3-2000

Using RSM, DOE, and Linear Regression to Develop a Metamodel to Predict Cargo Delivery of a Time Phase Force Deployment Document

Kenneth S. Browne

AFIT/GOA/ENS/00-01

Using RSM, DOE, and Linear Regression to Develop a Metamodel to Predict Cargo Delivery of a Time Phase Force Deployment Document

THESIS

Ken S. Browne
Captain, USAF

AFIT/GOA/ENS/00-01

Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 2000	Master's Thesis

4. TITLE AND SUBTITLE

Using RSM, DOE, and Linear Regression to Develop a Metamodel to Predict Cargo Delivery of a Time Phase Force Deployment Document

5. FUNDING NUMBERS

6. AUTHOR(S)

Ken S. Browne, Captain, USAF

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 P Street, Building 640
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GOA/ENS/00M-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Lt Col Robert Brigantic
HQ AMC/XPY
402 Scott Dr. Unit 3L3
Scott AFB, IL 62225
Robert.Brigantic@scott.af.mil DSN 576-8713

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Major Paul Murdock, Asst. Professor, AFIT/ENS
(937) 255-6565 x4339 William.Murdock@afit.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

ABSTRACT (*Maximum 200 Words*)

Air Mobility Command (AMC) uses the Airlift Flow Model as their primary tool to estimate the amount of cargo delivered in a Time Phase Force Deployment Document (TPFDD). The primary objective of this research was an exploratory investigation in the development of a metamodel to predict the amount of cargo delivered from a TPFDD by AMC into a theater. In creating a valid metamodel the analyst would be able to quickly provide the decision-maker with accurate insights should input parameters change. This would save valuable time and replace the need to physically change the input parameters and re-run the simulation. Techniques that were applicable to create this metamodel include DOE, RSM, and Linear Regression. Using the techniques outlined in this research a second metamodel was constructed using a separate set of data to validate the procedure. In both cases, the results substantiated good predictive capability between the simulation and the metamodel. The analysis procedures outlined in this effort allows the researcher to identify the salient factors to the metamodel in a timely, efficient manner.

14. SUBJECT TERMS	15. NUMBER OF PAGES
RSM, Metamodel, Linear Regression, Design of Experiment	83
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNC

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

AFIT/GOA/ENS/00-01

Using RSM, DOE, and Linear Regression to Develop a Metamodel to Predict Cargo Delivery of a Time Phase Force Deployment Document

THESIS

Presented to the Faculty of the Graduate School of Engineering and Management
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Operations Analysis

Ken S. Browne, B.S.
Captain, USAF

March 2000

AFIT/GOA/ENS/00M-01

THESIS APPROVAL

Student: Ken S. Browne, Captain, USAF Class: GOA-00M

Title: Using RSM, DOE, and Linear Regression to Predict Cargo Delivery of a Time Phase Force Deployment Document

Defense Date: 24 February 2000

Committee:	Name/Title/Department	Signature
Advisor	W. Paul Murdock, Maj, USAF Assistant Professor Department of Operational Sciences	W. Paul Murdock Jr
Reader	John Crown, Maj, USAF Assistant Professor Department of Mathematics and Statistics	John S. Crown

Acknowledgements

I would not have completed this thesis without the support of many people. I owe a large debt of gratitude to my advisor, Maj Paul Murdock, and my reader, Maj John Crown, for allowing me the freedom to pursue this topic in a manner that I felt would be the most beneficial.

I would also like to thank all the good friends that pulled me through over the last 18 months. Julia, Jeff, Bruce, Roger, Mike, Dave, and Gary (A.K.A. The Tau), I wouldn't have been able to do it without you. I would be remiss if I didn't thank the AFIT faculty as a whole for taking the extra time to answer my many questions.

None of this would have been possible without the love and support of my family. I thank and love you with all my heart. To my wonderful children Abigail and Jeremy, I appreciate all the sacrifices that you have made in order for this to happen for us. And to the beautiful lady in my life; Elizabeth, you pulled me through the tough times, even when you didn't realize you were doing it. Forever will I be thankful. I love you.

Ken S. Browne

Table of Contents

Page

Acknowledgements iii

List of Figures vii

List of Tables viii

Abstract ix

I. **Introduction** 1-1

1.1 **Background** 1-1

1.2 **Problem Statement** 1-3

1.3 **Objective** 1-4

1.4 **Methodology Overview** 1-4

1.5 **Summary** 1-6

II. **Literature Review** 2-1

2.1 **Introduction** 2-1

2.2 **Computer Simulation** 2-1

2.3 **Verification and Validation** 2-3

2.4 **Design of Experiment** 2-4

2.4.1 **Box-Behnken Design** 2-4

2.4.2 **Central Composite Design** 2-6

2.4.3 **Plackett-Burman Design** 2-7

2.4.4 **Aliasing** 2-8

2.4.5 **De-aliasing** 2-9

2.4.6 **Response Surface Methodology** 2-10

		Page
	2.5 **Linear Regression**	2-11
	2.6 **Metamodels**	2-12
	2.6.1 **Constructing Metamodels**	2-12
	2.6.2 **Metamodel Validation**	2-14
	2.7 **Summary**	2-15
III.	**Methodology**	3-1
	3.1 **Introduction**	3-1
	3.2 **Definition of the Problem**	3-1
	3.3 **The Model**	3-2
	3.4 **Statement of Assumptions**	3-4
	3.5 **Definition of Variables**	3-5
	3.6 **Design of Experiment**	3-6
	3.7 **Simulation Runs**	3-6
	3.8 **Output Analysis**	3-8
	3.9 **Additional Insights**	3-8
	3.10 **Summary**	3-9
IV.	**Results and Additional Insights**	4-1
	4.1 **Introduction**	4-1
	4.2 **Assumptions**	4-1
	4.2.1 **Mandatory**	4-1
	4.2.2 **Necessary**	4-1
	4.3 **Preliminary Analysis**	4-2
	4.4 **Full Model Regression Analysis**	4-7
	4.4.1 **Normality**	4-7
	4.4.2 **Constant Variance**	4-7
	4.4.3 **Lack of Fit Test**	4-9

		Page
	4.5 **Metamodel Verification**	4-10
	4.6 **Metamodel Validation**	4-11
	4.7 **Variable Screening**	4-13
	4.8 **Full Model Summary**	4-13
	4.9 **Reduced Model Analysis**	4-13
	4.10 **Stationary Points**	4-15
	4.11 **Optimization**	4-19
	4.12 **Additional Insights**	4-21
	4.13 **Full Model Regression Analysis IMS**	4-22
	4.13.1 **Normality IMS**	4-24
	4.13.2 **Constant Variance IMS**	4-27
	4.14 **Metamodel Verification IMS**	4-29
	4.15 **Metamodel Validation IMS**	4-30
	4.16 **System Comparison**	4-31
	4.17 **Conclusion**	4-32
V.	**Conclusions and Recommendations**	5-1
	5.1 **Introduction**	5-1
	5.2 **Conclusions**	5-1
	5.3 **Design Setting**	5-1
	5.4 **Recommendations**	5-2
	5.4.1 **Introduction**	5-2
	5.4.2 **Fuel and MOG**	5-2
	5.4.3 **Convert Passengers to Tons**	5-3
	5.5 **Conclusion**	5-3
Appendix A.	Glossary of Acronyms and Abbreviations	A-1
Bibliography		BIB-1

List of Figures

Figure		Page
1.1.	AMC's Through-put pipeline	1-3
3.1.	AFM Delivery Capabilities	3-3
3.2.	AFM High Level Block Diagram	3-3
4.1.	Full Model Least Squares Regression Plot	4-4
4.2.	Full Model Outlier Plot	4-6
4.3.	Full Model Residual Normality Plot	4-7
4.4.	Full Model Residual Plot	4-8
4.5.	Maximum Cargo throughput Response	4-18
4.6.	Maximum Cargo Throughput Response	4-19
4.7.	Cargo Throughput Optimization IP	4-20
4.8.	Optimal Throughput Response	4-21
4.9.	Full Model Least Squares Regression Plot - IMS	4-23
4.10.	Full Model Outlier Plot - IMS	4-24
4.11.	Full Model Residual Normality Plot - IMS	4-25
4.12.	Residual Stem and Leaf Plot - IMS	4-26
4.13.	Residual Box Plot - IMS	4-26
4.14.	Full Model Residual Plot - IMS	4-28
4.15.	Distribution of $Z_j's$	4-32

List of Tables

Table		Page
2.1.	Box-Behnken Design	2-5
2.2.	Full 2^3 BBD	2-5
2.3.	Central Composite Design	2-6
2.4.	2^{3-1} Half Fraction Design with Interaction	2-8
2.5.	2^3 Full Factorial Design with Interaction	2-10
3.1.	Number of Runs	3-7
4.1.	Full Model Parameter Estimates, P-values, and VIF	4-3
4.2.	Coded Variable Conversion Chart	4-3
4.3.	Full Model Regression ANOVA	4-4
4.4.	Test Data Verification	4-10
4.5.	Mean Absolute Percent Error of Validation Data	4-12
4.6.	Full Model Parameter Estimates, P-values, and VIF - IMS	4-22
4.7.	Full Model Regression ANOVA - IMS	4-23
4.8.	Quantiles and Moments	4-27
4.9.	Test Data Verification - IMS	4-29
4.10.	Mean Absolute Percent Error of Validation Data - IMS	4-30
5.1.	Coded and Uncoded Design Settings for Max Throughput	5-2

AFIT/GOA/ENS/00-01

Abstract

Air Mobility Command (AMC) uses the Airlift Flow Model as their primary tool to estimate the amount of cargo delivered in a Time Phase Force Deployment Document (TPFDD). The primary objective of this research was an exploratory investigation in the development of a metamodel to predict the amount of cargo delivered from a TPFDD by AMC into a theater. In creating a valid metamodel the analyst would be able to quickly provide the decision maker with accurate insights should input parameters change. This would save valuable time and replace the need to physically alter the input parameters and re-run the simulation. Techniques that were applicable to create this metamodel include DOE, RSM, and Linear Regression. Using the techniques outlined in this research, a second metamodel was constructed using a separate set of data to validate the procedure. In both cases, the results substantiated good predictive capability between the simulation and metamodel. The analysis procedures outlined in this effort allows the researcher to identify the salient factors to the metamodel in a timely, efficient manner. Once the metamodel has been constructed and validated, it may be possible to optimize it using integer programming techniques or some other software package. By doing this, it may be possible to examine the difference between the optimal solution and your current solution. This difference may be the decisive factor that warrants further experimentation of the system or provides additional verification that you are operating within some pre-established tolerance level.

Using RSM, DOE, and Linear Regression to Develop a Metamodel to Predict Cargo Delivery of a Time Phase Force Deployment Document

I. Introduction

1.1 Background

Since its birth, strategic airlift has been one of the main functions of the United States Air Force. Aircraft then carried much less and flew shorter distances than today's fleet. Consequently, in the past we have had to maintain many more en route air bases to service this fleet. With the advancement of the aircraft, i.e. the jet engine, we can fly much further and carry greater payloads. As a result of being able to fly farther and faster, the Air Force has lost many of its en route bases. This loss in en route infrastructure has forced the Air Force to be more effective and efficient in the scheduling of its assets.

Air Mobility Command Headquarters (AMC) located at Scott AFB, IL implements a simulation tool, which falls under the Mobility Analyst Support System (MASS). MASS is an umbrella of models designed to give decision-makers a quick look at different problems. One of the models that fall under this umbrella is the Airlift Flow Model (AFM). AFM is a FORTRAN based stochastic model (which is being converted to an object oriented design) designed to transport the airlift portion of a Time Phased Force Deployment Document (TPFDD) to its final offload location - for instance in Southwest Asia.

There are volumes of output that can be obtained from this simulation. One statistic that is output is total tons delivered per day, meaning how much cargo was actually delivered in theater on that day. It is from this output that we can see how delivery is affected by numerous factors such as the addition or subtraction of airframes, capacity of en route bases, and the positioning of aircrews to name a few. Even with our reduced footprint of overseas en route bases, and an even smaller working MOG[1] for our offload delivery locations, the amount of cargo that we can deliver is fairly linear - to a point. With the addition of more aircraft you reach a point where you level off and can actually deliver less cargo than a scenario with fewer aircraft. Why does this happen? It happens, in part, because AFM is a campaign level model; therefore, individual details are left out or more elegantly put, they are aggregated into the stochastic process. In fact, individual detail for moving an entire TPFDD would be very difficult to achieve, thus difficult to program without using distributions to aggregate the minuscule tasks that need to be accounted for and not assumed away.

Another reason that cargo delivery begins to taper off is because of the amount of airframes that are in the system. With few aircraft in the system, scheduling (for the simulation) is relatively easy. There are enough resources available (e.g. working MOG, fuel, material handling equipment, etc.) to efficiently service the aircraft in the system. Also, if we follow the path of a single aircraft from its home station in a sparsely populated aircraft scenario, we see that when it lands, it is immediately serviced due to the fact that there are few other aircraft competing for the resources at its current base. After it completes its ground time, it departs and heads to its next station. AFM plans missions from recovery base to recovery base. This sequence of events occurs until the aircraft reaches its offload destination, off-loads its cargo, then flies to the recovery station and is ready to receive either a new mission or if the simulation can find no new mission, the aircraft is routed back to its home station.

[1]MOG stands for Maximum on Ground. There are several different types of MOG - parking, working, fueling, etc. For my research I will refer only to Working MOG, which is the number of planes that may land on an airfield and be immediately serviced

When you add more aircraft to the scenario bottlenecks occur. That one aircraft that we followed around the system is now going head to head with other aircraft all vying for the same base resources of which fuel and working MOG shortages seem to be the driving factors. Figure 1.1 below, illustrates the problem. As the flow of aircraft leaves the CONUS through the pipeline, they are constrained by the few number of en route bases that they can land and be serviced at. This limits the amount of aircraft that can be effectively put into the system and ultimately impedes cargo delivery.

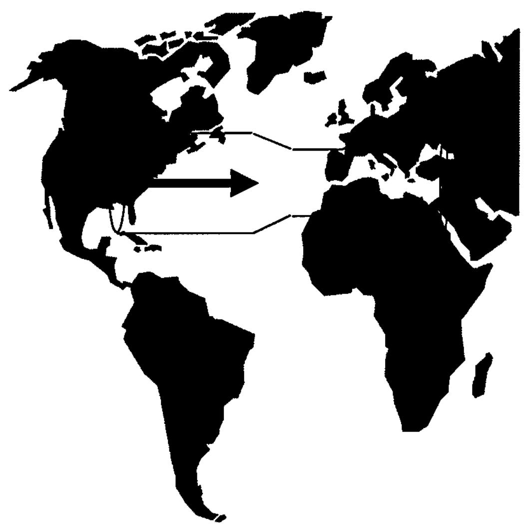

Figure 1.1 AMC's Through-put pipeline

1.2 Problem Statement

As stated before, AMC's concern is the amount of cargo that is delivered in theater per day. This is proposed to be a linear function until the transportation system becomes saturated with aircraft at which time it is believed that cargo delivery stagnates and may even become nonlinear.

1.3 Objective

The purpose of my research will be to develop a metamodel that can be used to predict a response (in this case amount of tons delivered in theater) when certain parameters are changed. By adjusting these parameters, an analyst will be able to predict the total cargo delivered for the scenario without re-running the simulation. An offshoot of this research will be to propose a method to optimize the metamodel. Optimization, however unlikely, will provide an upper bound to the problem.

1.4 Methodology Overview

The problem was attacked in the following way. First, and most importantly, the variables for the experiment were defined. These variables were scaled down from: number of aircraft, amount of fuel, en route MOG, and offload MOG, to just varying the amount of aircraft utilized. The reason for this is the allotted overseas MOG and fuel available to AMC will be dictated to AMC. The only factor in AMC's control is the amount of aircraft used in the scenario. Within the aircraft variable lie the sub variables of aircraft types. This thesis effort focuses on the following types of aircraft: Wide Body Passenger (WBP), Wide Body Cargo (WBC), and Narrow Body Passenger (NBC) all of which make up the Civil Reserve Air Fleet (CRAF). On the military side of the house, C-5s, C-17s, KC-10s, KC-135s, and the ultra heavy airlifter (UHA) make up the organic portion of the airlift system. The UHA is a proposed aircraft that will help to bridge the gap between airlift and sealift. The C-141B will not be used in this study as it is being retired and will be phased out of service completely by the year 2006.

The next step performed was to design a ten seed 2^5 full factorial central composite design with ten axial runs and 8 center runs. This design would be used as a template to run the simulation and conduct a normal linear regression on the data. The reason for this is simple, if the linear process shows a good fit, then there is no need to attempt nonlinear regression. However, if there is not a good linear fit to the data, then an investigation of either a transformation of variables, or a nonlinear approach would be examined.

The second step that would be performed would again involve using regular linear regression, but this time with transformed data. This procedure will only be necessary if the non-transformed data proves to be a poor fit for the linear regression model. This will produce different results than using the data without transformation. If the residual error decreases after fitting the data to a least squares regression line then further scrutiny of this approach may be explored.

If non-linearity seemed to be prevalent in the data, the next step would try a piecewise linear or nonlinear regression approach to the problem. Unless the data is unmanageable, such as some sort of exponential function, we should be able to use a transformation technique to make the data linear.

Dave Merrill from HQ AMCSAF remarks that if you give him a rickshaw and an infinite amount of time, he can deliver an entire TPFDD. Well, time is the one luxury that AMC does not possess. When the balloon goes up and bombs and bullets are needed in Theater X, unless there is a good pre-positioning system in place, it takes too long for the current air lift system to move the required amount of personnel and equipment from their home station or aerial port of embarkation (APOE) to the aerial port of debarkation (APOD). It takes a ship approximately 18-20 days to go from the Continental United States to a sea port of debarkation in theater. More amazing is the fact that one Large Medium Speed Roll-on Roll-off (LMSR) ship can carry approximately 175-185 C-5 loads. At present there are 19 LMSR's in the naval inventory. If we assumed that 10 of these were ready and available to be used for any scenario, and they are loaded immediately, then by day 20 there could be anywhere from 1,750 - 1,850 C-5 loads of cargo in theater just from sealift. Which equates to about 113,750 to 120,250 tons. If we look at what AMC delivers in those same 20 days we would see that the total amount of cargo moved by air is right around 175,000 tons. How can we get much needed troops and equipment in place in a shorter amount of time?

Enter the UHA. The UHA is a proposed aircraft that will have the capacity of eight to ten C-5s and will require as much as four to eight times the amount of parking space. This MOG capacity will severely limit where the UHA can offload its cargo since the MOG at most of the off-loads in theater are not capable of supporting more than two or three C-5s. Fuel and maintenance requirements are undisclosed at this time.

Since en route MOG, offload MOG, and fuel variables are going to be predetermined no matter what scenario is run, those variables were treated as constants and therefore, not varied. This study did not concern MOG efficiency or fuel consumption. Rather it looked at aircraft utilization. After the factors were determined, a design of experiment using response surface methodology techniques was formulated. Even though there are actually eight factors, only a 3^5 design of experiment was prepared. The reason for this is that the total number of NBC aircraft that are allotted to AMC are insignificant to the problem. Also left out were the KC-10s and KC-135s because their role would change depending on the severity of the conflict. What this means is that for smaller contingencies some of these aircraft would be used for cargo utilization capabilities. For larger contingencies, more aircraft (fighters and cargo) would need to be ferried across the ocean, thereby reducing the numbers that could be used for cargo delivery. This leaves only five aircraft, the C-5, C-17, UHA, WBC, and the WBP to be evaluated.

1.5 ***Summary***

Even though the presence of overseas en route bases has dwindled from 39 to 13, the requirement to carry out the mission of the Air Force has grown. The need to get people and equipment to austere locations in an expedited manner still exists. But, is throwing all available aircraft into the system the answer? More may not necessarily be better. The purpose of this research was to devise a metamodel that will be able to predict cargo throughput. As a by product of this research, the number of aircraft that gives the greatest amount of delivery will be disclosed.

II. Literature Review

2.1 Introduction

The purpose of this research is to develop a metamodel that can be used to estimate the tons of cargo delivered for a TPFDD without re-running the simulation. The approach implemented to achieve this objective uses the following tools: computer simulation, verification and validation of the computer simulation, design of experiments (DOE), response surface methodology techniques (RSM), linear regression, construction of a metamodel, and finally, validation of the metamodel.

2.2 Computer Simulation

A computer simulation model often portrays the dynamic behavior of a system over time. Models are constructed to provide information about real systems when conditions prevent the real system from being exercised. Examples of these conditions include but are not limited to: the cost of running an actual experiment, vastness of the system that is being studied (for instance trying to study the effects of changing a company's overseas shipping routes), or when an experiment is hazardous to the environment such as a nuclear explosive test. Virtually any system may be simulated via a computer. Although the more complex the simulation, the more likely you are to omit detail in favor of stochastic representation. Consider, for example, a manufacturing situation with the goal of maximizing production. What if management has narrowed the decision on increasing production to two options? Option one is to purchase new, more efficient, equipment. And option two is to hire a third shift for 24-hour plant operation. Management would like a cost effective way to analyze the pros and cons of both actions. These options may be evaluated two ways. First, compare the two alternatives by altering the actual system. This would entail hiring a third shift of employees, training them and evaluating their output once they have been trained. But profit could only be measured after all expenses for the new employees have been deducted. Once a steady state in production has been reached and a baseline of profit established, it is time to test option two. However, to test option two we have to get rid of option one. In simpler text, management would have to layoff our newly trained third shift in favor of purchasing the new equipment.

Even with new equipment, there may be the added cost of training the present crew on using it before you can reach a steady state and establish another profit baseline. The final step in option one would be to compare the profit margins of both methods and implement the one with the greatest yield. If the hiring of a third shift were the better alternative, then you would have to recall the shift that you previously terminated, assuming that none of those employees have found alternate work and no retraining were necessary. Now that you have hired all the employees back, the hardware that you have purchased is now a sunk cost. Conversely, if you decide to keep the hardware, there are still employee benefits to be paid out to the shift that was laid off.

A third alternative is to simulate both of the previous options without physically changing the system. By using simulation, you can "reproduce" the system you are interested in without affecting current production. Such experimentation is generally conducted with either physical or mathematical models [Law and Kelton, 1991:3-7]. The above example is a case where a physical model would not be of service due to interruption of the present system and its associated costs. There are, however, systems that avail themselves to physical models. For example, aeronautical engineers use scaled-down models to study airflow patterns in wind tunnels.

As an alternative to physical models, mathematical models use quantitative and logical relationships to characterize the system. For relatively simple mathematical models, analytical solutions can be calculated in order to characterize the performance of the system [Taylor, 1994: 2-2].

In the manufacturing example described before, it would be a simple task to compute the yield of either alternative if a direct relationship between the number of new machines and the number of additional workers could be established. But since there are qualitative factors involved in each, such a relationship cannot be easily obtained.

2.3 Verification and Validation

Once we have a simulation built for our system, we must ensure we have the correct simulation for our system. Did we build the correct model? One of the ways to verify our model is to dissect each section or subroutine of the model and determine if the logic is correct. This approach works well for small simulations or large ones without much detail. For the manufacturing example, the time between two stations (perhaps from the lathe to the sander) could be physically measured on the actual system and this value checked against the models results. If the average of these values are within some established tolerance you have set, assuming that you allow for some variance in travel speeds, that section of the model can be deemed verified. The rest of the model can be verified in the same manner. In reality you would only measure the time between machines if it directly impacted the simulation. Otherwise you could just assume it away as negligible travel time. A common mistake often made by beginning modelers is to include an excessive amount of model detail [Law and Kelton, 1991: 301].

The next step in the process is to validate our model. Did we build the model right. A simulation is a surrogate of an actual or proposed system. Keeping this in mind we must make sure that our simulation model is robust enough to make decisions about the system similar to those that would be made if it were feasible and cost-effective to experiment with the real system itself.

One should keep in mind that a simulation is developed for a particular purpose. One model that is valid for one purpose may not be valid for another. Refer back to the manufacturing example. The production of widgets and the production of automobiles both involve machining parts and using conveyor belts, but the times spent at each machine and on the belts would be different.

Speaking to subject matter experts is one way to validate the model. Rarely will the analyst know all the details about what he/she is trying to simulate. By discussing what he/she hopes to accomplish with people who know the specifics about the "real world", the analyst can implement more accurate detail in the simulation. Once the simulation is complete, output must be checked for accuracy. One way to accomplish this would be to use a Turing test [Banks, Carson, and Nelson 1999: 423]. To use the Turing test, one would take output data from the real system and an equal amount of data from the simulation, shuffle them together and give them to a subject matter expert. If the subject matter expert can consistently discern the simulated reports from the actual reports, then more work needs to be done on the simulation. However, if the expert cannot differentiate between the real and simulated reports, the modeler can conclude that the test provides no evidence of model inadequacy and the model may be used for its intended purpose.

Once the simulation has passed the verification, validation, and accreditation checklist, the analyst can construct a simulation experiment to cover the experimental region of interest.

2.4 ***Design of Experiment***

There are several types of experimental designs that are widely used today. A few of the more common ones are the Box - Behnken, Central Composite, and the Plackett Burman designs.

2.4.1 ***Box-Behnken Design.*** Box-Behnken designs (BBD) are an efficient measure to fit a three level design for a second order model [Myers and others, 1995: 318]. The way that a BBD design works is to form balanced but incomplete block designs. An example of a balanced but incomplete design with three treatment levels is shown in Table 2.1.

Treatment			
	x_1	x_2	x_3
Block 1	X	X	
Block 2	X		X
Block 3		X	X

Table 2.1 Box-Behnken Design

If we were to pair the X's from this design in block one and make them a full 2^2 experiment while the third factor remains zero and do the same for blocks two and three the resulting design matrix would look like this

x_1	x_2	x_3
-1	-1	0
1	-1	0
-1	1	0
1	1	0
-1	0	-1
1	0	-1
-1	0	1
1	0	1
0	-1	-1
0	1	-1
0	-1	1
0	1	1
0	**0**	**0**

Table 2.2 Full 2^3 BBD

The last line in the design matrix of Table 2.2 is a vector of center runs. When the number of parameters equals four or seven, then center runs are necessary to avoid singularity in the matrix [Myers and others, 1995: 319]. One of the downfalls of the BBD is that it is a spherical design, meaning if you were to inscribe that sphere on the inside of a cube, you would soon realize that you cannot reach the corner points of the cube. Therefore the BBD is not a good design if you need to corner points of the cube. Therefore the BBD is not a good design if you need to predict responses at the extreme or corner points. Equation 2.1 shows how the values of this matrix are obtained.

2.4.2 ***Central Composite Design.*** Table 2.3 shows the Central Composite Design (CCD). The CCD is widely used when fitting a second order response surface. If we were to envision a 2^2 experiment using the manufacturing example, and our factors were the lathe (x_1) and the sander (x_2) the design structure would resemble this:

Run	x_1	x_2	**Response (Min)**
1	-1	-1	20
2	1	-1	21
3	-1	1	64
4	1	1	47
5	-1.414	0	42
6	1.414	0	66
7	0	-1.414	68
8	0	1.414	25
9	0	0	38
10	0	0	35
11	0	0	32
12	0	0	43

Table 2.3 Central Composite Design

You can see that run sequence one through four alone would be a full 2^2 factorial design. The coded values for the design (± 1, 0, and ± 1.414) can be obtained using the following formula

$$x_i = \frac{\xi_i - [\max(\xi_i) - \min(\xi_i)]/2}{[\max(\xi_i) - \min(\xi_i)]/2} \tag{2.1}$$

where:

ξ_i = the amount of resource ξ_i that you are currently using in the experiment

$\max(\xi_i)$ = the maximum amount of resource ξ_i that you possess

$\min(\xi_i)$ = the minimum amount of resource ξ_i that you possess

Design points five through eight constitute what are known as axial or star points. The term axial comes from the fact that the points lie on the x_1 or the x_2 axes at a radius of 1.414 from the center of the design. This distance from the center of the design to where the axial point is, changes with the number of factors that you have in the design. However, you will always have $2 \times n$ number of axial points where n is the number of factors in the model. The last four design points, nine through twelve, are center or baseline runs. They are used to minimize variance in the output. Center runs result when $\xi_i = [\max(\xi_i) - \min(\xi_i)]/2$. Central composite designs are extremely useful when fitting second order response surfaces.

2.4.3 ***Plackett-Burman Design.*** In 1946 R. L. Plackett and J. P. Burman gave designs for the minimum possible size of experiment with p^n factorials and pointed out their utility in physical and industrial research. These designs have become so popular in industry and research because they enable five or more factors to be included simultaneously in an experiment of a feasible size. PB designs are two level designs which are useful for studying up to k = N - 1 variables, with N being the number of runs. Also note that N must be a multiple of four. For designs with N equaling 12, 20, 24, 28, and 36, the PB design is complicated with a very complex alias structure. In the 12 run experiment, every main effect is aliased with all two-factor interactions that do not involve it. For instance if k =11, then AB would be aliased with C, D, E,..., K [Cochran and Cox, 1957: 244].

In addition, each main effect is aliased with 45 two-factor interactions. The aliasing structure becomes even more convoluted in larger designs. The beauty of this seemingly confusing chain of events is that it allows the analyst to quickly discover which factors are important and which ones may be screened out of the design. This reduction in size does not come without a price tag. This price tag comes in the form of the aforementioned aliasing.

Treatment Combination	I	A	B	C	AB	AC	BC	ABC
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
b	+	-	-	+	+	-	-	+
abc	+	+	+	+	+	+	+	+

Table 2.4 2^{3-1} Half Fraction Design with Interaction

2.4.4 ***Aliasing.*** In order to see what happens when an experiment contains just part of a design, let's look at a 2^3 factorial design in which only the four treatment combinations *a, b, c,* and *abc* are examined. If we relate factors *a, b,* and *c* to the manufacturing experiment, then factor *a* could be the lathe, factor *b*, perhaps the drill press, *c* may be the sander and *abc* is the three way interaction. This, in effect is a half fractional experiment (2^{3-1}). Factorial designs allow multiple comparisons to be made to facilitate model creation, provide highly efficient estimates of model parameters, and usually involve simple calculations [Box and Draper 1987: 106]. Box and Draper also note that two level designs are especially useful in the exploratory stages of an investigation when little is known about the system and the model structure is relatively unknown.

Table 2.4 has a couple of interesting properties. When any other column multiplies the identity column, there is no change in the original column. For example if column I is multiplied by column AC (I * AC), the result is column AC. And the product of any two columns in the matrix will result in a third column in the matrix, such as:

$$B * ABC = AB^2C = AC \qquad (2.2)$$

This relationship shows that exponents are defined using modulus 2 arithmetic.

As you can see from Table 2.4, it is impossible to differentiate factor A from BC, B from AC, and C from AB. That is to say when we run our experiment we are unsure if the results are because of factor A or because of the BC interaction.

In fact when we estimate the main effects A, B, and C we are really estimating A + BC, B + AC, and C + AB. Two or more effects that have this property are called aliases, and they are said to be confounded with each other. In the manufacturing example A and BC are aliased, B and AC are aliased to each other and C and AB are aliased as well.

The aliasing structure for this design can be easily determined by using what is called a defining relationship. If we let I = ABC, then multiplying any column by our defining relationship (ABC)[1] will give the aliases for that effect. For example:

$$A * (I) = A * (ABC) = A^2BC \tag{2.3}$$

Remember any column multiplied by the identity (I) column is just the column and by definition of modulus 2 arithmetic, the square of any column is the identity so we are left with

$$A = I * BC = BC \tag{2.4}$$

$$A = BC \tag{2.5}$$

The alias structure for the main effects B and C may be found in the same manner.

2.4.5 ***De-aliasing.*** One way to avoid the aliasing effects is to perform a full factorial experiment if time and funding permit. As seen in Table 2.5, no two columns are identical to one another and therefore all factorial effects may be distinguished from each other.

[1] In general, the defining relationship for a fractional factorial is the set of all columns that are equal to the Identity (I) column.

Treatment Combination	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	-	+	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	+	-	+	-	-
bc	+	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+	+

Table 2.5 2^3 Full Factorial Design with Interaction

2.4.6 ***Response Surface Methodology.*** Response surface methodology (RSM) is a collection of statistical and mathematical techniques useful for developing, improving, and optimizing processes [Myers and Montgomery, 1995: 1]. RSM comprises a group of statistical techniques for empirical model building and model exploration. By careful design and analysis of experiments, it seeks to relate a *response*, or *output* variable to the level of a number of *predictors*, or *input* variables, that affect it [Box and Draper 1987: 1].

There are three typical uses for response surfaces methods.

1. To approximate the response of a system given a set of input parameters.

2. To assist in finding the particular input settings to produce a desired yield.

3. To give the settings that will produce the optimal yield for a system [Box and Draper, 1987: 17-19].

RSM has been used in a wide variety of studies to include, Captain Tim Smetek's thesis effort using experimental design and RSM to fit first order surface response equations to several measures of effectiveness using a simulation, which he designed, that modeled a real-world maintenance system [Smetek 1998: xi]. R. Garrison Harvey, Kenneth W. Bauer Jr., and Joseph R. Litko have also used RSM in military force allocation models where they used a nontraditional approach to optimizing a stochastic response surface subject to constraints [R. Garrison Harvey and others 1992: 1121]. Captain James L. Donovan took advantage of RSM and applied it to a macroeconomic model to facilitate better analysis with the model [Donovan 1985: viii]. These three citations are just a few of the ways that RSM has provided analysts insights to varying problems.

2.5 Linear Regression

Regression analysis is the statistical methodology that utilizes the relationship between two or more quantitative variables so that one variable can be predicted from the other, or others. This methodology is widely used in business, the social and behavioral sciences, the biological sciences, and many other disciplines [Neter and others 1996: 3]. What we wish to do, in simpler terms, is fit the output data from the real world system or a simulation to a function that is based on the inputs to the system. The regression model can also be used to determine the relative importance of each factor (including the interactions) that are in the original design matrix. As with all models, the objective of linear regression is to form a parsimonious model. The general form of a linear regression model with k regressor variables is

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon_i \tag{2.6}$$

The parameters β_j are the regressor coefficients, x_j are the regressor variables for $j = 0, 1, 2, ..., k$, and ε_i is the error term associated with the model for $i = 1, 2, ..., n$. β_j represents the expected change in the response variable y for every unit change in the regressor variable x_j while holding all other variables constant [Myers and Montgomery, 1995: 16-17].

2.6 Metamodels

Even though a simulation is not as detailed as a real world scenario, constructing a simulation can lead to a very labor intensive, time demanding, and complicated process of defining a relationship between variables that are input to a simulation and responses that may be seen in the real world. Sometimes a less complicated model may help bridge the gap between the real world and the simulation. This secondary model, called a metamodel will help bridge that gap and can be used to better understand and explore a simulation.

2.6.1 ***Constructing Metamodels.*** Several authors have pointed out the need for an analytic metamodel to aid in the interpretation of a more detailed model: Geoffrion was concerned with mathematical programming models; Blanning proposed the use of metamodels for all kinds of management science models; Lawless et al. made explicit use of metamodels for sensitivity analysis; Ignall et al. advocated that one take advantage of the potential benefits of both simulation and metamodels [Friedman and Friedman 1985:144].

One of the simplest models favored by some simulation researchers is the general linear model, or more commonly known as the linear regression model, whose general form you will recall as being displayed in equation 2.6. Using metamodels enables us to interpret a simulated system especially with regard to performing sensitivity analysis, by evaluating the effects of specific changes in our X variables on our response variable. They can also help to answer quick "what-if" and "back of the envelope" questions without having to rerun the simulation. Once the metamodel has been constructed and validated, further examination of the real system is less costly than conducting additional simulation runs. As stated, validation of the metamodel is a must. Remember, by using a metamodel, we are now two steps away from the actual system and therefore this metamodel must be a reasonable approximation.

Let us return once again to the manufacturing example. Suppose we want to develop a metamodel that represented the demand for several pieces waiting to be serviced at the lathe (LQ). Assume the factors that were found to be salient to the problem were arrival rate (ARR), service rate (SVC), and the number waiting in the service queue (NSVR). Our first inclination is to try and fit a linear model to the problem. There is a great temptation to fit a linear model as it makes our calculations easy. With great temptation comes the need for great care. Since we have replicated data (several samples of each of our three variables), linear regression techniques would give us a lack of fit test as well as a pure error term. A significant lack of fit term would be indicative of an inappropriate model selection (linear vs. quadratic or cubic) or the omission of one or more crucial variables in our equation. Assume that the true function is represented by the following equation:

$$LQ_i = \alpha \frac{(ARR)_i^{\beta_1}}{(SVC)_i^{\beta_2} * (NSVC)_i^{\beta_3}} v_i \tag{2.7}$$

for i = 1, 2, 3,..., n.

We may be able to make it into a linear function by performing a variable transformation. If we take the natural log of both sides we get:

$$\ln(LQ)_i = \ln(\alpha) + \beta_1 \ln(ARR)_i - \beta_2 \ln(SVC)_i - \beta_3 \ln(NSVR)_i + \ln(v)_i \tag{2.8}$$

By substituting variables we get the linear equation:

$$y_i = \beta_0 + \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} + \varepsilon_i \tag{2.9}$$

In doing this transformation of variables it is possible to turn a seemingly difficult equation into one that is tractable and thus use linear regression techniques to gain inferences to the real world system.

2.6.2 ***Metamodel Validation.*** There are two types of metamodel validation that will be discussed here. The first method requires 80% of your observations, selected at random, which are used to build the metamodel. The remaining 20% of the observations are used to test and validate the model. The R^2 value computed on the non-selected cases gives the analyst an indication of how the model performs on new data. A low R^2 for the un-selected cases indicates that our metamodel lacks predictive validity since it does not explain the variation in new data. Conversely, a high R^2 value shows that we have constructed an accurate metamodel [Friedman and Friedman, 1985: 145].

The R^2 of a model measures the proportionate reduction of total variation in Y associated with the use of the set of X variables $X_1, X_2, ..., X_{p-1}$ and can only take values ranging from $0 \leq R^2 \leq 1$ and is computed in the following manner [Neter and others 1996: 230].

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \tag{2.10}$$

$$\begin{aligned} SSR &= \text{Regression Sum of Squares} \\ SSE &= \text{Sum of Squares of Error} \\ SSTO &= \text{Total Sum of Squares} \end{aligned}$$

The second way to validate a metamodel is a technique called double cross-validation. This technique involves randomly dividing the data roughly in half and building a model out of both sets of data. After the models are complete, the data from the second model is used to validate the first and vice versa. By doing this, the analyst can compute two R^2 values for each portion of the data. One for the data used to create the model and one for the test data. These R^2 values are then compared. Again, if the R^2's are very dissimilar this would indicate that the metamodel is not valid. Even if the R^2 values are similar, the metamodel has to pass one more test. The coefficients of both metamodels should be examined. The coefficients of the metamodels should be of the same sign and magnitude. This indicates that the models are consistent and may be considered reliable.

2.7 Summary

Metamodels are an excellent tool for exploring a system without incurring the cost of additional simulation runs. Careful validation of the metamodel needs to take place in order to provide the decision maker with accurate insights. The analyst must also make sure that he/she stays within the design space of the experiment. Failure to do so could lead to inaccurate results.

Careful consideration should come to mind when attempting to optimize the metamodel. As stated before,optimization provides only an upper bound to the problem. This upper bound should be strived for in any analysis that you perform. However it is emphasized here that in the real world optimization rarely occurs. Chapter 3 discusses the methodology that implements the topics that have been reviewed here.

III. Methodology

3.1 Introduction

Chapter three explains the methodology used to meet the objective of devising a metamodel that will predict the cargo delivered from a notional TPFDD. The first section will be used to define the problem. This is the first step to the solution of any problem. Section two focuses on the statement of assumptions. Defining the variables will be the topic of the third section. The next topic will be to design an experiment to fully explore the region of interest. The fifth section will be used to determine if the necessary amount of simulation runs, for statistical purposes, have been accomplished . Data gathering and output analysis will be the topic of section six. Specifically, determining what output data is significant to the solution of the problem. Finally, additional insights will be discussed in the last section.

3.2 Definition of the Problem

No problem can be explored properly until it has been explained by the user and understood by the analyst. Several meetings with the sponsor may be needed. The purpose of these interactions are two-fold. First, it keeps the end user in the loop and it gives them a sense of quality control. The second purpose allows the analyst to fully understand the problem. With this in mind it is a good idea to put in writing what the user expects from the analyst and what the analyst will provide the user. AMC has asked that a metamodel be designed that will predict the amount of cargo that would be delivered from a TPFDD without re-running the AFM simulation model. The more direct question that AMC is concerned about in this effort is: Is there a linear connection between MOG and throughput?

3.3 The Model

AFM is owned by the HQ AMC Studies and Analysis Flight located at Scott AFB, IL. Although they are the primary users of the model, other groups such as AFOTEC, Air Force Studies and Analysis, and even AFIT have a copy. There are approximately 60,000 lines of code written in FORTRAN, C, and C++. Currently the model is under revision to make it compliant to the High Level Architecture that the Air Force is requiring of all its simulation models. Depending on the constraint set in the model, i.e. amount of fuel, MOG, cargo to be delivered, aircraft in the system, AFM can complete a 45 day scenario in as little as five minutes.

AFM simulates the AMC global airlift system and is capable of simulating AMC policies, procedures, operations, aircraft, air bases, cargo, passengers, and support resources as they relate to the airlift system. AFM simulates a fleet of aircraft moving a given amount of cargo and passengers from any number of on-load points, through any needed en-route stops, to any number of off-load points, then recovering and returning to home station for another mission. The model can continue this process for as many simulated days as desired, or until all requirements have been airlifted to their destination. Figure 3.1 illustrates these features on a macro scale.

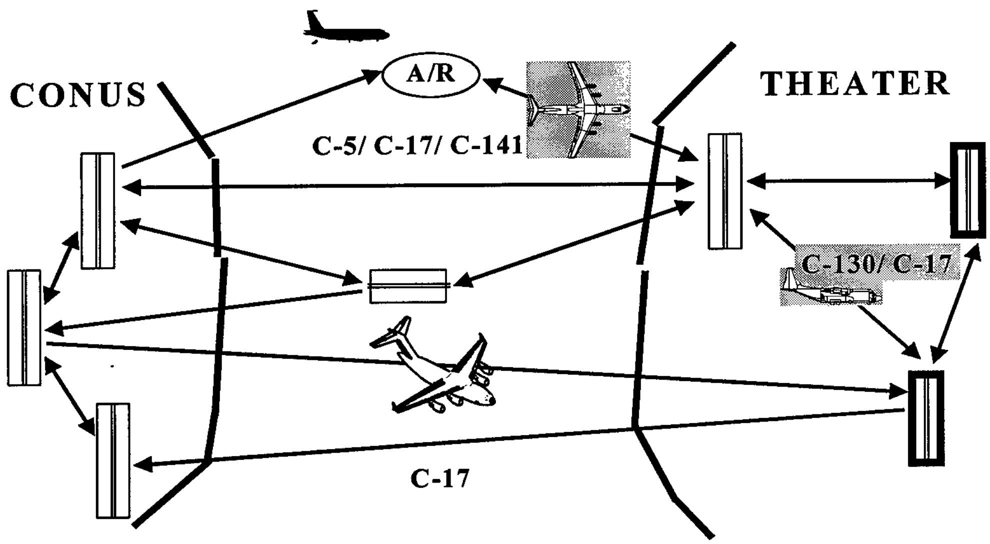

Figure 3.1 AFM Delivery Capabilities

On a micro scale, Figure 3.2 shows the series of steps that the model passes through for each aircraft.

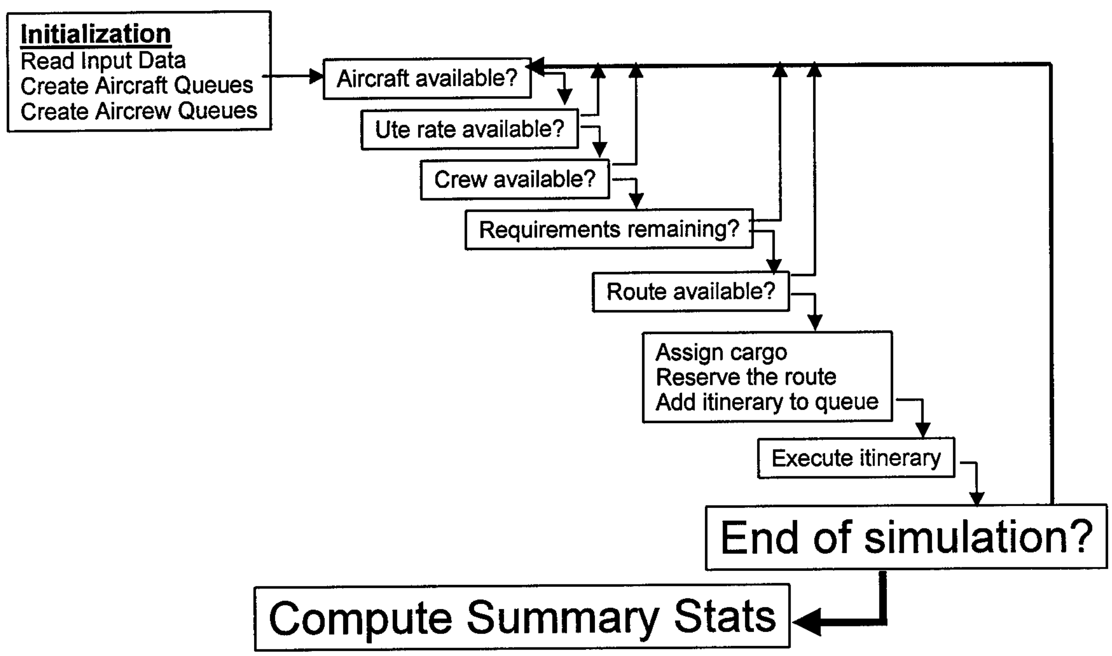

Figure 3.2 AFM High Level Block Diagram

3.4 Statement of Assumptions

There are several assumptions that can be made when running AFM. The assumptions that were made and the rationale behind them for this thesis effort are explained below.

1. It is assumed that a crew for every aircraft is ready and available. That is, the system (the world in this case) is starting empty and idle. This is not a crew ratio study, therefore the amount of crews should not be the limiting factor. One can argue that in the real world the number and availability of crews would be an issue, but with the other assumptions that are listed below, the reader should be in agreement that the assumption of crew readiness is valid. Furthermore, outside of doing crew ratio studies, AMC assumes full crew availability in planning day to day operations.

2. The second assumption made in this study was that at the start of the simulation all aircraft were ready and available for use. This is not how AMC does business. At any point in time most of AMC's air fleet is spread across the globe. Consequently they are not available for immediate use. When a contingency arises, a large portion of the fleet, that has been apportioned by the Joint Strategic Capabilities Plan (JSCP), are recalled to their home station base to receive any maintenance they may need. After servicing they are then released to perform their mission of transporting troops and cargo to the war fight. The rationale for assuming all aircraft are ready at the start of the simulation will be discussed in assumption three.

3. Assumption three shirttails on assumption two in the following manner. A standard AMC TPFDD has a ramp-up phase to account for aircraft being dispersed all over the globe. During this ramp-up phase in the early days of cargo movement in the TPFDD, the demand for cargo at the contingency area is considerably smaller. This reduction in required cargo delivery takes into

account AMC's assets being dispersed over the globe. The notional TPFDD that was used for this thesis effort had no such ramp-up schedule. The reason that a notional TPFDD was used is twofold. A TPFDD is a war plan and therefore classified. Also, with the assumption that all aircraft are readily available for cargo movement, the amount of cargo that is available to be delivered in the early days of the scenario was substantially greater than what is available to be moved in a "real-world" TPFDD.

4. The fourth assumption that was made was the removal of the C-141 from AMC's air fleet. This study was a look ahead as to what AMC's throughput capability would be after 2006. In 2006 the last of the C-141s will be retired from service. The C-141 will be replaced by the C-17.

5. NBC were not used in the study since the number that is allotted to AMC is negligible in moving material to the war fight.

6. This assumption centers around the KC-10s and KC-135s. Typically, they have a dual use role as air refuelers and strategic cargo airlift aircraft. In this scenario, they are to be used as air refueling vehicles only and not as cargo transports. Depending on the severity of the contingency, these aircraft would be needed more for ferrying fighters, bombers, and strategic airlift forces across the ocean(s) rather than using them as cargo carrying aircraft.

7. The last assumption that was made for this study was the use of the UHA. This notional aircraft is being proposed by the army as a way to help shorten the length of time it takes to get bombs and bullets into theater. Although sketchy at best, the capabilities of this aircraft are such that it is capable of carrying as much as ten C-5s.

3.5 Definition of Variables

There are many variables to consider in this experiment. Among them are: material handling equipment (MHE), MOG, fuel, and maintenance personnel to name a few. The only factor that was chosen to build a design of experiment around were the actual aircraft assets themselves. AMC has direct control over these assets and can use them as they see fit. MHE, MOG, fuel, and maintenance personnel are already in place or transported to where they need to be. Five sub-factors result when you break down the aircraft factor. These are the number of C-5, C-17, UHA, WBP, and WBC aircraft.

3.6 ***Design of Experiment***

The design and resolution will be dependent upon the number of significant variables that will be represented in the experiment. When the number of factors are small, then a full factorial experimental design may be accomplished. In a full factorial experiment the effects of all the different factors can be investigated simultaneously. There are several different types of designs that can be used. A Central Composite Design (CCD) was used in this experiment for the following reasons. The CCD is a widely used design of experiment because it allows for the fitting of a second order response surface. The CCD design involves the use of as many as five levels for each variable. These levels are represented by -1, 0, 1 in coded units, and the use of axial points. It is the axial points that allow for the estimation of pure quadratic terms. Two important features to note concerning the use of axial points are they lie only on the x^i axes where i = 1, 2,..., n for the number of factors to be represented. Secondly, in the axial portion of the design, the factors are not varying simultaneously but rather one factor at a time. As a result, no information concerning any factor interaction is obtained from this section of the design matrix.

3.7 ***Simulation Runs***

Now that the variables and the design of experiment have been determined, it is important to insure that the correct number of simulation runs will be accomplished. Statistical tests rely on a certain confidence level. For example, forming a 95%

confidence interval allows us to state that we are 95% certain that our response will lie within a given range. For the problem being studied in this thesis, we wish to estimate at the 95% confidence level that the total tons of cargo delivered in our 45 day time window will fall within our 2,500 ton threshold. To be sure that we reach that 95% level we must perform the correct number of simulation runs. Formula 12.28 from Banks, Carson, and Nelson's Discrete-Event System Simulation states the equation to determine the correct number of simulation runs as

$$R_f \geq \left(\frac{t_{(\frac{\alpha}{2},R-1)} * S_0}{\varepsilon} \right)^2 \tag{3.1}$$

where:

R_f = total number of runs needed to conduct statistical analysis

$t_{(\frac{\alpha}{2},R-1)}$ = test statistic

R = initial set of sample runs performed

α = confidence level

S_0 = Max standard deviation of the sample runs

ε = tolerance level of cargo delivered (arbitrarily set at 2,500 tons)

Initially, five runs of a 50 point central composite design matrix were accomplished in order to determine a sample standard deviation at each of the design points. From these 50 points, the maximum standard deviation of 3,200 tons was obtained and the calculations for R_f at the 95% confidence level are based on this value.

	$\mathbf{R_f}$	**5**	**6**	**7**	**8**	**9**
t-value	$t_{(.025,\ R-1)}$	2.776	2.571	2.447	2.365	2.306
# of runs needed	$\frac{t^2_{(.025,\ R-1)} S_0^2}{\varepsilon^2}$	12.626	10.830	9.810	9.164	8.712

Table 3.1 Number of Runs

Table 3.1 illustrates this stepwise iterative approach. When R_f is greater than the number of runs needed, we can be sure at whatever confidence level we choose that we have the required amount of runs for statistical purposes. In this example we see that the number of runs necessary for statistical purposes is nine. A total of ten replications were accomplished at each design point. Therefore, we can estimate at the 95% confidence level, the total tons delivered will fall within our 2,500 ton threshold.

3.8 ***Output Analysis***

The output from AFM is extremely detailed and readily available. It ranges from aircrew scheduling to utilization rates for aircraft. A mock scenario depicting the defense of Turkey was created for the AFM model to use, and the output that was of interest for this experiment was the total amount of cargo that could be delivered in a 45 day time window at each design point. Since the data was relatively well behaved, meaning there were no erratic data points, linear regression techniques were used to fit the data to a curve.

3.9 ***Additional Insights***

The purpose of performing these types of analyses (also referred to as sensitivity analysis) is to see what effects small changes to our input parameters (X^i) have on our response variable (Y). The sensitivity analysis that was conducted for this experiment was a simple one. The number of MOG spaces available were expanded at key bases by one third. Key bases are ones which have been determined to have a high volume of airlift traffic. The purpose behind this is that there is a lot of effort in AMC today to build up these key bases by increasing the available MOG for AMC use and the amount of fuel that is available to AMC. By increasing AMC's allotment of parking space, we are asking the simple question of can we do better with more? If by increasing the MOG at key bases we see no change in the amount of cargo delivered, further study may be warranted to expand the horizon for AMC to consider that the airlift problem may not be related to the repairs and construction that are slated to be completed in 2006. Rather, it may be an aircraft in the system problem.

3.10 ***Summary***

Determining a methodology is an important step in any experimental process. Having a methodology prevents an analyst from doing steps out of order or, even worse, omitting steps altogether. The road map that is contained in this methodology was used to conduct the analysis that is displayed in chapter four.

IV. Results and Additional Insights

4.1 Introduction

This chapter includes a summary of the results and sensitivity analysis from the Airlift Flow Model simulation. The experiment was conducted as it was outlined in Chapter 3. Analysis of the output data was used to formulate a metamodel to avert the need for additional simulation runs of the model. This metamodel would be used to predict cargo delivery given changes to the input parameters. The analysis was conducted using linear regression techniques.

4.2 Assumptions

4.2.1 ***Mandatory.*** In order to perform linear regression analysis on the output data, there are three mandatory input data criteria that must be satisfied. These are:

1. The data needs to be independent and identically distributed
2. Normally distributed error terms
3. Constant variance

Violation of any of these three criteria might prevent us from performing an analysis of variance test (ANOVA) on the output.

4.2.2 ***Necessary.*** These assumptions were made on both the input and output data. First is the assumption that reliable data was input into the simulation. This means no "fat fingering" of the data or other operator error confounded the data. Making sure that the output data is gathered and stored in chronological order is the second assumption. Finally, removing, at random, 20% of the output data to use as test data does not drastically affect the resulting metamodel.

4.3 *Preliminary Analysis*

This step is used to verify the appropriateness of the data. Three steps need to be accomplished in this preliminary analysis. First the correct variables to use in the analysis need to be identified. This is accomplished by performing a linear regression fit of the output data and verifying the impact of each variable to the model by checking its associated p-value and variance inflation factor (VIF). A VIF is used to detect the presence of multicollinearity. Multicollinearity exists when predictor variables are correlated among themselves. Having a VIF greater than 10 suggests high multicollinearity [Neter and others, 1996: 387].

The the full model analysis, the first attempt was to fit a model that included up to four way factor interactions. This resulted in a model with an R^2 of 0.98648. Although this is an excellent fit of our data, four way interactions are difficult to envision and hard to explain. In addition, the fewer the amount of terms in the model, the wider the range of data that may be fit to it. When reducing the number of terms in the model, care must be taken to ensure that we keep enough to fully explain the data. By reducing the previous model to include only two way interactions, we showed a decrease in the R^2 of only 0.008462. This negligible decrease in the R^2 value provides us with enough security to reduce the number of terms in the model. The model and ANOVA table displayed in tables 4.1 and 4.3 show the resulting values of the two factor regression.

Term	Estimate	t Ratio	Prob>|t|	VIF
Intercept	279,823.430	958.72	0	0
C-5	5,906.926	47.30	<.0001	1.007384
C-17	9,289.214	74.36	<.0001	1.007858
C-5*C-17	-2,436.464	-16.59	<.0001	1.013022
UHA	5,558.726	44.19	<.0001	1.005743
C-5*UHA	-1,709.033	-11.64	<.0001	1.011891
C-17*UHA	-1,771.266	-12.07	<.0001	1.011040
WBP	2,036.573	16.02	<.0001	1.010112
C-5*WBP	-385.505	-2.62	0.0091	1.015127
C-17*WBP	-1,680.886	-11.42	<.0001	1.017345
UHA*WBP	667.449	4.54	<.0001	1.011911
WBC	7,951.317	63.04	<.0001	1.010229
C-5*WBC	-2,507.189	-17.13	<.0001	1.005342
C-17*WBC	-2,376.538	-16.23	<.0001	1.007542
UHA*WBC	-2,125.416	-14.52	<.0001	1.005477
WBP*WBC	-395.740	-2.70	0.0073	1.010827
$C\text{-}5^2$	-1,523.227	-14.00	<.0001	1.063238
$C\text{-}17^2$	-2,341.755	-21.53	<.0001	1.063164
UHA^2	-2,156.767	-19.44	<.0001	1.056327
WBP^2	-1,106.649	-9.76	<.0001	1.050255
WBC^2	-3,275.149	-29.47	<.0001	1.059731

Table 4.1 Full Model Parameter Estimates, P-values, and VIF

From the adjusted R^2 value obtained in table 4.3, we see that we have an excellent fit for our model. The larger the R^2 is, the more the total variation of Y is reduced by introducing the predictor variable X into the model. The adjusted R^2 modifies the R^2 by dividing each sum of squares by its associated degrees of freedom. The adjusted R^2 may actually become smaller when another X variable is introduced into the model. It is also noted to the reader that all VIF values were well below 10 indicating a lack of multicollinearity.

	-2.37841	-1	0	1	2.37841
C-5	44	65	80	95	116
C-17	44	65	80	95	116
UHA	1	12	20	28	39
WBP	44	65	80	95	116
WBC	29	50	65	80	101

Table 4.2 Coded Variable Conversion Chart

Table 4.2 shows the actual number of aircraft that would be used in the regression equation. For example if the experiment was set up to analyze a center run, all factors would be at the coded level of zero, then 80 C-5s, 80 C-17s, 20 UHAs, 80 WBP, and 65 WBC aircraft would be used in the scenario.

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	20	9.19E+10	4.60E+09	843.1076
Error	379	2,065,992,281	5,451,167	Prob>F
C Total	399	9.40E+10		<.0001
RSquare	0.978018			
RSquare Adj	0.976858			
Root Mean Square Error	2,334.77			
Mean of Response	270,865.60			
Observations (or Sum Wgts)	400			

Table 4.3 Full Model Regression ANOVA

The ANOVA of Table 4.3 provides an analytical view of the regression model. Note that the model possesses a large R^2 and Adj R^2 indicating that the model does have predictive capability and warrants further investigation of its validity.

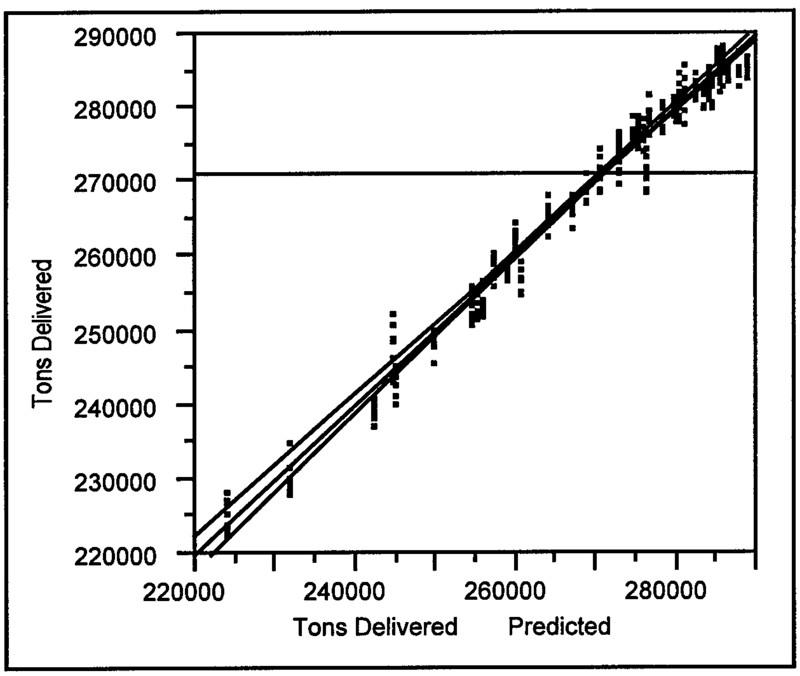

Figure 4.1 Full Model Least Squares Regression Line Plot

Figure 4.1 shows the regression plot for the data. The solid center line, called the least squares regression line, represents the predicted value of the model. The two solid outer lines show the confidence bands surrounding the predicted line. The points show the actual value that was obtained by the simulation. The difference between the actual value and the predicted value of the model is the error term. It can be seen how well the data points fit the least squares regression line indicating a small amount of error. This reinforces the assumption that we seem to have a proper fitting model. Now we must check for the presence of outliers and influential data points. To check for outliers, we plot the studentized residuals versus run time. If a point lies outside of plus or minus three standard deviations (four for large groups of sample data), then we would investigate that point for possible outlier consideration [Neter and others 1996: 103]. Another test for outliers is to check each individual studentized deleted residual. If any of the absolute values of the studentized deleted residuals are greater than the t-distribution [Neter and others 1996: 374]

$$t(1 - \frac{\alpha}{2n}; n - p - 1) \tag{4.1}$$

where:

α = the associated level of confidence with the test

p = the number of parameters in the model

n = the total number of samples

or numerically, t(0.9988,379) which lends a value of 3.266 then they are classified as an outlier.

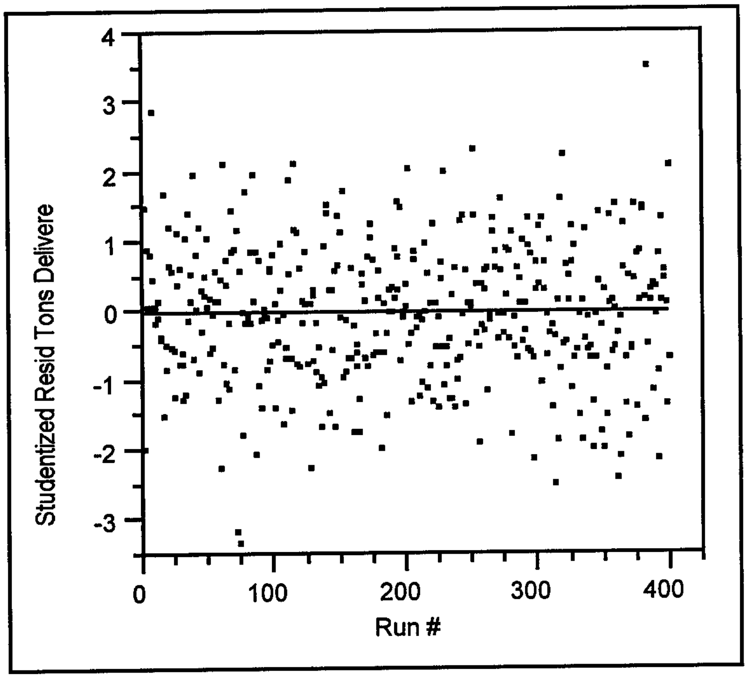

Figure 4.2 Full Model Outlier Plot

Figure 4.2 above shows the studentized residuals plotted against the run time. We see that only one out of the four hundred data points is close to being an outlier. Outliers can create great difficulty when using least squares analysis. Outliers can cause a fitted line to be pulled disproportionately towards them in order to minimize the sum of squared deviations. The presence of outliers can be attributed to several things such as inaccurate readings, bad sampling techniques, or bad input data. Removal of the point is unwarranted since the data collection process was automated. This provides data integrity. In addition, as stated before, for large data sets (number of samples greater than 50) a point needs to be outside of four standard deviations to warrant further outlier testing. Figure 4.2 is also an indication of the independence of the residuals. It is necessary to have independence in the residuals in order to continue using linear regression techniques. The lack of any patterns in the dispersal of the residuals is an indication of independence.

4.4 Full Model Regression Analysis

4.4.1 ***Normality.*** The first step in regression analysis is to check the residuals for normality. Residuals result from the difference in the predicted value versus the actual value. A lack of normality would suggest that a transformation of the response is necessary. To check for normality in the raw residuals, we plot the residuals and perform a Shapiro-Wilk test [Sall and Lehman 1996: 112]. In Figure 4.3, a visual inspection of the residual plot and use of the Shapiro-Wilk test confirms that the residuals are normally distributed. The obtained p-value of 0.9891 allows us to fail to reject H_0 at the 0.05 level.

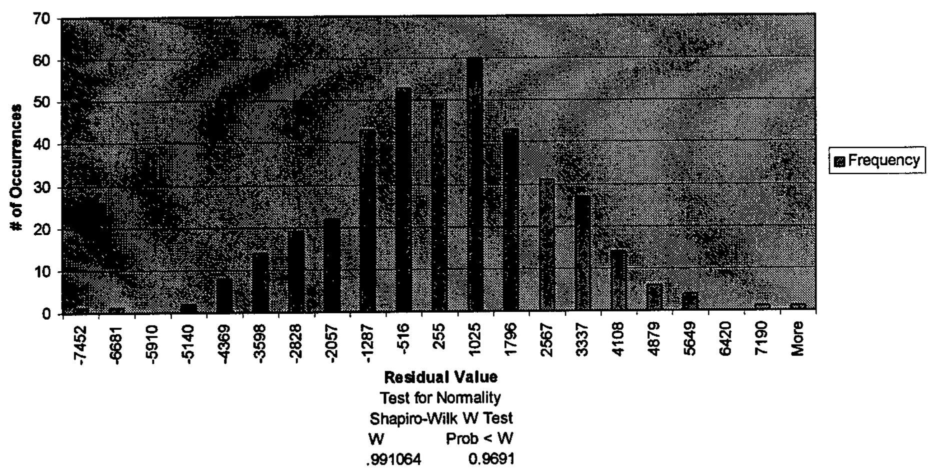

Figure 4.3 Full Model Residual Normality Plot

4.4.2 ***Constant Variance.*** Step two in regression analysis is to check the residuals for constant variance. Figure 4.4 shows a plot of the residuals for the regression model. What is meant by constant variance? In a visual inspection of the raw residual plot, we are looking for evenly spaced and distributed residuals. Patterns such as megaphoning, football, or any other type of trends are undesirable and indicate the need for possible variance-stabilizing transformations.

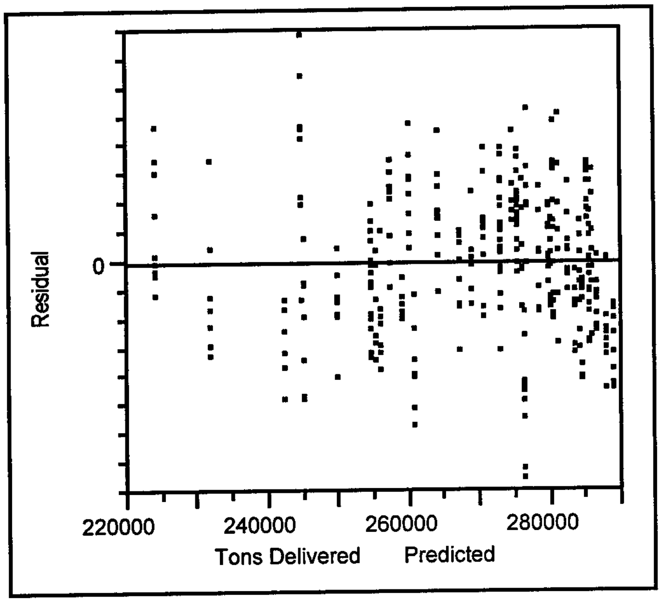

Figure 4.4 Full Model Residual Plot

As can be seen in Figure 4.4 there doesn't appear to be any patterns present in the residuals.

To quantitatively test for constant variance in the residuals we can use the Breusch-Pagan Test [Neter and others, 1996: 115]. This is a large sample test which assumes that the error terms are independent and normally distributed. The test statistic is denoted:

$$\chi^2_{BP} = \frac{SSR^*}{2} \div \left(\frac{SSE}{n}\right) \tag{4.2}$$

where:

SSR^* = regression sum of squares when regressing e^2 on X_i

SSE = error sum of squares when regressing Y on X_i

n = number of parameters in the model

To conduct the Breusch-Pagan test at the 95% confidence level with 20 degrees of freedom for the data that was obtained in this experiment, we refer back to table 4.3 for our SSE, and n values. We use JMP to determine the SSR^* value. We see that the test statistic is equal to:

$$\chi^2_{BP} = \frac{6.4971\text{x}10^{15}}{2} \div \left(\frac{2,065,992,281}{20}\right)^2 = 0.3044$$

Looking up the table value for the χ^2 at the 95% confidence level and 20 degrees of freedom we obtain the value of 31.41. Since the test statistic value is much smaller than the table lookup, we fail to reject the null hypothesis of constant variance at the 95% confidence level.

4.4.3 ***Lack of Fit Test.*** This test is used to determine whether the model adequately fits the data. The assumptions that the lack of fit makes is that the response(s) are (1) independent, (2) normally distributed, and (3) the distributions of Y have the same variance σ^2 [Neter and others 1996: 116]. In order to perform the lack of fit test, multiple observations of at least one design point are required.

Referring back to the model in Table 4.3 we see that the P-value is <0.0001. This is a clear indication that we reject the null hypothesis H_0, (H_0 being we have a linear model), in favor of the alternative, H_a, (the model as represented poorly fits the data). Why does the model reject the null hypothesis? Ten replications of a 50 point design of experiment were performed since multiple replications are necessary to achieve a lack of fit test. This resulted in the regression line having to pass through 50 different means. The likelihood of this happening to the point where it statistically passes the lack of fit test is a minimum. Even though the model fails lack of fit, with the R^2 that is produced, we can still have confidence in the model.

4.5 *Metamodel Verification*

Now that we have constructed the metamodel, we need to fit the 20% of the data that has been separated from it. The reason that we do this is to make sure the model fits an independent set of data. Table 4.4 shows how the regression model of Table 4.1 fits the test data. Note that Table 4.4 displays a combination of the information that is contained in Tables 4.1 and 4.3.

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	20	2.36E+10	1.18E+09	239.3199
Error	79	389,477,838	4,930,099	Prob>F
C Total	99	2.40E+10		<.0001
RSquare	0.983763			
RSquare Adj	0.979652			
Root Mean Square Error	2,220.383			
Mean of Response	270,576.700			
Observations (or Sum Wgts)	100			
Parameter Estimates				
Term	Estimate	t Ratio	Prob>|t|	VIF
Intercept	279,996.590	523.12	<.0001	0
C-5	6,340.599	23.68	<.0001	1.175805
C-17	8,838.736	32.85	<.0001	1.184609
C-5*C-17	-2,545.233	-8.31	<.0001	1.218404
UHA	5,888.283	23.14	<.0001	1.135253
C-5*UHA	-1,735.199	-5.54	<.0001	1.273288
C-17*UHA	-2,447.812	-7.86	<.0001	1.257180
WBP	1,809.686	7.19	<.0001	1.172543
C-5*WBP	-397.137	-1.23	0.2235	1.338952
C-17*WBP	-1,360.797	-4.17	<.0001	1.368292
UHA*WBP	384.075	1.22	0.2262	1.258032
WBC	7,901.784	29.90	<.0001	1.203806
C-5*WBC	-2,139.790	-7.31	<.0001	1.106506
C-17*WBC	-2,984.041	-10.09	<.0001	1.136479
UHA*WBC	-2,039.424	-6.92	<.0001	1.124256
WBP*WBC	-864.468	-2.77	0.007	1.254359
C-5^2	-1,541.121	-6.60	<.0001	1.045001
C-17^2	-1,928.224	-8.26	<.0001	1.044537
UHA^2	-1,963.494	-9.33	<.0001	1.050706
WBP^2	-1,169.985	-6.00	<.0001	1.071581
WBC^2	-3,893.931	-17.90	<.0001	1.122838

Table 4.4 Test Data Verification

By inspecting the R^2 and adj. R^2 from the table above, we see that we have a good fit of the test data using the regression model of Table 4.1. The coefficients of the test data are of the same magnitude and sign as Table 4.1. Taking note of the p-values for the test data regression model, it is observed that two of the coefficients may be insignificant to the model due to their high p-value. However, since we wish to fit the test data to the metamodel that has already been established, we shall leave these two terms in the model.

4.6 *Metamodel Validation*

In order to test the predictive validity of the metamodel, another regression model was developed using the twenty percent of the observations that were selected at random. The test model was created using the same parameters the metamodel used. The R^2 value computed on the test data gives an indication of how well the model performs on new data. A very low R^2 for the test data regression model would indicate that the model lacks predictive validity since it does not sufficiently explain the variation in the test data set. For the data that was used for this project, the R^2 for both the metamodel and validation set were very close. The R^2 for the metamodel data was 0.9867, while the test set achieved an R^2 of 0.9898. This attests to the fact that the metamodel does possess predictive capability.

Another way to test the validity of the model is to plug the design matrix values into the regression equation to determine what the predicted value would be at that design point. Then, by using the following equation,

$$\text{Absolute Error} = \frac{|\text{Simulation - Metamodel}|}{\text{Simulation}} \tag{4.3}$$

a mean absolute percent error (MAPE) can be established for the system [Friedman and Friedman 1985: 146]. Table 4.5 contains the data for the absolute error for each simulation run.

Run	Simulation	Metamodel	Deviation	Absolute % Error	Run	Simulation	Metamodel	Deviation	Absolute % Error
1	281,405	279,821	1,584	0.562957	51	239,838	242,357	-2,519	1.050244
2	279,399	284,348	-4,949	1.771287	52	245,899	244,686	1,213	0.493450
3	277,599	276,416	1,183	0.426230	53	247,697	250,361	-2,664	1.075541
4	281,267	279,821	1,446	0.514170	54	244,889	244,686	203	0.083053
5	277,726	278,123	-397	0.142948	55	268,214	268,694	-480	0.178905
6	281,050	280,202	848	0.301801	56	262,231	258,131	4,100	1.563576
7	286,384	284,764	1,620	0.565642	57	278,560	279,821	-1,261	0.452617
8	271,747	270,832	915	0.336569	58	231,534	242,357	-10,823	4.674426
9	284,261	282,433	1,828	0.643125	59	279,766	279,821	-55	0.019591
10	278,560	279,821	-1,261	0.452617	60	225,962	224,424	1,538	0.680537
11	281,267	279,821	1,446	0.514170	61	277,164	276,145	1,019	0.367538
12	282,866	283,220	-354	0.125165	62	279,766	279,821	-55	0.019591
13	285,712	285,320	392	0.137066	63	278,560	279,821	-1,261	0.452617
14	286,567	285,926	641	0.223607	64	280,168	278,123	2,045	0.729918
15	248,877	244,541	4,336	1.742178	65	275,399	274,454	945	0.343118
16	279,155	276,416	2,739	0.981250	66	283,965	287,195	-3,230	1.137375
17	285,942	286,671	-729	0.254924	67	252,356	254,252	-1,896	0.751194
18	241,889	242,357	-468	0.193429	68	274,983	276,145	-1,162	0.422687
19	286,004	284,348	1,656	0.579028	69	269,399	268,694	705	0.261750
20	269,480	267,419	2,061	0.764684	70	253,224	256,118	-2,894	1.142924
21	279,134	276,255	2,879	1.031526	71	231,069	229,856	1,213	0.524770
22	286,495	284,764	1,731	0.604167	72	281,758	279,201	2,557	0.907386
23	278,736	276,416	2,320	0.832404	73	280,591	282,433	-1,842	0.656417
24	282,407	280,963	1,444	0.511418	74	276,750	274,454	2,296	0.829609
25	229,654	229,856	-202	0.088141	75	275,610	274,454	1,156	0.419413
26	252,531	255,569	-3,038	1.203093	76	287,102	288,590	-1,488	0.518412
27	258,249	259,669	-1,420	0.549968	77	249,088	250,361	-1,273	0.511098
28	276,452	276,416	36	0.013098	78	256,350	254,252	2,098	0.818536
29	275,810	274,484	1,326	0.480618	79	275,042	276,145	-1,103	0.401145
30	278,167	278,398	-231	0.083058	80	253,889	254,424	-535	0.210818
31	278,370	279,821	-1,451	0.521180	81	282,852	284,510	-1,658	0.586252
32	269,363	268,694	669	0.248420	82	280,619	279,821	798	0.284439
33	230,629	229,856	773	0.334989	83	284,466	280,963	3,503	1.231528
34	280,040	284,510	-4,470	1.596281	84	256,996	255,569	1,427	0.555190
35	246,500	250,361	-3,861	1.566363	85	259,482	263,292	-3,810	1.468480
36	278,331	274,484	3,847	1.382021	86	283,784	287,195	-3,411	1.201881
37	250,287	244,541	5,746	2.295717	87	281,395	283,220	-1,825	0.648572
38	277,085	274,484	2,601	0.938554	88	257,408	258,131	-723	0.280807
39	277,465	278,398	-933	0.336273	89	281,734	284,348	-2,614	0.927810
40	286,460	285,320	1,140	0.397827	90	278,192	275,057	3,135	1.126813
41	279,250	274,484	4,766	1.706569	91	278,560	279,821	-1,261	0.452617
42	281,405	279,821	1,584	0.562957	92	284,333	287,195	-2,862	1.006477
43	277,755	278,123	-368	0.132492	93	281,267	279,821	1,446	0.514170
44	254,934	254,424	510	0.199956	94	279,855	279,821	34	0.012217
45	279,279	278,123	1,156	0.413922	95	278,705	279,821	-1,116	0.400355
46	282,465	282,433	32	0.011383	96	260,856	263,292	-2,436	0.934018
47	283,367	287,330	-3,963	1.398512	97	276,520	278,123	-1,603	0.579706
48	258,170	259,669	-1,499	0.580737	98	256,766	257,072	-306	0.119315
49	258,675	258,131	544	0.210372	99	280,619	279,821	798	0.284439
50	281,405	279,821	1,584	0.562957	100	260,415	258,131	2,284	0.877131

Table 4.5 Mean Absolute Percent Error of Validation Data

To calculate the overall MAPE for the system, simply sum up all of the absolute percent errors and divide by the total number of samples in the system, in this case one hundred. The MAPE for this system is 0.6756%, thus providing a second check that the metamodel appears to be a reliable and valid approximation.

4.7 Variable Screening

The comparison of R^2 for both models indicates that we have a good model. But, have we really constructed the correct model for the problem? Looking back to table 4.1, we see that our model uses two way factor interaction. All combinations of those two way interactions were kept in the model and their associated p-values are listed as well. Upon further inspection of these p-values, we see that they are all below our tolerance level of 0.05. Looking deeper than that, we have an extremely high R^2 for a model that uses two factor interactions. This is a clear indication that all the variables in Table 4.1 are pertinent to the model. But we must keep in mind that we are not looking for a metamodel for this specific problem. We are looking for a parsimonious model, one that contains the fewest amount of variables possible, yet is complete and robust enough to provide an accurate prediction.

4.8 Full Model Summary

The model as represented, gives a fairly accurate numerical description of the data. With the large R^2 value that we have obtained, we are fairly certain, at least 95% certain, that the model will predict cargo throughput as long as we experiment within the given design space.

4.9 Reduced Model Analysis

By removing terms that are close to being insignificant we can perform a full versus reduced test. Even if the p-value of the term warrants the term to be kept in the model, this test will determine if there are any additional terms that can be removed without degradation of the model. The null hypothesis, H_0, for this test is $\beta_i = 0$ meaning that the i^{th} term is in fact insignificant to the regression model. The alternative, H_a, is that $\beta_i \neq 0$, or the i^{th} term is significant to the model. The test statistic for this is [Neter and others, 1996: 80]:

$$F^* = \frac{SSE_R - SSE_F}{df_R - df_F} \div \frac{SSE_F}{df_F} \tag{4.4}$$

The decision rule for this test is:

$$\begin{aligned} If\ F^* &\leq F(1-\alpha; df_R - df_F, df_F), conclude\ H_0 \\ If\ F^* &> F(1-\alpha; df_R - df_F, df_F), conclude\ H_a \end{aligned}$$

where:

SSE_R = Sum of Squares of Error in the reduced model

SSE_F = Sum of Squares of Error in the full model

df_R = degrees of freedom in the reduced model

df_F = degrees of freedom in the full model

Several variables were removed from the model one at a time according to their absolute t-value from table 4.1 in order to create a reduced model. The absolute t-value is indicative of a term's contribution to the model. The smaller the t-value, the less the contribution. Reduced models were in turn tested against the full model to determine if individual terms may be excluded from the model. For example by removing the WBP*WBC interaction term, the full versus reduced model comparison is illustrated below.

$$F^* = \frac{2,105,630,252 - 2,065,992,281}{380 - 379} \div \frac{2,065,992,281}{379} = 7.2714$$

We have an F statistic with the following parameters F(0.05, 1, 379) which equates to a value of 3.8661. Our F* value of 7.2714 is greater than our F statistic of 3.8661 therefore we conclude the alternative. As it turns out, $F^* > F(0.05,1,379)$ in all cases, concluding at the 95% level that all terms are significant to the model.

4.10 Stationary Points

Now that we have obtained a viable second order model, we can use RSM to see how close the model comes to the optimal solution to our problem. We begin by finding the stationary point of our design space. If we could minimize and physically transport ourselves onto the design space, the stationary point would represent where we would be actually standing on the design space. There are three cases that can happen when we find the stationary point on a surface.

1. We can be located at a local (or global) minimum where all values of x_s are positive

2. We can be located at a local (or global) maximum where all values of x_s are negative

3. We can be located at a saddle point the signs of x_s are mixed (both positive and negative)

If we were at a min point, stepping in any direction (remember we are on the surface of the design space) would improve our solution. Conversely, at a max point we would decrease our value if we stepped in any direction. And finally if our stationary point is a saddle point, then we could either increase or reduce or value depending on which way we stepped.

In order to find the stationary point (and in doing so we will discover the nature of the system at that point, either min, max, or saddle) we use the following formula [Myers and Montgomery, 1995: 218]:

$$x_s = -\hat{\mathbf{B}}^{-1} * \mathbf{b}/\mathbf{2} \tag{4.5}$$

where $\mathbf{b}$ and $\hat{\mathbf{B}}$ contain estimates of the linear and second order coefficients respectively. In matrix format, they are represented as thus:

$$\hat{\mathbf{B}} = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \cdots & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{12} & b_{22} & \cdots & \frac{1}{2}b_{2k} \\ . & . & \cdots & . \\ \frac{1}{2}b_{1k} & \frac{1}{2}b_{2k} & \cdots & b_{kk} \end{bmatrix} \tag{4.6}$$

and

$$\mathbf{b} = [(b_1, b_2, \ldots, b_k)]^T \tag{4.7}$$

By applying equation 4.5 to the regression model of Table 4.1, we obtain the following values for the stationary point.

$$x_s = [84.634, 90.219, 27.230, 103.651, 71.901]^T$$

Inspecting these values we see that we are at a local min because all of the signs in the vector are positive. A local min tells us that if we step in a certain direction we improve our solution and going the opposite way decreases our solution. But inspecting the vector in more detail, we observe the magnitude of the values. We note that the third value is much smaller than any of the other values. This indicates that we may be standing on or close to a ridge in the design space. A technique known as ridge analysis would be beneficial at this point. Ridge analysis is the steepest ascent method applied to a second order model. The intent of steepest ascent method is to provide a path to an improved solution to a system that has not been well studied. Steepest ascent is generally a low-cost first order experiment. Ridge analysis is used when the analyst feels confident that he or she is close to the region where the contains the optimal solution. However, due to the fact that no further data is available ridge analysis computations are not able to be exercised here.

Another way to "see" where we are on the design space is to create a response surface. A response surface allows us to represent the curvature of the response and affords us with a visual image of what we would see if we were minimized and placed on the surface. Out of the 400 data points that are represented in the metamodel, the design point that produced the maximum amount of cargo was selected to be representative of the system. A response was built about this point. The stationary point that is represented above could not be mapped to this response since it is outside of our design space, and hence of no predictive value. It only serves to inform us that a new experiment needs to be conducted at that point. Since it is impossible to represent more than a 2D image on paper, the coded value for the UHAs, WBPs, and WBC aircraft were held constant. These constant values that were used were for those three aircraft types were the settings that were used at the maximum response. Only the C-5s and C-17s were allowed range from the low axial value through their high axial value.

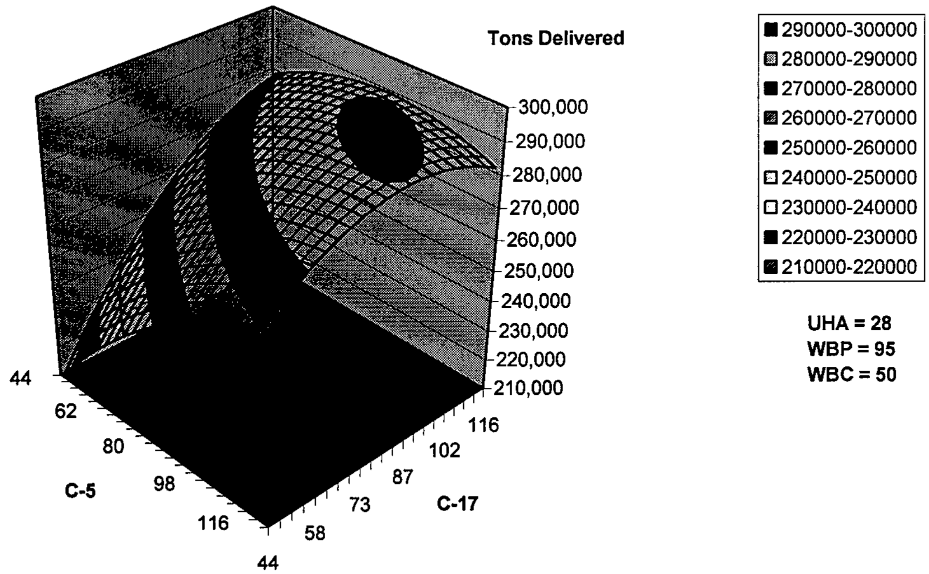

Figure 4.5 Maximum Cargo Throughput Response

Figures 4.5 and 4.6 show the response. Figure 4.5 shows how quickly the response rises and levels off and has a large region of maximum throughput. It also has a rather large region where throughput, although not optimal, would be a very acceptable delivery region.

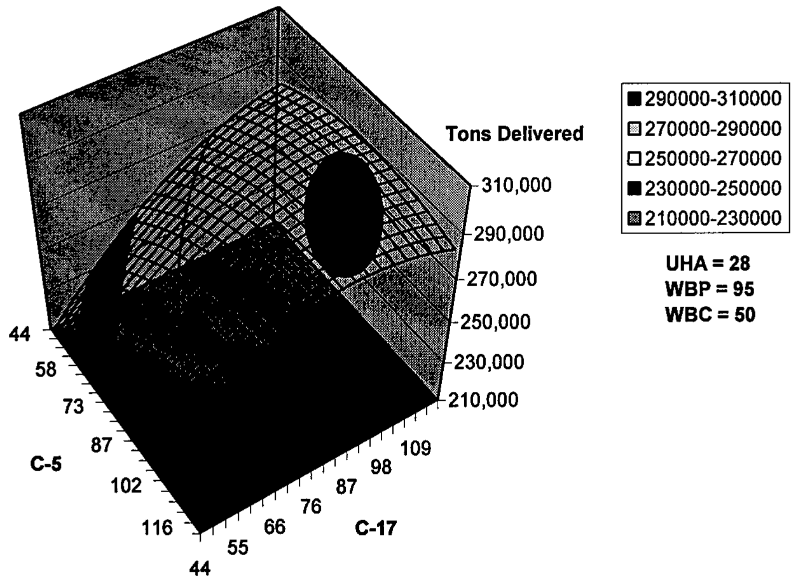

Figure 4.6 Maximum Cargo Throughput Response

Figure 4.6 displays this large region in greater detail. Any mix of C-5s, C-17s, and the remaining three aircraft types set at their uncoded values that keep us in this region provide a high throughput of cargo. But is this throughput optimal?

4.11 Optimization

Optimization is more of a pipe dream than a reality. When the best laid plans are put to the test, the slightest deviation now makes you sub-optimal. However, optimization techniques do provide a good estimate of the upper bound (or lower bound depending on how the problem is constructed). Using Lingo to solve the problem, Figure 4.7 shows the constructed integer program (IP) and its results in uncoded units.

A = C-5, B = C-17, C = UHA, D = WBP, E = WBC

```
Max = -387457.554 - 6.77*A*A - 10.408*B*B -33.699*C*C - 4.18*D*D - 14.556*E*E
  + 3489.483*A + 4730.242*B + 5069.32*C + 1660.522*D + 4653.776*E
  - 10.829*A*B - 14.242*A*C - 1.713*A*D - 11.143*A*E - 14.761*B*C
  - 7.471*B*D - 10.562*B*E + 5.562*C*D - 17.712*C*E - 1.759*D*E;


  -387457.554 - 6.77*A*A - 10.408*B*B -33.699*C*C - 4.18*D*D - 14.556*E*E
  + 3489.483*A + 4730.242*B + 5069.32*C + 1660.522*D + 4653.776*E
  - 10.829*A*B - 14.242*A*C - 1.713*A*D - 11.143*A*E - 14.761*B*C
  - 7.471*B*D - 10.562*B*E + 5.562*C*D - 17.712*C*E - 1.759*D*E > 288902;

!! AIRCRAFT CONSTRAINT
 A + B + C + D + E > 0;
 A + B + C + D + E < 488;

!! WIDE BODY AIRCRAFT CONSTRAINT
 A > 2*E;
 B > 2*E;
 A > B;
 D > 75;
 D < 83;

@GIN(A);
@GIN(B);
@GIN(C);
@GIN(D);
@GIN(E);
```

Variable	IP Value	SIM Value
C-5	104	95
C-17	104	95
UHA	24	28
WBP	83	95
WBC	52	50
TONS	290722	288902

Figure 4.7 Cargo Throughput Optimization IP

This is just a rough sketch at improving cargo delivery. Other factors need to be considered when constructing the IP. For instance, all aircraft were given equal weight. Not in cargo capacity but aircraft desirability for cargo. An example of this follows, WBC and C-5s both carry large amounts of cargo, but not the same type. Weighting functions and perhaps goal programming methods would need to be implemented in order to get an accurate aircraft mix. Passenger aircraft would also need to be converted to tons carried rather than passengers and weighted in the same manner.

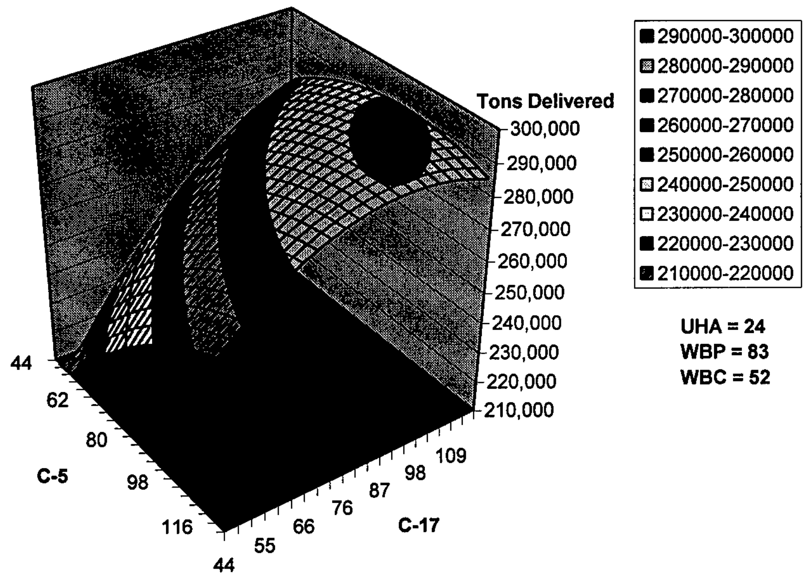

Figure 4.8 Optimal Throughput Response

Inspection of the response in Figure 4.8 shows that we have not increased our region of maximum throughput a substantial amount by implementing the results from the IP. It is stressed again that the solution represented by the IP may not be optimal due to certain constraints being left out. The actual optimal region could be larger or smaller.

4.12 Additional Insights

A separate set of data was collected using the same design matrix as in the previous experiment. Again, as in the previous experiment, the same amount of samples were collected at each design point for a total of 500 samples. The only difference between the two experiments is that the MOG values at crucial bases for the second experiment were increased. This increase will be noted as IMS (increased

MOG scenario). Crucial bases were determined by examining traffic patterns from the output of the first experiment. MOG was increased by 20% for those bases having a high flow of aircraft. Fuel was not increased at any of the bases.

4.13 Full Model Regression Analysis IMS

Table 4.6 represents the final regression model of this data. The insignificant terms have already been removed from the model. This was accomplished by removing variables with large p-values, then performing a full versus reduced test on the remaining ones. Again, note that our VIF factors are well below 10, indicating a lack of multicollinearity between predictor variables.

Term	Estimate	t Ratio	Prob>|t|	VIF
Intercept	281,980.510	500.68	0	0
C-5	7,287.891	29.62	<.0001	1.0096698
C-17	3,715.291	14.99	<.0001	1.0068280
UHA	6,047.431	24.61	<.0001	1.0070985
C-5*UHA	-1,992.959	-6.88	<.0001	1.0129846
C-17*UHA	1,452.773	5.03	<.0001	1.0091683
WBP	7,944.796	32.06	<.0001	1.0073892
C-5*WBP	-2,045.934	-7.06	<.0001	1.0152550
C-17*WBP	-1,455.306	-5.03	<.0001	1.0099856
UHA*WBP	-1,633.573	-5.66	<.0001	1.0054868
WBC	6,948.790	27.87	<.0001	1.0204690
C-5*WBC	-1,792.193	-6.18	<.0001	1.0120150
UHA*WBC	-1,864.064	-6.43	<.0001	1.0144451
WBP*WBC	-2,256.427	-7.78	<.0001	1.0162575
C-5^2	-1,756.503	-8.27	<.0001	1.0597127
C-17^2	-2,249.597	-10.39	<.0001	1.0527594
UHA^2	-2,507.733	-11.81	<.0001	1.0594988
WBP^2	-1,771.223	-8.18	<.0001	1.0527594
WBC^2	-3,477.003	-15.98	<.0001	1.0631796

Table 4.6 Full Model Parameter Estimates, P-values, and VIF - IMS

Examining the ANOVA in Table 4.7 next, we observe that it has an Adjusted R^2 of approximately 0.91. With a value this high we can say that this model, as well as the first, possesses predictive capability.

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	18	8.87E+10	4.93E+09	235.5281
Error	381	7,974,855,780	20,931,380	Prob>F
C Total	399	9.67E+10		<.0001
RSquare	0.917541			
RSquare Adj	0.913646			
Root Mean Square Error	4,575.083			
Mean of Response	272,273.300			
Observations (or Sum Wgts)	400			

Table 4.7 Full Model Regression ANOVA - IMS

Visual verification of how well the model fits can be viewed in the least square regression line plot of figure 4.9.

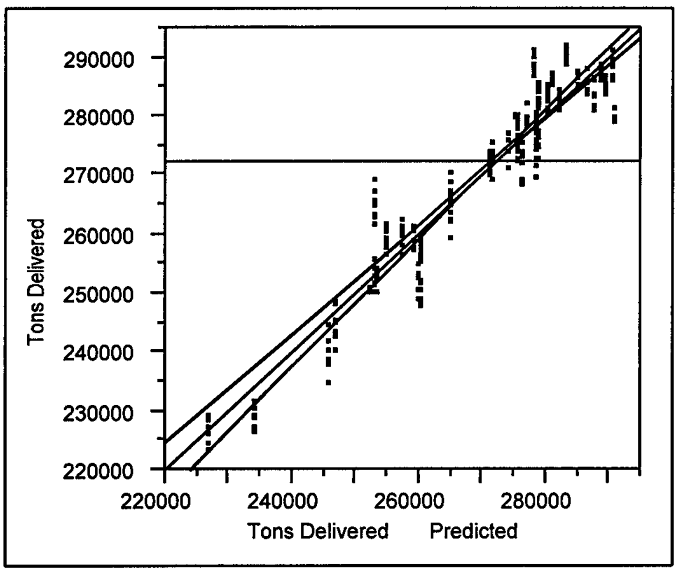

Figure 4.9 Full Model Least Squares Regression Line Plot - IMS

As you recall, to check for outliers we plot the studentized deleted residuals against the run order. Figure 4.10 displays these results.

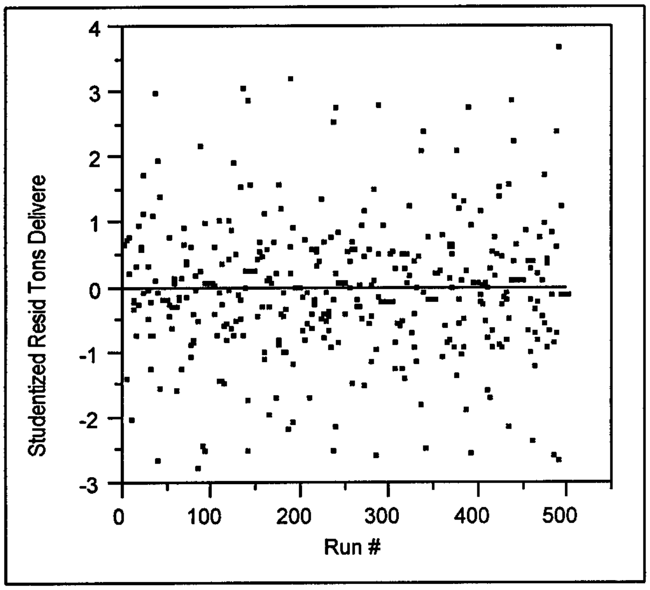

Figure 4.10 Full Model Outlier Plot - IMS

Again we see that only one point is close to being an outlier. Next we move to checking the residuals for normality.

4.13.1 ***Normality IMS.*** The p-value of <0.0001 in Figure 4.11 implies that we fail to meet our normality assumption. This is a violation of assumption number two in section 4.2.1. Further investigation of the residuals is now necessary to see if regression analysis techniques are applicable to this situation.

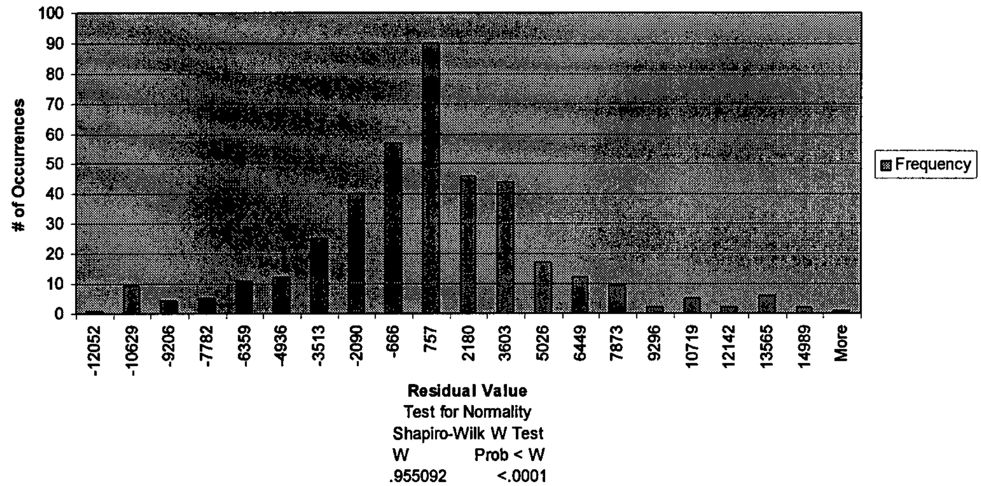

Figure 4.11 Full Model Residual Normality Plot - IMS

Normality not only depends on the distribution of the residuals, it also depends on the number of samples that you have. For a small sample, you are more likely to accept a data set as being normal. This happens because there is not enough data to influence the regression line in any one direction. Conversely, normality is usually rejected for large sets of sample data. The slightest deviation can influence the regression line away from normality [Crown, 1999].

However, this does not mean we need to stray away from using normal linear regression techniques. Visual inspection, and scrutiny of the box plot, quantiles, and moments of the residuals can provide enough confidence to continue using linear regression techniques.

Stem	Leaf	Count
16	4	1
14	3	1
12	3458947	7
10	0773	4
8	78458	5
6	0344580022389	13
4	00122335555688023456788	23
2	00000011112333444445566677788888999900011222233345566678	56
0	0001111111122233344444444445555555555555555677777777777788990001111222223333333344555677789	91
-2	00111111123333334445777888899900001112222333333333333333333333344444555556777777778889999999	92
-4	00001112222344444555566677889999112222334445667778888899	56
-6	015667701124557778	18
-8	26677114466899	14
-10	6881259	7
-12	55770122357	11
-14	9	1

Multiply Stem.Leaf by 10^3

Figure 4.12 Residual Stem and Leaf Plot

The stem and leaf plot in Figure 4.12 appears to have a normal distribution about it. Looking at the box plot of Figure 4.13 we see that both tails seem to be heavy with residuals. Since the residuals seem to be distributed evenly in both tails, we continue with the quest to validate pseudo-normality.

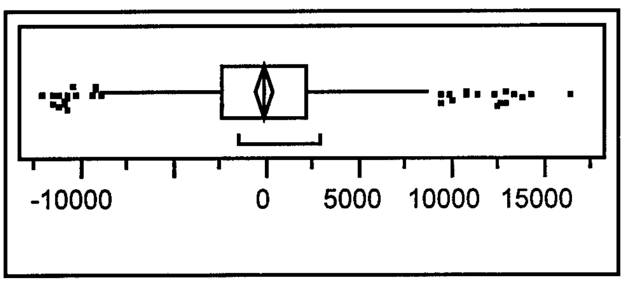

Figure 4.13 Residual Box Plot

Table 4.8 allows us to view the residuals through their quantiles and moments.

Quantiles			Moments	
maximum	100.00%	16412	Mean	0
	99.50%	14312	Std Dev	4470.694
	97.50%	11325	Std Error Mean	223.535
	90.00%	4960	Upper 95% Mean	439.46
quartile	75.00%	2217	Lower 95% Mean	-439.46
median	50.00%	-11	N	400
quartile	25.00%	-2226	Sum Weights	400
	10.00%	-5244	Sum	0
	2.50%	-10743	Variance	19987107
	0.50%	-11518	Skewness	0.272
minimum	0.00%	-12052	Kurtosis	1.56
			CV	3.84E+15

Table 4.8 Quantiles and Moments

Inspecting the quantiles we see that the data is fairly evenly distributed about the median. We see in the moments that the mean is located at zero and the skewness (the shift left or right of the mean) is 0.272. A symmetric distribution has a skewness of zero. The only statistic that may be of question is the kurtosis, which is a measure of "tail weight" of the distribution [Law and Kelton, 1991: 360]. We have an obtained value of 1.56. For a normal distribution, the kurtosis should be equal to 3. However, kurtosis has not been found to be very useful for discriminating among distributions [Law and Kelton 1991: 360].

From the visual observation of the stem and leaf plot, box plot, quantiles, and moments, we can conclude that our residuals, while not normal, are normal enough to proceed with the use of linear regression techniques.

4.13.2 ***Constant Variance IMS.*** As before, we use the Bruesch-Pagan test to check for constant variance in the residuals. The plot of Figure 4.14 shows the distribution of the residuals.

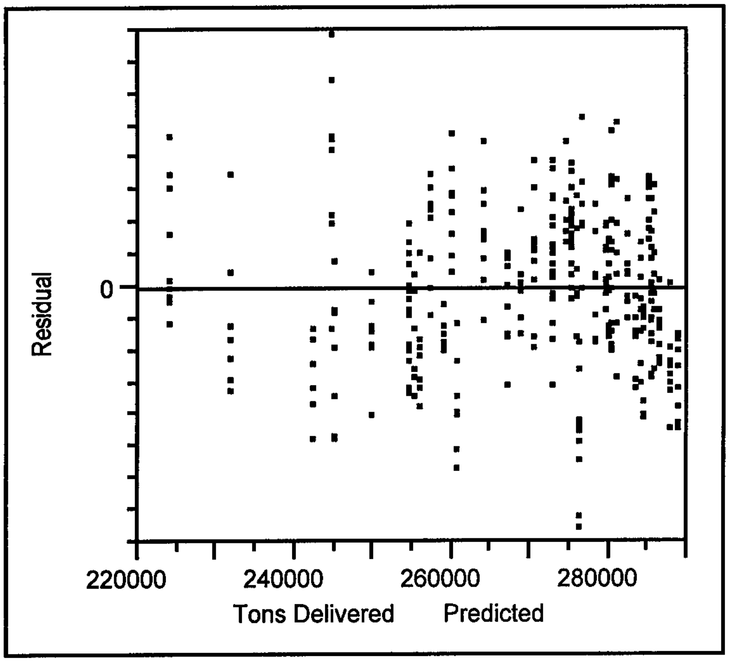

Figure 4.14 Full Model Residual Plot - IMS

Again the residuals appear to lack any sort of a pattern. To quantitatively test this we again use the Breusch-Pagan test.

$$\chi^2_{BP} = \frac{3.19347 \times 10^{17}}{2} \div \frac{7,974,855,780}{18} = 0.8135$$

Looking up the χ^2 value at the 95% confidence level with 18 degrees of freedom, we obtain the value of 28.87. Again, this value is much greater than our obtained value so we fail to reject our hypothesis of constant variance at the 95% level.

For the lack of fit test for the increased MOG scenario, we refer to Table 4.7 and observe again that we reject the null hypothesis for lack of fit with a p-value of <0.0001. As before, we fail the lack of fit test, but conclude that the model is valid for the same reason as previously stated in section 4.5.

4.14 *Metamodel Verification IMS*

Again we check how well the test data fits the regression model that it was designed for. In this case we are referring to the regression model of Table 4.6. Table 4.9 displays how well the metamodel fits the test data. The reader should again note that Table 4.9 contains the same data as Tables 4.6 and 4.7.

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	18	2.67E+10	1.48E+09	54.2971
Error	81	2,215,062,043	27,346,445	Prob>F
C Total	99	2.89E+10		<.0001
RSquare	0.923466			
RSquare Adj	0.906458			
Root Mean Square Error	5,229.383			
Mean of Response	269,633.400			
Observations (or Sum Wgts)	100			
Parameter Estimates				
Term	Estimate	t Ratio	Prob>|t|	VIF
Intercept	281,890.300	210.12	<.0001	0
C-5	7,489.988	12.10	<.0001	1.172510
C-17	2,989.045	4.99	<.0001	1.160429
UHA	7,247.852	11.76	<.0001	1.166141
C-5*UHA	-2,181.637	-3.16	0.0022	1.156555
C-17*UHA	1,298.954	1.85	0.0677	1.205213
WBP	8,133.552	13.71	<.0001	1.142193
C-5*WBP	-2,369.832	-3.29	0.0015	1.257788
C-17*WBP	-1,828.484	-2.54	0.0131	1.269767
UHA*WBP	-1,576.305	-2.30	0.0239	1.109654
WBC	6,184.909	9.35	<.0001	1.433978
C-5*WBC	-1,449.455	-2.00	0.0488	1.230378
UHA*WBC	-1,878.091	-2.61	0.0109	1.269497
WBP*WBC	-2,112.785	-2.89	0.0049	1.303534
$C5^2$	-1,137.575	-2.03	0.0459	1.064403
$C\text{-}17^2$	-2,092.574	-4.15	<.0001	1.066821
UHA^2	-2,236.083	-3.99	0.0001	1.060906
WBP^2	-1,613.077	-3.20	0.002	1.066821
WBC^2	-4,766.729	-8.56	<.0001	1.299841

Table 4.9 Test Data Verification - IMS

We have both a high R^2 and adj. R^2 which means we have an excellent fit of the data. All the coefficients of the test data are of the same magnitude and sign as the model. We are suspect of only one term in the regression model, the C-17 * UHA interaction. However, since we are trying to fit data to the model and removal of the term results in an insignificant loss in the adj. R^2, as before, we shall leave this term in the model.

4.15 *Metamodel Validation IMS*

We test the predictive nature of this metamodel by establishing its MAPE in the same fashion as before. Table 4.10 gives an analytic view of the error associated with this model.

Run	Simulation	Metamodel	Deviation	Absolute % Error	Run	Simulation	Metamodel	Deviation	Absolute % Error
1	286,072	284,941	1,131	0.003952	51	284,710	287,661	-2,951	0.010366
2	243,581	250,334	-6,753	0.027724	52	283,965	280,236	3,729	0.013132
3	267,517	249,588	17,929	0.067020	53	274,499	281,848	-7,349	0.026771
4	277,322	278,370	-1,048	0.003779	54	254,526	249,934	4,592	0.018040
5	289,092	290,575	-1,483	0.005130	55	280,493	278,420	2,073	0.007389
6	256,558	260,360	-3,802	0.014820	56	276,829	274,056	2,773	0.010017
7	277,712	278,371	-659	0.002371	57	259,331	258,973	358	0.001380
8	266,181	268,600	-2,419	0.009089	58	286,461	286,556	-95	0.000331
9	266,847	233,396	33,451	0.125356	59	282,751	283,713	-962	0.003402
10	241,488	261,976	-20,488	0.084842	60	286,167	285,901	266	0.000930
11	275,018	274,056	962	0.003498	61	273,095	278,370	-5,275	0.019315
12	264,046	265,124	-1,078	0.004082	62	250,031	255,887	-5,856	0.023419
13	288,744	281,077	7,667	0.026552	63	252,553	256,941	-4,388	0.017376
14	227,412	230,164	-2,752	0.012100	64	249,634	259,845	-10,211	0.040905
15	282,085	278,504	3,581	0.012697	65	275,821	274,056	1,765	0.006400
16	258,661	258,973	-312	0.001206	66	266,808	265,124	1,684	0.006313
17	278,338	278,890	-552	0.001983	67	289,427	283,149	6,278	0.021690
18	281,677	281,981	-304	0.001078	68	258,344	263,581	-5,237	0.020271
19	284,774	282,178	2,596	0.009116	69	283,271	278,504	4,767	0.016830
20	281,318	278,504	2,814	0.010005	70	255,957	256,402	-445	0.001737
21	273,893	276,221	-2,328	0.008500	71	281,318	281,981	-663	0.002355
22	274,590	282,391	-7,801	0.028411	72	281,318	278,504	2,814	0.010005
23	289,950	274,614	15,336	0.052891	73	256,818	260,104	-3,286	0.012794
24	287,544	288,418	-874	0.003041	74	284,290	292,945	-8,655	0.030444
25	261,796	251,234	10,562	0.040346	75	278,975	287,380	-8,405	0.030128
26	247,206	252,410	-5,204	0.021049	76	286,855	281,077	5,778	0.020142
27	261,469	265,124	-3,655	0.013978	77	288,469	290,575	-2,106	0.007300
28	282,122	281,897	225	0.000796	78	268,897	271,479	-2,582	0.009601
29	282,419	278,504	3,915	0.013864	79	255,952	260,360	-4,408	0.017223
30	243,201	252,410	-9,209	0.037864	80	288,937	283,149	5,788	0.020030
31	225,998	226,687	-689	0.003047	81	286,427	281,077	5,350	0.018678
32	251,690	259,845	-8,155	0.032402	82	290,635	283,149	7,486	0.025756
33	256,509	260,360	-3,851	0.015014	83	226,616	237,599	-10,983	0.048466
34	287,188	281,077	6,111	0.021278	84	262,700	251,234	11,466	0.043648
35	275,948	277,533	-1,585	0.005743	85	252,899	259,845	-6,946	0.027466
36	282,460	281,981	479	0.001698	86	288,386	294,052	-5,666	0.019647
37	282,460	262,312	20,148	0.071332	87	282,463	281,981	482	0.001708
38	237,017	261,976	-24,959	0.105307	88	282,463	278,504	3,959	0.014018
39	250,765	259,845	-9,080	0.036210	89	254,936	252,925	2,011	0.007890
40	252,245	256,402	-4,157	0.016478	90	287,911	281,077	6,834	0.023736
41	281,238	262,312	18,926	0.067296	91	287,823	291,980	-4,157	0.014442
42	242,916	261,976	-19,060	0.078465	92	279,837	290,857	-11,020	0.039380
43	268,848	265,124	3,724	0.013853	93	253,303	253,411	-108	0.000428
44	266,106	268,600	-2,494	0.009373	94	281,299	281,981	-682	0.002423
45	282,704	278,504	4,200	0.014858	95	291,081	278,091	12,990	0.044625
46	283,906	287,661	-3,755	0.013227	96	248,746	256,941	-8,195	0.032947
47	249,813	236,218	13,595	0.054422	97	230,855	234,122	-3,267	0.014153
48	241,590	261,976	-20,386	0.084384	98	285,309	280,236	5,073	0.017781
49	274,349	271,479	2,870	0.010462	99	250,825	255,887	-5,062	0.020179
50	268,701	265,123	3,578	0.013315	100	281,238	281,981	-743	0.002640

Table 4.10 Mean Absolute Percent Error of Validation Data - IMS

Solving this analytically, we find that we have a MAPE of 0.0216%. We can conclude that this model also has a tendency for high predictive capability.

4.16 System Comparison

We can determine which of the two systems performs better by utilizing the paired t-test. With this test, we can assume unequal and unknown variances of the samples. In order to perform a paired t-test, we must ensure that the number of samples (X_{1i}, X_{2i}) for each experiment are equal. We need to define the parameter for the test [Law and Kelton, 1991: 587].

$$Z_i = X_{1i} - X_{2i} \; for \; i = 1, 2, ..., n \tag{4.8}$$

where

X_{1i} = Tons of cargo delivered in 45 days by system one

X_{2i} = Tons of cargo delivered in 45 days by the increased MOG scenario

Now we compute the estimated mean $(\bar{Z})$, the estimated variance (S^2), and the confidence interval (CI) about Z using the t-test as the test statistic.

If the confidence interval contains zero, then there is no distinction between which system would be preferred. If the confidence interval is greater than zero, system one performs better. Conversely, if the confidence interval is less than zero, it implies that system two performs better. For this experiment, the system with the normal set of MOG values is system one and the increased MOG scenario will be considered system two. The results for the paired t-test computed at the 95%

confidence level are shown below.

$$\bar{Z} \pm t_{(1-(\frac{.05}{2}),499)} * (\frac{S}{\sqrt{N}})$$
$$-937 \pm (1.96 * \frac{2795}{22.36})$$
$$-937 \pm 245$$

The 95% confidence interval for the paired t-test yields (-1182, -692). This indicates that system two performs better than system one. We chose the paired t-test in this instance because if the $Z_j's$ are normally distributed, then this confidence interval is exact [Law and Kelton, 1991: 587]. Figure 4.15 shows that the distribution of the $Z_j's$ are normally distributed.

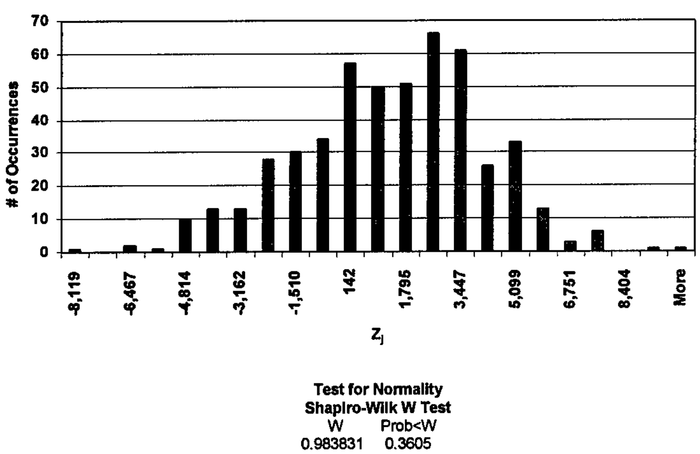

Figure 4.15 Distribution of $Z_j's$

4.17 Conclusion

This chapter has provided an exploration of the methodology discussed in Chapter 3. The two metamodels that were constructed in this chapter showed they possess good predictive capability. The reader should note that the original model contained 4-way instead of 2-way factor interactions. Changing the model from a

4-way to a 2-way factor interaction resulted in negligible loss in the adjusted R^2. This insignificant loss is important since it allows reduction in the model to fewer terms, thereby making a more parsimonious model capable of predicting a wider range of data. This reduction in variables may or may not be feasible with other metamodels due to their significance to the problem.

The primary benefit of a metamodel is that it can be studied using straightforward mathematical analysis [Barry 1992]. Metamodels become useful when the appropriate values are assigned to the unknown parameters. Experimental design is needed to ensure good parameter estimates. Once the experiments have been performed, least squares regression is a standard procedure for estimating the parameters.

V. *Conclusions and Recommendations*

5.1 Introduction

As stated previously, the purpose of this research effort was to develop a valid metamodel that would be able to accurately predict the amount of cargo that would be delivered in a given scenario. Conclusions from the previous chapters are presented here. In addition, recommendations for additional research are presented.

5.2 Conclusions

The graphical and ANOVA analyses of Chapter 4 clearly address the problem statement and research objective of Chapter 1. It is important to remember that model building is an art not a science. Although there are specifications on when to add and remove variables, there are no hard and fast "rules" as to which variables remain in the model. Building a parsimonious model, one that accurately explains the system at hand while using the fewest variables, is the key concept.

The predictability of the two metamodels that were developed in Chapter 4 were validated by evaluating their mean absolute percent error. In both cases, the MAPE was less than 1% supporting the research and the decisions made concerning which variables to keep.

5.3 Design Setting

In order to find out if we are taxing the airlift system to its maximum potential, we would like to know where our maximum delivery occurs. If all our variables are at their high settings, meaning all five aircraft types are set at +1, then we can conclude that we have not saturated the system with aircraft.

Coded Values						
	C-5	C-17	UHA	WBP	WBC	Tons Delivered
Normal MOG	1	-1	1	1	-1	288,902
Increased MOG	1	-1	1	1	1	292,608
Uncoded Values						
Normal MOG	95	65	28	95	50	288,902
Increased MOG	95	65	28	95	95	292,608

Table 5.1 Coded and Uncoded Design Settings for Max Throughput

The above table shows the coded and uncoded values and the associated tons of delivery for both data sets. The Joint Capabilities Strategic Planning guide sets forth the amount of aircraft that are available to be used in the scenario. Typically, AMC uses all available assets to deliver cargo. Using a design of experiment may uncover a theoretically "better" mix of aircraft to use.

Looking at the maximum output of each of the models in Table 5.1 we see that increasing MOG values at various airfields, does not reap a substantial benefit in cargo delivery. This could prompt us to believe that this is a fuel, not MOG sensitive system. However, it only stands to reason that increasing both system attributes would serve well in increasing cargo delivery. Further research would need to be accomplished in order to bring this development to light.

5.4 Recommendations

5.4.1 ***Introduction.*** Although the scope of this research was relatively limited due to the small number of variables, additional research is warranted in the areas depicted below.

5.4.2 ***Fuel and MOG.*** Another design of experiment can be implemented to include Fuel and MOG as factors. This would provide a better understanding as to what would be the limiting factor. By adding these two variables a true sensitivity analysis could be conducted. Complications arise, however, by adding more factors.

By changing from a 2^5 full factorial experiment to a 2^7 full factorial experiment, you increase the number of runs necessary from 50 to 152. You can reduce the number of simulation runs necessary needed by doing a fractional experiment, but then you encounter the problem of aliasing.

5.4.3 ***Convert Passengers to Tons.*** The experiment, as performed, did not take into account the required delivery of passengers. The reason for this omission is that AMC "assumes" that the delivery of soldiers to the theater will take place. As such, the number of passengers delivered to a theater is seldom reported. By multiplying the number of passengers delivered by 350 you can effectively convert the number of passengers delivered into tons. AMC estimates the weight of each individual soldier plus his/her gear to be 350 pounds. By collecting the number of passengers delivered, converting them to tons, and adding this value to the total number of tons that are delivered by cargo planes, may make the WBP more significant to the problem. As it displayed in this effort, the WBP aircraft are little more than place holders. Their only role is to "take up space and resources" to make the cargo aircraft report more accurate data.

5.5 ***Conclusion***

This research provided proof of a relatively simple axiom. Work smarter and not harder. What is meant by that is, if you can develop a good metamodel to replace additional simulation runs you will save time and effort. As long as you, the analyst stay within the established bound or design space of your experiment, you will be able to provide the decision maker with options. Remember, as an analyst you provide insight not answers.

Appendix A. Glossary of Acronyms and Abbreviations

AFM Airlift Flow Model

AMC Air Mobility Command

ANOVA Analysis of Variance

APOD Aerial Port of Debarkation

APOE Aerial Port of Embarkation

BBD Box-Behnken Design

CCD central composite design

CRAF Civil Reserve Air Fleet

DOE design of experiment

JSCP Joint Stategic Capabilities Plan

LMSR Large Medium Speed Roll-On Roll-Off

MAPE Mean Absolute Percent Error

MASS Mobility Analysis Support System

MOG maximum on the ground

NBC Narrow Body Cargo aircraft

RSM response surface methodology

TPFDD Time-Phased Force Deployment Data document

UHA Ultra Heavy Airlifter

VIF Variance Inflation Factor

WBC Wide Body Cargo aircraft

WBP Wide Body Passenger aircraft

Bibliography

1. Banks, Jerry, et al. *Discrete-Event System Simulation.* Upper Saddle River, New Jersey: Prentice-Hall Inc., 1999.

2. Box, George E. P. and Norman R. Draper. *Empirical Model-Building and Response Surfaces.* New York, NY: John Wiley Sons, 1987.

3. Cochran, William G. and Gertrude M. Cox. *Experimental Design.* New York: John Wiley Sons, Inc., 1957.

4. Donovan, James L. *An Application of Response Surface Methodology to A Macroeconomic Model.* MS thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December, 1985.

5. Friedman, Linda W. and Hershey H. Friedman. "Validating the Simulation Metamodel: Some Practical Approaches," *Simulation, 45:3, 144-146* September 1985.

6. Harvey, R. Garrison, et al. "Force Allocation Through Constrained Optimization Of Stochastic Response Surfaces," *Proceedings of the 1992 Winter Simulation Conference, 1121-1129* Winter 1992.

7. Law, Averill M. and W. D. Kelton, editors. *Simulation Modeling Analysis.* New York: McGraw-Hill, Inc., 1991.

8. Myers, Raymond H. and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* New York: John-Wiley Sons, Inc, 1995.

9. Nelson, Barry L. *Statistical Analysis of Simulation Results* (Second Edition). New York: John Wiley Sons, Inc., 1992.

10. Neter, John, et al. *Applied Linear Statistical Models* (Fourth Edition). New York: McGraw-Hall, Inc., 1996.

11. Sall, John and Ann Lehman. *JMP Start Statistics.* Cary, NC: Duxbury Press, 1996.

12. Smetek, Timothy E. *Alternative Assesment For An Aircraft Maintenance System Using Simulation Modeling And Response Surface Methodology.* MS thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, February, 1998.

13. Taylor, Michael K. *Class notes, Stat 696, Applied Linear Statisitcal Models.* MS thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, May, 1999.

14. Taylor, Michael K. *An Response Surface Methodology Study Of The Effects Of Simulation Work And Metamodel Specification On the Statistical Quality Of Metamodel Estimates*. MS thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March, 1994.