guitouni, "The design of robust value-creating supply chain networks: A critical review," *European Journal of Operational Research*, vol. 203, pp. 283–293, June 2010.

[52] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios, "Robust optimization of Large-Scale systems," *Operations Research*, vol. 43, pp. 264–281, Mar. 1995.

[53] P. Vincke, "Robust solutions and methods in decision-aid," *Journal of Multi-Criteria Decision Analysis*, vol. 8, no. 3, p. 181–187, 1999.

[54] K. Sorensen, "A framework for robust and flexible optimisation using metaheuristics," *4OR: A Quarterly Journal of Operations Research*, vol. 1, no. 4, pp. 341–345, 2003.

[55] J. Rosenhead, M. Elton, and S. K. Gupta, "Robustness and optimality as criteria for strategic decisions," *Operational Research Quarterly (1970-1977)*, vol. 23, no. 4, pp. pp. 413–431, 1972.

[56] P. Kouvelis and G. Yu, *Robust Discrete Optimization and Its Applications*. Springer, 1997.

[57] H.-Y. Wong and J. Rosenhead, "A rigorous definition of robustness analysis," *The Journal of the Operational Research Society*, vol. 51, no. 2, pp. pp. 176–182, 2000.

[58] F. Mosteller and J. W. Tukey, *Data Analysis and Regression: A Second Course in Statistics*. Pearson, 1 ed., Jan. 1977.

[59] R. Andersen, *Modern Methods for Robust Regression*. SAGE Publications, Inc, Sept. 2007.

[60] G. Taguchi, *System of experimental design : engineering methods to optimize quality and minimize costs /*. White Plains, N.Y. :: UNIPUB-/Kraus International Publications ;.

[61] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. pp. 591–611, 1965.

[62] T. R. Foundation, "The r project for statistical computing." http://www.r-project.org/, 2010.

[63] F. Bingen, C. Siau, and P. Rousseeuw, "Applying robust regression techniques to institutional data," *Research in Higher Education*, vol. 25, no. 3, pp. pp. 277–297, 1986.

[64] J. Fox, *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications, Inc, Feb. 1997.

[65] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. pp. 871–880, 1984.

[66] D. L. Massart, L. Kaufman, P. J. Rousseeuw, and A. Leroy, "Least median of squares: a robust method for outlier and model error detection in regression and calibration," *Analytica Chimica Acta*, vol. 187, pp. 171–179, 1986.

[67] J. A. Doornik, "Robust estimation using least trimmed squares," Mar 2011.

[68] W. Stevenson, *Operations Management*. McGraw-Hill/Irwin, 10 ed., Oct. 2008.

[69] J. Engel and A. F. Huele, "A generalized linear modeling approach to robust design," *Technometrics*, vol. 38, no. 4, pp. pp. 365–373, 1996.

[70] W. Welch, T. Yu, S. Kang, and J. Sacks, "Computer experiments fo quality control by parameter design," *Journal of Quality Technology*, pp. 15–22, 1990.

[71] A. C. Shoemaker, K.-L. Tsui, and C. F. J. Wu, "Economical experimentation methods for robust design," *Technometrics*, vol. 33, no. 4, pp. pp. 415–427, 1991.

[72] G. Vining and R. Myers, "Combining taguchi and response surface philosophies: A dual response approach," *Journal of Quality Technology*, vol. 22, pp. pp. 38–45, 1990.

[73] R. H. Myers, A. I. Khuri, and G. Vining, "Response surface alternatives to the taguchi robust parameter design approach," *The American Statistician*, vol. 46, no. 2, pp. pp. 131–139, 1992.

[74] J. A. Nelder and Y. Lee, "Generalized linear models for the analysis of taguchi-type experiments," *Applied Stochastic Models and Data Analysis*, vol. 7, no. 1, p. 107–120, 1991.

[75] S. M. Sanchez, "Design of experiments: robust design: seeking the best of all possible worlds," in *Proceedings of the 32nd conference on Winter simulation*, WSC '00, (San Diego, CA, USA), p. 69–76, Society for Computer Simulation International, 2000.

[76] A. C. Atkinson and M. Riani, "Building regression models with the forward search.," *Journal of Computing and Information Technology*, vol. 15, no. 4, pp. 287 – 294, 2007.

[77] "The general algebraic modeling system." http://www.gams.com/, 2012.

[78] "Gurobi optimization," Feb. 2013.

[79] "IBM - mathematical programming: Linear programming, mixed-integer programming and quadratic programming - IBM ILOG CPLEX optimizer - software." http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/, Feb. 2013.

[80] G. E. P. Box and N. R. Draper, *Empirical model-building and response surfaces*. Wiley series in probability and mathematical statistics., Oxford, England: John Wiley & Sons, 1987.

[81] H. E. Haber, "Physics 116a course notes."

[82] N. Vasconcelos, "Dimensionality and dimensionality reduction dimensionality reduction."

[83] A. E. Lawrence, "The volume of an n-dimensional hypersphere."

[84] S. R. Keller, "On the surface areas of the ellipsoid," *Mathematics of Computation*, Jan 1979.

[85] G. J. Tee, "The volume of an n-dimensional hypersphere." http://www.math.auckland.ac.nz/Research/Reports/Series/539.pdf.

[86] D. C. Montgomery, "Experimental design for product and process design and development.," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 48, no. 2, p. 159, 1999.

[87] G. Michon, "Spheroids and scalene ellipsoids." http://www.numericana.com, 2013.

[88] E. W. Weisstein, "Power mean – from wolfram MathWorld." http://mathworld.wolfram.com/PowerMean.html.

[89] G. Michon, "Geometry." http://www.numericana.com, 2013.

[90] R. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[91] E. Teague, T. Warner, and D. Brown, "Evaluating infrastructure resource allocation in support of regional stability," *International Journal of System of Systems Engineering*, vol. 3, no. 2, pp. 154–167, 2012.

[92] J. Lambert, C. Karvetski, D. Spencer, B. Sotirin, D. Liberi, H. Zaghloul, J. Koogler, S. Hunter, W. Goran, R. Ditmer, and I. Linkov, "Prioritizing infrastructure investments in afghanistan with multiagency stakeholders and deep uncertainty of emergent conditions," *Journal of Infrastructure Systems*, vol. 18, no. 2, pp. 155–166, 2012.

[93] U.S.Army, *Stability Operations: Field Manual 3-07.* Headquarters United States Army, field manual ed., Oct. 2008.

[94] M. Ryan, "The military and reconstruction operations.(Viewpoint essay)," *Parameters (Carlisle, Pa.)*, vol. 37, p. 58, Dec. 2007.

[95] T. Warner, J. Bowling, C. Erikson, D.Hui, C. Marsh, E. Teague, and D. Brown, "Agent-Based modeling of public health infrastructure projects in jalalabad, afghanistan," Jan. 2011.

[96] "Maslows hierarchy of needs." http://communicationtheory.org/maslow%e2%80%99s-hierarchy-of-needs/.

[97] A. H. Maslow, "A theory of human motivation.," *Psychological Review*, vol. 50, no. 4, pp. 370–396, 1943.

[98] A. N. D. S. Secretariat, *Afghanistan National Development Strategy: 1387-1391 (2008-2013).* Islamic Republic of Afghanistan, internet ed., 2008.

[99] J. D. Montgomery and D. A. Rondinelli, *Beyond Reconstruction in Afghanistan: Lessons from Development Experience.* Palgrave Macmillan, first edition ed., July 2007.

[100] "Operation enduring freedom casualties." http://icasualties.org.

[101] E. Kennedy, "Poll: More afghans support attacks on american troops,"

[102] A. Rashid, *Taliban: Militant Islam, Oil and Fundamentalism in Central Asia, Second Edition.* Yale University Press, 2 ed., Apr. 2010.

[103] M. J. Hogan, *The Marshall Plan: America, Britain and the Reconstruction of Western Europe, 1947-1952.* Cambridge University Press, Jan. 1989.

[104] R. Stackelberg, *The Routledge Companion to Nazi Germany.* Routledge, Jan. 2008.

[105] K. Didow and J. Jacob, "Ten years after september 11: An analysis of public opinion in the muslim world," *Belfer Center for Science and International Affairs*, 2011.

[106] T. Christensen and P. Lægreid, "Trust in government: The relative importance of service satisfaction, political factors, and demography," *Public Performance & Management Review*, vol. 28, pp. 487–511, June 2005. ArticleType: research-article / Full publication date: Jun., 2005 / Copyright 2005 M.E. Sharpe, Inc.

[107] S. KULL, C. RAMSAY, S. WEBER, E. LEWIS, and E. MOHSENI, "Public opinion in the islamic world on terrorism, al qaeda, and us policies." http://www.worldpublicopinion.org, 2009.

[108] S. Weber, "Afghan public opinion amidst rising violence." http://www.worldpublicopinion.or, 2006.

[109] N. I. o. Justice, "Factors that influence public opinion of the police." http://www.nij.gov/pubs-sum/197925.htm, 2003.

[110] N. Malleson, "repastcity - a simple repast simphony virtual city model. - google project hosting." http://code.google.com/p/repastcity/, 2012.

[111] T. S. Project, *The Sphere Handbook 2011: Humanitarian Charter and Minimum Standards in Humanitarian Response.* Practical Action, Pap/Cdr ed., Apr. 2011.

[112] "Tijuana workforce." http://tijuana-edc.com/why-tijuana.

[113] R. Biederman, "Immex manufacturing and investment opportunities," (Chicago, IL), U.S. - Mexico Chamber of Commerce, February 2012.

[114] *International trade Mexico's maquiladora decline affects U.S.-Mexico border communities and trade : recovery depends in part on Mexico's actions : report to congressional requesters /.* Washington, D.C. :: U.S. General Accounting Office,, 2003.

[115] D. Darlin, "Maquiladora-ville.," *Forbes*, vol. 157, no. 9, pp. 111 – 112, 1996.

[116] P. Cooney, "The mexican crisis and the maquiladora boom: A paradox of development or the logic of neoliberalism?," *Latin American Perspectives*, vol. 28, no. 3, pp. pp. 55–83, 2001.

[117] T. Fullerton, R. Tinajero, and J. M. Cota, "An empirical analysis of tijuana water consumption," *Atlantic Economic Journal*, vol. 35, no. 3, pp. 357–369, 2007.

[118] "Border 2012: U.s.-mexico environmental program." http://www.epa.gov/border2012/.

[119] J. Hadjimarcou, L. E. Brouthers, J. P. McNicol, and D. E. Michie, "Maquiladoras in the 21st century: Six strategies for success," *Business Horizons*, no. 0, pp. –, 2012.

[120] Y. H. Yang, "Post-NAFTA production and sourcing development in the maquiladora program," *Industrial Management & Data Systems*, vol. 98, pp. 269–268, Sept. 1998.

[121] R. Rhoda and T. Burton, *Geo-Mexico: the geography and dynamics of modern Mexico*. Sombrero Books, first ed., Jan. 2010.

[122] J. Williams, "Demographics and trends in the california mexico border region." http://www.energy.ca.gov/2005publications/CEC-600-2005-011/CEC-600-2005-011.PDF, 2005.

[123] O. Sosa, "Border planning in the san diego-tijuana region: Local planning and national policy.," *Berkeley Planning Journal*, vol. 21, pp. 165 – 181, 2008.

[124] "Public utilities infrastructure for manufacturing facilities in tijuana." http://tijuana-edc.com/why-tijuana/infrastructure/public-utitlies.

[125] "Baja california's community based needs." http://www.icfdn.org/NA/volume1/main.php.

[126] "North american free trade agreement (NAFTA) (canada-united states-mexico [1992]) – britannica online encyclopedia." http://www.britannica.com/EBchecked/topic/418784/North-American-Free-Trade-Agreement-NAFTA.

[127] P. Majkut, "The other chiapas: Llabor unrest in soconusco, chiapas, mexico," *Multinational Monitor*, vol. 16, June 1995.

[128] "Mexico's maquiladora labor system keeps workers in poverty," *McClatchy Newspapers*.

[129] M. de la Luz Inclan, "From the ya basta to the caracoles: Zapatista mobilization under transitional conditions," *American Journal of Sociology*, vol. 113, no. 5, pp. pp. 1316–1350, 2008.

[130] D. Bacon, "Tijuana maquiladora labor strike deepens," *LA Weekly*.

[131] S. Dibble, "Brief strike at maquiladora is over," *San Diego Union Tribune*.

[132] S. Babones, "The middling kingdom.," *Foreign Affairs*, vol. 90, no. 5, pp. 79 – 88, 2011.

[133] M. Thompson, "Foxconn's china workers to get more union rights." http://money.cnn.com/2013/02/04/investing/foxconn-china-union/index.html, Feb. 2013.

[134] "Chinese pollution protesters turn violent in clash with police." http://worldnews.nbcnews.com/_news/2012/07/28/13002897-chinese-pollution-protesters-turn-violent-in-clash-with-police.

[135] G. G. Chang, "Is foxconn fleeing china? sure looks like it.." www.forbes.com/sites/gordonchang/2013/02/24/is-foxconn-fleeing-china-sure-looks-like-it/, Feb. 2013.

[136] Maciamo, "Japanese society economy and politics wapedia." http://www.wapedia.com/economy/japan_postwar_economic_miracle.shtml, May 2004.

[137] R. Florida, "Which countries pay blue collar workers the most?," *The Atlantic*, January 2012.

[138] R. E. Cole, *Japanese Blue Collar: The Changing Tradition*. University of California Press, first edition ed., Mar. 1973.

[139] S. D. Targum and J. Kitanaka, "Overwork suicide in japan," *Innovations in Clinical Neuroscience*, vol. 9, pp. 35–38, Feb. 2012. PMID: 22468243 PMCID: PMC3312902.

[140] J. M. Ottino, "Engineering complex systems," *Nature*, vol. 427, pp. 399–399, Jan. 2004.

[141] F. M. De Leon, F.D. and U. Wilensky, "Tijuana bordertowns model - center for connected learning and computer-based modeling." http://ccl.northwestern.edu/netlogo/models/UrbanSuite-TijuanaBordertowns, 2007.

[142] G. Taguchi, S. Chowdhury, and Y. Wu, *Taguchi's Quality Engineering Handbook*. Wiley-Interscience, 1 ed., Nov. 2004.

[143] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Springer, 2nd ed., Mar. 2002.

[144] J. Durbin, *Time Series Analysis by State Space Methods*. Oxford University Press, 2 ed., July 2012.

[145] M. L. George, J. Maxey, D. Rowlands, and M. Price, *The Lean Six Sigma Pocket Toolbook: A Quick Reference Guide to 100 Tools for Improving Quality and Speed*. McGraw-Hill, 1 ed., Aug. 2004.

[146] E. Koren, John, *The History of Statistics. Their Development and Progress in Many Countries*. MacMillan & Co LTD, 1St edition ed., 1918.

[147] N. K. Avkiran and N. Thoraneenitiyan, "Purging data before productivity analysis," *Journal of Business Research*, vol. 63, pp. 294–302, Mar. 2010.

[148] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models(text only)1st (First) edition[Paperback]2006*. Cambridge University Press, 1st (First) edition ed., 2006.

[149] OECD, *Impact of Transport Infrastructure Investment on Regional Development*. Paris: Organisation for Economic Co-operation and Development, May 2002.

[150] "Rebuilding the roads of afghanistan." www.transition.usaid.gov.

[151] "On the roads again in afghanistan." www.pbs.org/wnet/need-to-know/opinion/on-the-roads-again-in-afghanistan/3223/.

[152] "Provincial and peri-urban water supply and sanitation project, royal kingdom of cambodia." www-wds.worldbank.org.

[153] "UNICEF - afghanistan - afghanistan: water and sanitation services under fire." www.unicef.org/infobycountry/afghanistan/7167.html?p=printme.

[154] "Athi water services board environmental impact assessment report." www.wds.worldbank.org.

[155] M. Hallett, "Distributed power in afghanistan: The padisaw micro-hydro project," *Renewable Energy*, vol. 34, pp. 2847–2851, Dec. 2009.

[156] "Import of power from uzbekistan, tajikistan, and turkmenistan."

[157] "Policy brief: Health and the environment." www.oecd.org.