

**SOME NEW IDEAS ON FRACTIONAL FACTORIAL
DESIGN AND COMPUTER EXPERIMENT**

A Thesis
Presented to
The Academic Faculty

by

Heng Su

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2015

Copyright © 2015 by Heng Su

SOME NEW IDEAS ON FRACTIONAL FACTORIAL DESIGN AND COMPUTER EXPERIMENT

Approved by:

Professor C. F. Jeff Wu, Advisor
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Jianjun Shi
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Roshan Joseph Vengazhiyil
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Ben Haaland
H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology

Professor Godfried L. Augenbroe
School of Architecture
Georgia Institute of Technology

Date Approved: 5 Jan 2015

*To my beloved parents and my fiancée Xinxian Shao,
for their support, inspiration and encouragement.*

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank many people that have guided me, encouraged me and supported me through my Doctoral studies at Georgia Tech.

First and foremost, I would like to thank my advisor, Professor Jeff Wu, for being a tremendous mentor in both academics and my personal life. His patience, motivation, enthusiasm and immense knowledge have inspired and guided me throughout the study. I am thankful to his extensive care on both research and my personal life that have made my time at Georgia Tech so enjoyable and memorable. I could not have imagined having a better advisor and mentor for my Doctoral study.

I would like to thank Professor Jianjun Shi, Professor Roshan Vengazhiyil, Professor Ben Haaland and Professor Godfried Augenbroe for serving on my dissertation committee and for their valuable comments and suggestions.

I would also like to extend a special thanks to Professor Peter Qian and Dr. Vladimir Brayman. Professor Qian has advised me through my undergraduate studies at University of Wisconsin-Madison. His genuine passion has inspired me to become the researcher I am today. Dr. Brayman has provided valuable research topics and shared his deepest insights, which makes the project on funnel testing possible. My thanks also go to Webtrend Inc. for generously supporting me for the summer of 2013.

I am very thankful for my fellow students at Georgia Tech, especially Dr. Rui Tuo, Dr. Yuming Sun, Dr. Huizhi Xie, Dr. Yijie Wang, Dr. Matthias Tan, Fang Cao, Chengliang Zhang, Li Gu, Yuan Wang, Dianpeng Wang and Simon Mak, who have made my experience at Georgia Tech enjoyable. It is my honor to work with you.

Finally, I would like to leave my last but the most heartfelt appreciation to my beloved parents, my fiancée Xinxian Shao and my cat Muffin, for their companion, encouragement and support. This thesis is dedicated to them.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
SUMMARY	xi
I CME ANALYSIS: A NEW METHOD FOR UNRAVELING ALIASED EFFECTS IN TWO-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS	1
1.1 Introduction	1
1.2 Properties of CME	2
1.3 Method of Analysis	10
1.4 Examples	11
1.4.1 Example 1: Injection molding experiment	11
1.4.2 Example 2: Filtration experiment	14
1.4.3 Example 3: Aluminum experiment	16
1.5 Conclusions	19
II FUNNEL TESTING: REPRESENTATION, DESIGN AND ANALYSIS	21
2.1 Introduction	21
2.2 Representation and Design	23
2.3 Analysis	26
2.4 Toy Example	28
2.5 Simulated Example	37
2.6 Conclusions	49
III UNCERTAINTY QUANTIFICATION OF SOLAR DIFFUSE IRRADIATION ON INCLINED SURFACES FOR BUILDING ENERGY SIMULATION	51

3.1	Introduction	51
3.2	Approach	56
3.3	Uncertainty Quantification of the Perez Model	59
3.3.1	Perez Model Description	60
3.3.2	Experimental Data	61
3.3.3	Statistical Evidence for Model Inadequacy	63
3.3.4	Model Form Uncertainty Quantification	67
3.3.5	Model Validation	69
3.3.6	Prediction of the Modified Perez Model and Uncertainty Quantification	70
3.4	The Effect of Model Form Uncertainty on Building Energy Simulations	72
3.4.1	Case Study Description	72
3.4.2	Results and Discussions	73
3.5	Conclusions	78
IV	PIG PROCESS: JOINT MODELING OF POINT AND INTEGRAL RESPONSES IN COMPUTER EXPERIMENTS	79
4.1	Introduction	79
4.2	PIG Process Model	80
4.2.1	Point GP, Integral GP and PIG Process Models	80
4.2.2	Estimation and Prediction	84
4.2.3	Covariance Functions for Axis-Parallel Rectangular Regions	88
4.3	Examples	90
4.3.1	Example 1	91
4.3.2	Example 2	96
4.4	Conclusions	103
	REFERENCES	105

LIST OF TABLES

1	2^{4-1}_{VI} design with $\mathbf{I}=\mathbf{ABCD}$ and some cmes from the design.	2
2	Design Matrix and Response Data, Injection Molding Experiment. . .	12
3	Design Matrix and Response Data, Filtration Experiment.	14
4	Design Matrix and Response Data, Aluminum Experiment.	17
5	Design Matrix, Toy Example.	29
6	Functions for Decision Probabilities, Toy Example.	30
7	Conversion Rates for Different Funnels, Toy Example.	31
8	Design Matrix and Response Data, Conversion Rates for CF_1	33
9	Design Matrix and Response Data, Conversion Rates for CF_2	35
10	Design Matrix and Response Data, Conversion Rates for CF_3	37
11	Design Matrix, Simulated Example.	39
12	Functions for Decision Probabilities, Simulated Example.	44
13	Conversion Rates for Different Funnels, Simulated Example.	45
14	Conversion Rates and Related Factors, Simulated Example.	48
15	Fitted Models for Each Conversion Rate, Simulated Example.	48
16	Specifications of Instruments.	62
17	Average Annual Errors of the Perez Model in the Calculation of the Global Solar Irradiation on Four Tilted Surfaces.	67
18	Model Validation Statistical Results.	77
19	CPU Time Comparison.	91
20	Training Data, Example 1.	92
21	Training Data, Example 2.	98
22	Additional Four Point Data, Example 2.	102
23	Comparison of Prediction Errors for the Four Point Data, Example 2.	103

LIST OF FIGURES

1	Half-normal Plot, Injection Molding Experiment.	13
2	Half-normal Plot, Filtration Experiment.	15
3	Half-normal Plot, Aluminum Experiment.	18
4	Directed Graph Representation, Simulated Example.	24
5	Half-normal Plot, Conversion Rate for CF_1	32
6	Half-normal Plot, Conversion Rate for CF_2	34
7	Half-normal Plot, Conversion Rate for CF_3	36
8	Simulated Example.	38
9	Directed Graph Representation, Simulated Example.	39
10	Input-to-output Relationship, Perez (1990) Sky Irradiation Model. . .	61
11	Direct Measured and Derived Global Horizontal Irradiation from Two Components.	63
12	The Absolute MBE and RMSE of Global Solar Irradiation Predicted by the Perez Model at Different Sites.	66
13	Measured and Predicted Global Irradiation on Two Vertical Surfaces. .	67
14	Model Prediction Discrepancies against Solar Azimuth, Sky brightness, Direct Normal Solar Irradiation, and Surface Tilt Angle.	68
15	Results on a South-Facing Vertical Surface at Hourly Intervals over Two Days (June 1 and 2, 2011): (a) Perez Predictions, y_{Perez} , and Physical Observations, y_{obs} , and (b) Prediction of diff(\bullet) and 95% Con- fidence Interval.	71
16	Annul Total Cooling Electricity Consumption Predicted from the Use of the Original and the Modified Perez Model.	74
17	Sample-Paired Analysis for Cooling Electricity Use and Heating Nat- ural Gas Use.	76
18	True Function Curve, Example 1.	91
19	True and PIG Prediction Curves, Example 1.	93
20	True and GP(5) Prediction Curves, Example 1.	94
21	True and GP(5+2) Prediction Curves, Example 1.	95
22	True and GP(7) Prediction Curves, Example 1.	96

23	3D Plot and Contour Plot of the True Function. Example 2.	96
24	Plot of Training Data. Example 2.	97
25	True and Predicative Contour Plots, Example 2.	100
26	Scatter Plot of Ten Point Sites and Contour Plot of the True Function, Example 2.	101
27	Scatter Plot of Fourteen Point Sites and Contour Plot of the True Function, Example 2.	102

SUMMARY

This thesis consists of two parts. The first part is on fractional factorial design, and the second part is on computer experiment. The first part has two chapters. In the first chapter, we use the concept of conditional main effect, and propose the CME analysis to solve the problem of effect aliasing in two-level fractional factorial design. In the second chapter, we study the conversion rates of a system of webpages with the proposed funnel testing method. The second part also has two chapters. In the third chapter, we use statistical models to calibrate the Perez model. In the last chapter, we propose a new Gaussian process that can jointly model both point and integral responses.

Ever since the founding work by Finney, it has been widely known and accepted that aliased effects in two-level regular designs cannot be de-aliased without adding more runs. A surprising result by Wu in his 2011 Fisher Lecture showed that aliased effects can sometimes be de-aliased using a new framework based on the concept of conditional main effects (CMEs). In the first chapter, this idea is further developed into a methodology that can be readily used. Some key properties are derived that govern the relationships among CMEs or between them and related effects. As a consequence, some rules for data analysis are developed. Based on these rules, a new CME-based methodology is proposed. Three real examples are used to illustrate the methodology. The CME analysis can offer substantial increase in the R-squared value with fewer effects in the chosen models. Moreover, the selected CME effects are often more interpretable.

Nowadays, internet has become an important source of revenue for various companies. How to design the webpages to maximize the conversions is now a hot topic

in e-commerce. In the second chapter, we propose a new method called the funnel testing to simultaneously study a system of webpages and optimize its overall conversions. Directed graph is used to represent the system of webpages and identify its structure. Fractional factorial design is used to conduct the experiment systematically. A new method of analysis is proposed to maximize the total conversion rate of the system. A toy example is used to demonstrate the idea along the description of the method. Another more complicated simulated example is given to further illustrate the methodology.

Traditional uncertainty quantification (UQ) in the prediction of building energy consumption has been limited to the propagation of uncertainties in model input parameters. Models by definition ignore, at least to some degree, and, in almost all cases, simplify the physical processes that govern the reality of interest, thereby introducing additional uncertainty in model predictions that cannot be captured as input parameter uncertainty. Quantification of this type of uncertainty (which we will refer to as model form uncertainty) is a necessary step toward the complete UQ of model predictions. In the third chapter, we introduce a general framework for model form UQ and shows its application to the widely used sky irradiation model developed by Perez (1990), which computes solar diffuse irradiation on inclined surfaces. We collect a dataset of one-year measurements of solar irradiation at one location in the United States. The measurements were done at surfaces with different tilt angles and orientations, for a wide spectrum of sky conditions. A statistical analysis using both this dataset and published studies worldwide suggests that the Perez model performs non-uniformly across different locations and produces a certain bias in its predictions. Based on the same data, we then use a two-phase regression model, to express model form uncertainty in the use of the Perez model at this particular location. Using a holdout validation test, we demonstrate that the two-phase regression model considerably reduces the model bias errors and root mean

square errors for every tilted surface. Lastly, we discuss the significance of including model form uncertainty in the energy consumption predictions obtained with whole building simulation.

In some computer experiments, the quantity of interest may be the average value of the responses over a specific region. One example from building energy simulation is the diffuse solar irradiance on a building façade representing the integral of the irradiance over the sky dome that the façade is exposed to. Treating this information as point responses will lead to estimation efficiency loss. In the last chapter, we extend the standard point Gaussian process framework so that it can handle both point and integral responses. This new methodology is called the point-integral Gaussian process model, which is abbreviated as the PIG process model. A generic expression of the PIG process model is given with its complicated covariance functions. Parameter estimation and prediction following the frequentist approach is shown. Closed-form expressions of the covariance functions are derived for axis-parallel rectangular regions, whose computational time are compared with the numerical integration using quadrature. Two examples are given to demonstrate the use and the performance of the new methodology. Two point GP models, one ignores the integral responses and the other transforms the integral into point responses, are compared with the PIG process. In all cases, the proposed PIG process model obtains higher prediction accuracy.

CHAPTER I

CME ANALYSIS: A NEW METHOD FOR UNRAVELING ALIASED EFFECTS IN TWO-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS

1.1 Introduction

It is a traditional wisdom that aliased effects cannot be disentangled without adding more runs, where two effects are said to be aliased if they represent the same contrast (Finney, 1943). After nearly 70 years, this belief and practice was broken in the Fisher lecture paper by Wu (2014). He employed a concept called conditional main effect to reparametrize the space of aliased effects, and used variable selection to identify significant effects among the candidate set consisting of main effects, interactions, and selective conditional main effects. An example from GM of Canada (Brajac and Morey, 1987) was used to illustrate the new idea with promising results. The goal of this work is to further explore the concept in Wu (2014) and develop a systematic analysis strategy to de-alias aliased effects in two-level fractional factorial designs.

In this work, we consider only the 2^{k-p} designs where k factors, each at two levels denoted by $+$ and $-$ are being studied. It is a p^{-1} fraction of the 2^k full factorial design. The effects such as main effects, two-factor and higher order interactions considered in traditional analysis are referred to collectively as the *traditional effects* in contrast to the *conditional main effects* discussed in this work. Two effects that are neither orthogonal nor aliased are said to be *partially aliased*. To distinguish the concept of partially aliasing, we will call the aliasing relationship in traditional 2^{k-p} designs as *fully aliased*, which is in line with the terminology in Wu and Hamada (2009, p.363). For definitions and detailed discussions for the traditional effects, and full and partial

aliasing, the readers should consult the book by Wu and Hamada (2009).

In section 1.2, important definitions and properties related to the conditional main effects will be given. Rules of analysis will be derived from these definitions and properties, which form the basis for the method of analysis proposed in section 1.2. Three examples will be used to illustrate the analysis strategy in section 1.4. Concluding remarks on the examples will be given in the last section.

1.2 Properties of CME

Let us start by reviewing the definition of conditional main effects. Consider the first four columns of Table 1. It is a 2_{VI}^{4-1} design with 8 runs and four factors A, B, C and D. The defining relation of this design is $\mathbf{I=ABCD}$.

Table 1: 2_{VI}^{4-1} design with $\mathbf{I=ABCD}$ and some cmes from the design.

A	B	C	D	AB	CD	A B+	A B-	B A+	B A-	A C+	C D-	D C-	A D-
+	+	+	+	+	+	+	0	+	0	+	0	0	+
+	+	-	-	+	+	+	0	+	0	0	-	-	0
+	-	+	-	-	-	0	+	-	0	+	+	0	0
+	-	-	+	-	-	0	+	-	0	0	0	+	+
-	+	+	-	-	-	-	0	0	+	-	+	0	0
-	+	-	+	-	-	-	0	0	+	0	0	+	-
-	-	+	+	+	+	0	-	0	-	-	0	0	-
-	-	-	-	+	+	0	-	0	-	0	-	-	0

Suppose we consider only the first two factors A and B. The standard definition of main effects and two factor interactions (abbreviated as 2fi's thereafter) in texts on design of experiments (Box, Hunter and Hunter, 2005. Wu and Hamada, 2009) is given by:

$$ME(A) = \bar{y}(A+) - \bar{y}(A-), \quad (1)$$

$$ME(B) = \bar{y}(B+) - \bar{y}(B-), \quad (2)$$

$$INT(A, B) = \frac{1}{2}(\bar{y}(A+|B+) - \bar{y}(A-|B-)) - \frac{1}{2}(\bar{y}(A+|B-) - \bar{y}(A-|B+)), \quad (3)$$

$$INT(A, B) = \frac{1}{2}(\bar{y}(B + |A+) - \bar{y}(B - |A-)) - \frac{1}{2}(\bar{y}(B + |A-) - \bar{y}(B - |A+)). \quad (4)$$

where $\bar{y}(A+)$, $\bar{y}(A-)$, $\bar{y}(B+)$, $\bar{y}(B-)$ are the averages of the responses y at the level settings $A+$, $A-$, $B+$ and $B-$ respectively, and $\bar{y}(A + B+)$, $\bar{y}(A - B-)$, $\bar{y}(A + |B-)$ and $\bar{y}(A - |B+)$ are the averages of y at the level settings $A+B+$, $A-B-$, $A+B-$ and $A-B+$ respectively.

Notice that $\bar{y}(A+)$ can also be expressed as $\bar{y}(A+) = \frac{1}{2}(\bar{y}(A + B+) + \bar{y}(A + B-))$, and similarly for $\bar{y}(A-)$, $\bar{y}(B+)$ and $\bar{y}(B-)$. Then (1) and (2) can be written as:

$$ME(A) = \frac{1}{2}(\bar{y}(A + |B+) + \bar{y}(A + |B-)) - \frac{1}{2}(\bar{y}(A - |B+) + \bar{y}(A - |B-)), \quad (5)$$

$$ME(B) = \frac{1}{2}(\bar{y}(B + |A+) + \bar{y}(B + |A-)) - \frac{1}{2}(\bar{y}(B + |A+) + \bar{y}(B - |A-)). \quad (6)$$

In Wu and Hamada (2009), the conditional main effect, *which is henceforth abbreviated as cme*, of A given B at level $+$ is defined as:

$$cme(A|B+) = \bar{y}(A + |B+) - \bar{y}(A - |B+). \quad (7)$$

Similarly, the cme of A given B at level $-$ is defined as:

$$cme(A|B-) = \bar{y}(A + |B-) - \bar{y}(A - |B-). \quad (8)$$

By rearranging terms in (5), it is easy to show $ME(A) = \frac{1}{2}(cme(A|B+) + cme(A|B-))$.

Based on this observation, David Woods calls each cme as a half main effect (in a personal communication).

By interchanging the roles of A and B , we also have:

$$cme(B|A+) = \bar{y}(B + |A+) - \bar{y}(B - |A+), \quad (9)$$

and

$$cme(B|A-) = \bar{y}(B + |A-) - \bar{y}(B - |A-). \quad (10)$$

Thus far, we have defined the main effects, 2fi's and cme's all in terms of the average y values at specific level settings of A and B in (3)-(10). Now we will link

these three types of effects through some algebraic relationships. By adding (3) and (5), we get

$$ME(A) + INT(A, B) = \bar{y}(A + |B+) - \bar{y}(A - |B+) = cme(A|B+). \quad (11)$$

By subtracting (3) from (5), we get

$$ME(A) - INT(A, B) = \bar{y}(A + |B-) - \bar{y}(A - |B-) = cme(A|B-). \quad (12)$$

By adding (4) and (6), we get

$$ME(B) + INT(A, B) = \bar{y}(B + |A+) - \bar{y}(B - |A+) = cme(B|A+). \quad (13)$$

And finally, by subtracting (4) from (6), we get

$$ME(B) - INT(A, B) = \bar{y}(B + |A-) - \bar{y}(B - |A-) = cme(B|A-). \quad (14)$$

From (11)-(14), we can see that each cme is related to a main effect and a 2fi. We call this main effect its *parent effect* and the 2fi its *interaction effect*. The main effect being conditioned on is called the *conditioning effect*, and its corresponding level setting is called the *conditioning level*. The relationships in (11)-(14) can be summarized as the first property:

Property 1 *A conditional main effect is equal to the sum (and respectively, the difference) of its parent effect and its interaction effect, if its conditioning level is + (and respectively -).*

Now let us go back to the design and see if we can derive similar relationships between the columns representing these three types of effects. From now on, we will use the shorthand notation (A|B+), (A|B-), AB and A to represent $cme(A|B+)$, $cme(A|B-)$, $INT(AB)$ and $ME(A)$ respectively.

Again, take the first two factors A and B for illustration. We will show how to write down the column of (A|B+) in a standard way. By the definition of cme,

$(A|B+)$ is the effect of A given B at level +. For the rows with B at level +, the entries of $(A|B+)$ are the same as the entries of A. On the other hand, at B -, the entries of $(A|B+)$ are zero. This cme is represented in column 7 of Table 1. Similarly, we have $(A|B-)$ in column 8. By interchanging the roles of A and B, we have $(B|A+)$ and $(B|A-)$ in columns 9 and 10 respectively.

However, the above procedure is very tedious. To construct a cme, one has to go through the columns of its parent effect and conditioning effect entry by entry. As inspired by Property 1, we will try to find a simple relationship between these three columns.

Take $(A|B+)$ for example. Its parent effect is the main effect A, and its interaction effect is the 2fi AB. By the definition of 2fi, the column of AB (column 5), is constructed by multiplying columns 1 and 2 element by element. Thus, for rows with B at level +, the entries of AB are the same as the entries of A. On the other hand, for B -, the entries of AB have the opposite sign of the entries of A. So if we add A and AB in columns 1 and 5, and divide by 2, we get $(A|B+)$ in column 7. Similarly, if we subtract AB from A and divide by 2, we get $(A|B-)$ in column 8. Therefore we have the following algebraic relationships:

$$(A|B+) = \frac{1}{2}(A + AB), \quad (15)$$

$$(A|B-) = \frac{1}{2}(A - AB). \quad (16)$$

We call them the *construction definition* of cme.

To model a 2^{k-p} design with k factors, the set of candidate effects consists of $4 \times \binom{k}{2}$ cme's, k main effects and $\binom{k}{2}$ 2fi's. The set of candidate models is even larger, which consists of certain subsets of the previous set. Without any restriction, it would be hard to find a good model from such a large candidate set. In analyzing experiments with complex aliasing, Hamada and Wu (1992) have encountered similar situations, where they used the effect sparsity principle and the effect heredity principle to reduce

the size of the candidate set and exclude incompatible models. In this work, *we restrict the search to orthogonal models*, where all effects in a candidate model have to be orthogonal to each other. Here we only consider the following notion of orthogonality. Two columns are orthogonal if their inner product is zero. Let $u = (u_i)$ and $v = (v_i)$ be two column vectors of size n . Their inner product is defined as:

$$u \times v = \sum_{i=1}^n u_i v_i. \quad (17)$$

From this definition, it follows immediately that any two traditional effects are orthogonal if they are not fully aliased with each other. Additionally, the inner product of a traditional effect with itself or a fully aliased effect is the squared norm of that effect, i.e., the number of runs of the design. *These two properties will be used in (19)-(25) without specific referencing.*

First, we explore the orthogonality relationships between cmes and traditional effects. Let $(A|B+)$ be a cme and TE be a traditional effect. By (15), we can write their inner product as:

$$(A|B+) \times TE = \frac{1}{2}(A + AB) \times TE = \frac{1}{2}(A \times TE + AB \times TE). \quad (18)$$

If TE=A, i.e., $(A|B+)$'s parent effect, (18) becomes:

$$(A|B+) \times A = \frac{1}{2}(A \times A + AB \times A) = \frac{1}{2}|A|^2 + 0 \neq 0. \quad (19)$$

Similarly, if TE=AB, i.e., $(A|B+)$'s interaction effect, (18) becomes:

$$(A|B+) \times AB = \frac{1}{2}(A \times AB + AB \times AB) = 0 + \frac{1}{2}|AB|^2 \neq 0. \quad (20)$$

Otherwise, we have:

$$(A|B+) \times TE = \frac{1}{2}(A \times TE + AB \times TE) = 0 + 0 = 0. \quad (21)$$

Combining (19)-(21), we have the second property:

Property 2 *cme's are orthogonal to all the traditional effects except for their parent effects and interaction effects.*

Now we turn to a review of effect aliasing. Consider the same 2_{VI}^{4-1} design. The 2fi's AB and CD are shown in columns 5 and 6 of Table 1. By comparing these two columns, it is seen that they are exactly the same. This suggests that AB and CD represent the same vector (called contrast in design of experiment). Thus, we are not able to distinguish them in the traditional sense. However, by the construction definition in (15) and (16), we can write AB (=CD) in several ways in terms of the cme's:

$$AB = (A|B+) - (A|B-) = (B|A+) - (B|A-)$$

$$CD = (C|D+) - (C|D-) = (D|C+) - (D|C-)$$

Though AB and CD cannot be disentangled, as inspired by Wu (2014), we can use the above equations to reparametrize the space that represents AB and CD with cme's, and choose a subset of the cme's to represent or approximate the fully aliased 2fi's. Recall that we require the selected cmes in the same model to be orthogonal to each other.

Now let us construct some cmes, and check their orthogonality relationships by computing the pairwise inner products. Columns 11 to 14 of Table 1 are (A|C+), (C|D-), (D|C-) and (A|D-) respectively. Together with the other four cme's constructed before in columns 7 to 10, we take pairwise inner products among these eight columns. The computational results are summarized as:

- (i). (A|B+) is orthogonal to (A|B-);
- (ii). For (A|B+), (B|A+), (C|D-) and (D|C-), none of them are orthogonal to each other;
- (iii). For (A|B+), (A|C+) and (A|D-), none of them are orthogonal to each other.

These grouping relationships serve as the motivation for the definition of *twins*, *siblings* and *family* to be given below.

The first group of cme's differ only in their conditioning levels, such as (A—B+) and (A—B-). We call these two cme's the *twins*. As proved in (22), the twin cme's are orthogonal to each other:

$$\begin{aligned} (A|B+) \times (A|B-) &= \frac{1}{2}(A + AB) \times \frac{1}{2}(A - AB) = \frac{1}{4}(A \times A - AB \times AB) \\ &= \frac{1}{4}(|A|^2 - |AB|^2) = 0. \end{aligned} \quad (22)$$

Note that the 2d space of the twin cme's is exactly the same as the 2d space of their parent effect and interaction effect according to (15) and (16). If we include both the twin cme's in the same model, it is the same as having their parent effect and interaction effect and thus no effect de-aliasing can be achieved. Therefore, *only one* of the twin cmes can be included in the model. From Properties 1 and 2, we can substitute a pair of main effect, say A, and a 2fi involving A, say AK by one of the twin cme's (A|K+) and (A|K-). If A and AK have the same sign, we should choose (A|K+) according to Property 1; otherwise, we choose (A|K-). This strategy is especially effective if the effects A and AK have similar magnitudes, because from (13) and (14), the selected cme is much larger than both A and AK, and the ignored cme is much smaller. For simplicity of terminology, we will refer to AK as a 2fi and A as its *parental main effect*. Note that each 2fi has two parental main effects. This strategy can be summarized as the first rule of analysis.

Rule 1 *Substitute a pair of 2fi and its parental main effect that have similar magnitudes with one of the corresponding twin cme's.*

Secondly, we consider the group of cme's that have the same parent effect but not the interaction effects. Consider (A|B+) and (A|C+), which have the same parent effect A, but different interaction effects AB and AC respectively. We call these two

siblings. By (15), their inner product can be written as:

$$\begin{aligned}
(A|B+) \times (A|C+) &= \frac{1}{2}(A + AB) \times \frac{1}{2}(A + AC) \\
&= \frac{1}{4}(A \times A + A \times AC + AB \times AC + AB \times A) \\
&= \frac{1}{4}(|A|^2 + 0 + 0 + 0) \neq 0.
\end{aligned} \tag{23}$$

This is stated as the third property:

Property 3 *Sibling cme's are not orthogonal to each other.*

Next, we consider the group of cme's that have the same or fully aliased interaction effects. These cme's are said to belong to the same family. In Table 1, (A|B+), (A|B-), (B|A+) and (B|A-) have the same interaction effect AB and thus belong to the same family. Similarly (C|D+), (C|D-), (D|C+) and (D|C-) have the same interaction effect CD, which is fully aliased with AB. Therefore, they belong to the same family as the first four cme's. Note that in a family, the twin cme's are orthogonal. Without loss of generality, assume (A₁|B₁+) and (A₂|B₂+) are two non-twin cme's from the same family. This means A₁ ≠ A₂ and B₁ ≠ B₂, but A₁B₁ = A₂B₂. By (15), their inner product can be written as:

$$\begin{aligned}
(A_1|B_1+) \times (A_2|B_2+) &= \frac{1}{2}(A_1 + A_1B_1) \times \frac{1}{2}(A_2 + A_2B_2) \\
&= \frac{1}{4}(A_1 \times A_2 + A_1 \times A_2B_2 + A_1B_1 \times A_2B_2 + A_1B_1 \times A_2) \\
&= \frac{1}{4}(0 + 0 + 0 + |A_1B_1|^2) \neq 0.
\end{aligned} \tag{24}$$

This is summarized as the fourth property:

Property 4 *Non-twin cme's in a family are not orthogonal.*

Recall that we require the effects in a candidate model to be orthogonal to each other. From Properties 3 and 4, we have the second rule of analysis, which excludes non-orthogonal terms in the same model.

Rule 2 *Only one cme among its siblings can be included in the model. Only one cme from a family can be included in the model.*

Finally, we study cme's that have different parent effects as well as different interaction effects. Without loss of generality, let them be $(A|B+)$ and $(C|D+)$, where $AB \neq CD$. By (15), their inner product can be written as:

$$\begin{aligned}
 (A|B) \times (C|D) &= \frac{1}{2}(A + AB) \times \frac{1}{2}(C + CD) \\
 &= \frac{1}{4}(A \times C + A \times CD + AB \times CD + AB \times C) \\
 &= \frac{1}{4}(0 + 0 + 0 + 0) = 0.
 \end{aligned} \tag{25}$$

This gives us the last property:

Property 5 *cme's with different parent effects and different interaction effects are orthogonal.*

Because of the orthogonal modeling requirement, Property 5 leads to the third rule of analysis:

Rule 3 *cme's with different parent effects and different interaction effects can be included in the same model.*

These three rules serve as the basis for the method of analysis proposed in the next section.

1.3 Method of Analysis

Our analysis strategy is based on the following two ideas. First, we consider only orthogonal models. Rules 2 and 3 are used to select orthogonal effects in the model. Second, according to Rule 1, a pair of 2fi and its parental main effect with similar magnitudes can be replaced by one of the corresponding twin cme's. In orthogonal models, this cme is orthogonal to the rest of the effects. Therefore, if this 2fi is aliased with other 2fi's, by substituting it with this cme, the effect aliasing will be unraveled.

Based on the above two ideas, we propose a new method of analysis, called the CME analysis.

CME Analysis:

- (i). Use the traditional analysis methods such as analysis of variance or half-normal plot, to select significant effects, including aliased pairs of effects. Go to (ii).
- (ii). Among all the significant effects, use Rule 1 to find a pair of fully aliased 2fi and its parental main effect, and substitute them with an appropriate cme. Use Rules 2 and 3 to guide the search and substitution of other such pairs until they are exhausted.

In step (i), if the use of half-normal plot is considered too subjective or judgmental, a formal method like the Lenth method can be considered (Wu and Hamada, 2009, Ch.4). In the next section, three examples will be given to illustrate the analysis strategy.

1.4 Examples

In this section, we give three examples to illustrate the analysis strategies proposed in Section 3. All the examples are from real physical experiments using 2^{k-p} designs with Resolution IV. The CME analysis as applied to the data appears to work very well.

1.4.1 Example 1: Injection molding experiment

Shrinkage is a common problem in parts manufactured by injection molding and can reduce the efficiency in the upcoming assembly operations. A team of engineers conducted an experiment on the shrinkage from injection molding using a 2_{IV}^{5-2} design with 16 runs. The defining relations of the design are $\mathbf{I}=\text{ABCE}=\text{BCDF}=\text{ADEF}$ and the six factors are: mold temperature (A), screw speed (B), holding time (C), cycle time (D), gate size (E), and holding pressure (F). The design matrix and data are

given in Table 2, where the response is ten times the percent shrinkage (Montgomery, 1991, p.352). The goal of this experiment was to minimize the shrinkage. We apply the CME analysis to this data set.

Table 2: Design Matrix and Response Data, Injection Molding Experiment.

A	B	C	D	E	F	y
-1	-1	-1	-1	-1	-1	6
1	-1	-1	-1	1	-1	10
-1	1	-1	-1	1	1	32
1	1	-1	-1	-1	1	60
-1	-1	1	-1	1	1	4
1	-1	1	-1	-1	1	15
-1	1	1	-1	-1	-1	26
1	1	1	-1	1	-1	60
-1	-1	-1	1	-1	1	8
1	-1	-1	1	1	1	12
-1	1	-1	1	1	-1	34
1	1	-1	1	-1	-1	60
-1	-1	1	1	1	-1	16
1	-1	1	1	-1	-1	5
-1	1	1	1	-1	1	37

In step (i), we use half-normal plot to identify significant effects. From Figure 1, it is clearly seen that the main effect B (screw speed) is the most significant. It is followed by the main effect A and their 2fi AB, and the R^2 value for the three terms is 96.24%. Because the remaining effects are not significant, we include only these three terms in the first model, denoted as Model 1.1. The p values for B, A and AB are 2.39e-09, 5.38e-05 and 0.022% respectively.

In step (ii), among the significant effects in Model 1.1, we use Rule 1 to identify a pair of 2fi and its parental main effect with similar magnitudes. From Figure 1, the only such pair is A and AB. Therefore, we consider the twin cmes (A—B+) and (A—B-). Since A (=6.938) and AB (=5.938) have the same sign, by Rule 1, we should substitute them with (A|B+). This leads to the model with only two terms B

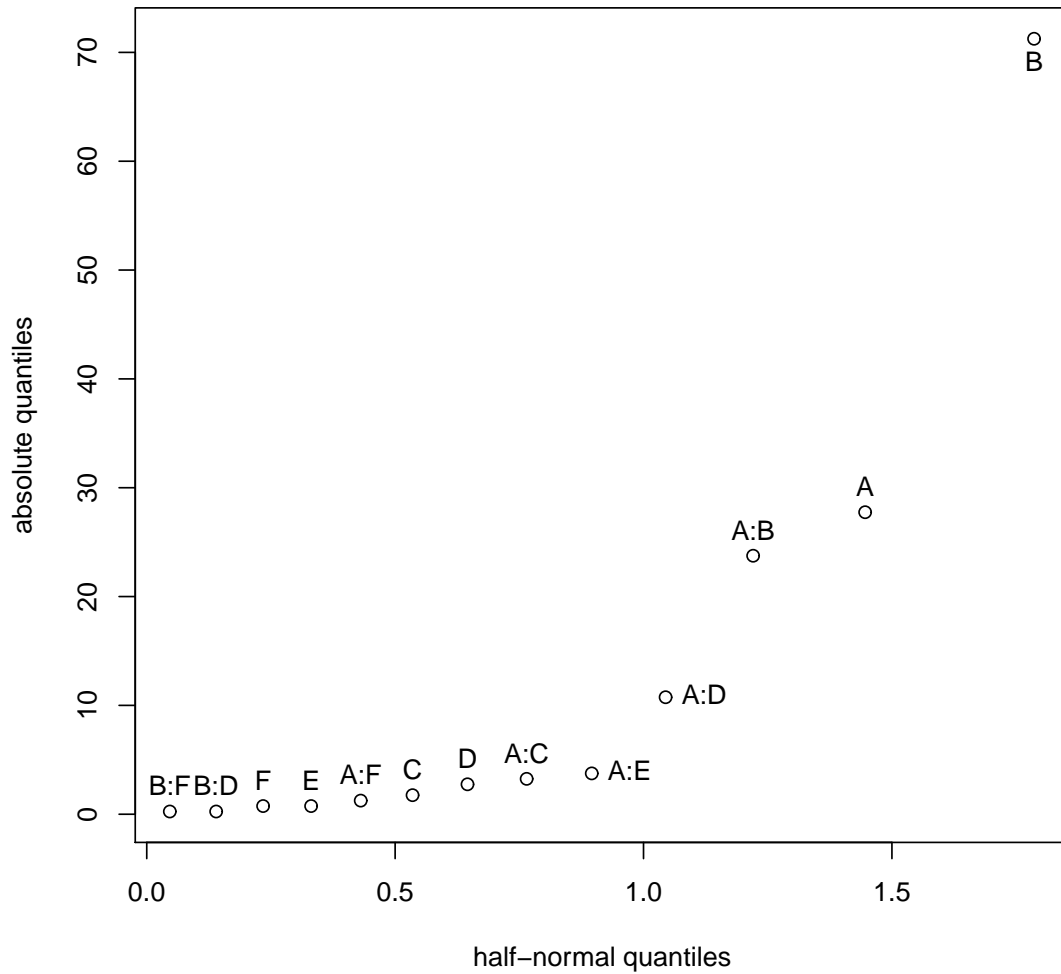


Figure 1: Half-normal Plot, Injection Molding Experiment.

and $(A|B+)$, denoted as Model 1.2. Its R^2 value is 96.14% and the p values for these two terms are $6.06e-10$ and $1.72e-06$ respectively, each of which is more significant than the corresponding term in Model 1.1. Clearly, Model 1.2 is better than Model 1.1. Moreover, unlike AB in Model 1.1, the cme $(A|B+)$ has a good interpretation, i.e., at high screw speed, pressure has a significant effect on shrinkage but not at low speed.

1.4.2 Example 2: Filtration experiment

A team of engineers conducted an experiment on the filtration rate of a chemical product using a 2_{IV}^{4-1} design with 8 runs. The defining relation of this design is $I=ABCD$, and the four factors are: temperature (A), pressure (B), concentration of formaldehyde (C) and stirring rate (D). The design matrix and data are given in Table 3, where the response is measured in gal/h (Montgomery, 1991, p.342). The goal of this experiment was to maximize the filtration rate. We apply the CME analysis to this data set.

Table 3: Design Matrix and Response Data, Filtration Experiment.

A	B	C	D	y
-1	-1	-1	-1	45
1	-1	-1	1	100
-1	1	-1	1	45
1	1	-1	-1	65
-1	-1	1	1	75
1	-1	1	-1	60
-1	1	1	-1	80
1	1	1	1	96

In step (i), we use half-normal plot to identify significant effects. From Figure 2, it is clearly seen that main effects A, D and C, and 2fi's AD (=BC) and AC (=BD) are significant. There is a huge gap between C and B. Therefore, we include these five terms in the first model, denoted as Model 2.1. Its R^2 value is 99.79% and the p values for A, AD, AC D and C are 0.45%, 0.45%, 0.47%, 0.59% and 0.82% respectively.

In step (ii), among the significant effects in Model 2.1, we use Rule 1 to identify a pair of 2fi and its parental main effect with similar magnitudes. From Figure 2, the first such pair are A and AD. Therefore, we consider the twin cmes (A|D+) and (A|D-). Since A (=9.5) and AD (=9.5) have the same sign, by Rule 1, we should substitute them with (A|D+). This leads to a model with four terms (A|D+), AC, D

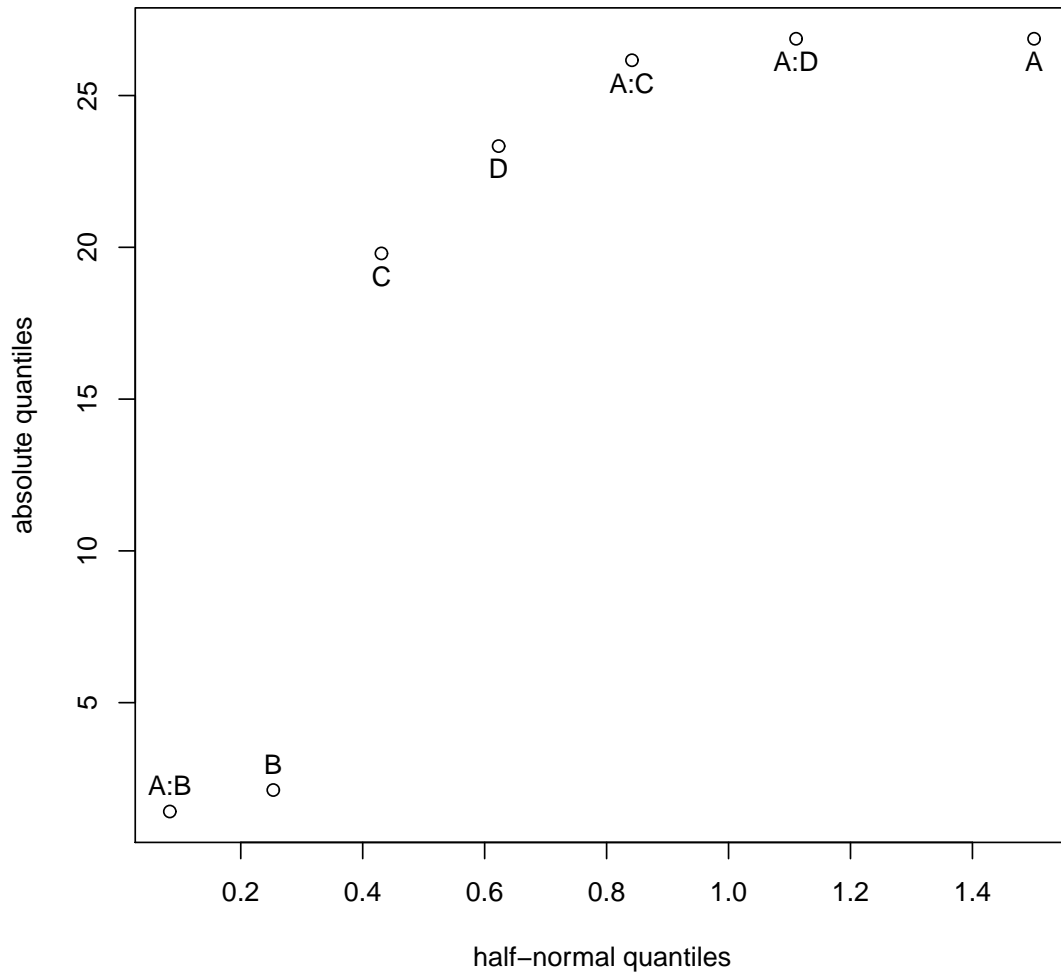


Figure 2: Half-normal Plot, Filtration Experiment.

and C, denoted as Model 2.2. Its R^2 value is 99.79%, and the p values for the four terms are 0.013%, 0.039%, 0.055% and 0.089% respectively, each of which is more significant than the corresponding term in Model 2.1. Clearly, Model 2.2 is better than Model 2.1. Moreover, unlike AD in Model 2.1, the cme (A|D+) has a good engineering interpretation, i.e., at high stirring rate, temperature has a significant effect on filtration rate, but not at low stirring rate.

Next, we search for other such pairs among the rest of the significant effects. In

Figure 2, the next such pair is BD (=AC) and D, and thus we consider the twin cme's (D|B+) and (D|B-). Since D (=8.25) and BD (=−9.25) have opposite signs, by Rule 1, we should substitute them with (D|B-). The two selected cme's (A|D+) and (D|B-) are neither siblings nor belonging to the same family. Therefore, by Rule 3, they are orthogonal to each other and can both be included in the same model. This further reduces the model to three terms: (A|D+), (D|B-) and C, denoted as Model 2.3. Its R^2 value is 99.66%, and the p values for the three terms are 1.96e-05, 2.72e-5 and 0.026%, each of which is more significant than the corresponding term in Model 2.1, and Model 2.2. Therefore, Model 2.3 is the best. Moreover, the cme (D|B-) has a good engineering interpretation, i.e., at low concentration of formaldehyde, the stirring rate has a significant effect on filtration rate, but not at high concentration. Since there are no further such pairs among the remaining significant effects, we conclude the CME analysis with Model 2.3.

1.4.3 Example 3: Aluminum experiment

The Iowa Aluminum Corporation manufactures aluminum sheets from recycled aluminum beverage containers. The molten aluminum was placed onto a continuous strip, and then went through three mills before the final packing. Coolant consisting of oil and water was applied to the metal as it entered the mill each time. The produced aluminum sheets had a rejection rate of 25 percent, so an experiment was undertaken to improve the quality. Due to the limited time and resource, a 2_{IV}^{6-2} design with 16 runs was used. The defining relations of the design are $\mathbf{I}=ABCE=ADEF=BCDF$, and the six factors are: coolant temperature (A), oil percentage (B), coolant volume 1 (C), coolant volume 2 (D), coolant volume 3 (E) and strip speed (F). The design matrix and data are shown in Table 4, where the response is the surface impurity score with a scale of 0-10 with 0 being no impurity and 10 high impurity (Neter et al., 1996, p.1259). The goal of this experiment was to minimize the impurity score.

We apply the CME analysis to this data set.

Table 4: Design Matrix and Response Data, Aluminum Experiment.

A	B	C	D	E	F	y
-1	-1	-1	-1	-1	-1	4
1	-1	-1	-1	1	-1	6
-1	1	-1	-1	1	1	7
1	1	-1	-1	-1	1	2
-1	-1	1	-1	1	1	3
1	-1	1	-1	-1	1	1
-1	1	1	-1	-1	-1	5
1	1	1	-1	1	-1	9
-1	-1	-1	1	-1	1	3
1	-1	-1	1	1	1	2
-1	1	-1	1	1	-1	8
1	1	-1	1	-1	-1	5
-1	-1	1	1	1	-1	4
1	-1	1	1	-1	-1	4
-1	1	1	1	-1	1	4
1	1	1	1	1	1	6

In step (i), we use half-normal plot to identify significant effects. From Figure 3, it is clearly seen that main effects B, F and E are much more significant than the rest. They are followed by two 2fi's AC (=BE) and AF (=DE), and the R^2 value for the five terms is 96.45%. Because the remaining effects are not significant, we include these five terms in the first model, denoted as Model 3.1. The p values for B, F, E, AC and AF are 3.17e-06, 8.56e-6, 8.56e-06, 0.032% and 0.135% respectively.

In step (ii), among the significant effects in Model 3.1, we use Rule 1 to identify a pair of 2fi and its parental main effect with similar magnitudes. From Figure 3, the first such pair are E and BE (=AC). Therefore, we consider the twin cme's (E|B+) and (E|B-). Since E (=1.0625) and BE (=0.6875) have the same sign, by Rule 1, we should substitute them with (E|B+). This leads to a model with four terms (E|B+), B, F and AF, denoted as Model 3.2. Its R^2 value is 94.93% and the p values for the four terms are 3.75e-06, 5.57e-06, 1.58e-05 and 2.68% respectively. By comparing

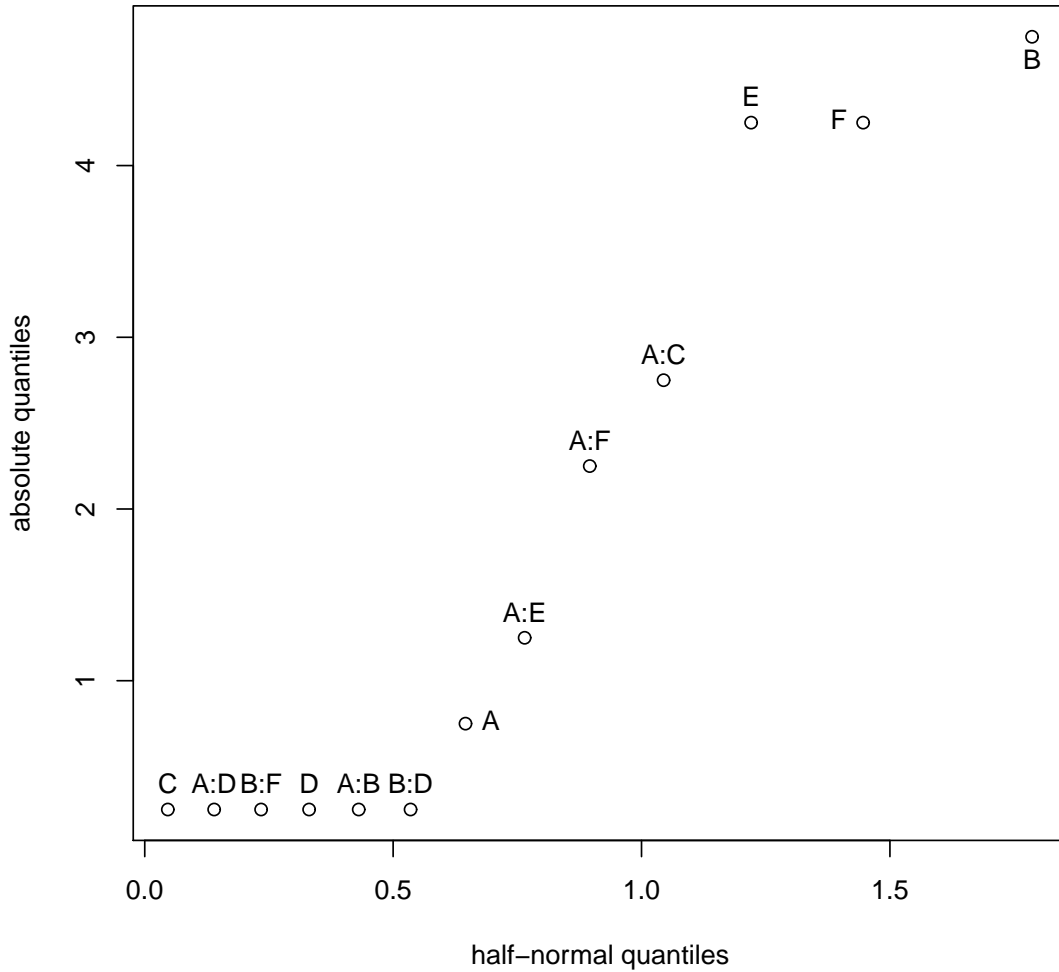


Figure 3: Half-normal Plot, Aluminum Experiment.

the R^2 values and the p values for the significant effects, we cannot say that Model 3.2 is better than Model 3.1. However, unlike BE in Model 3.1, the cme (E|B+) in Model 3.2 has a good engineering interpretation, i.e., at high oil percentage, coolant volume of the third mill has a significant effect on the impurity score, but not at low percentage.

Next, we search for other such pairs among the rest of the significant effects. From Figure 3, there are two pairs associated with AF (=DE): AF with F and DE with E.

If DE and E are chosen, the corresponding twin cme's (E|D+) and (E|D-) are siblings of the selected cme (E|B+) in Model 3.2. By Rule 2, they cannot both be included in the same model. For AF and F, we consider the twin cme's (F|A+) and (F|A-). Since F (= -1.0625) and AF (= -0.5625) have the same sign, by Rule 1, we replace them with (F|A+). The two selected cme's (E|D+) and (F|A+) are neither siblings nor belong to the same family. Therefore, by Rule 3, they are orthogonal to each other and can both be included in the same model. This further reduces the model to three terms: (E|B+), B and (F|A+), denoted as Model 3.3. Its R^2 value is 92.22% and the p values for the three terms are 1.16e-05, 1.75e-05 and 2.40e-05 respectively. By comparing the R^2 values and the p values for significant effects, Model 3.3 is not the best among the three models. However, unlike AF in Model 3.1 and Model 3.2, the cme (F|A+) in Model 3.3 has a good engineering interpretation, i.e., at high coolant temperature, strip speed has a significant effect on impurity score, but not at low temperature. In summary, Model 3.2 or Model 3.3 can be presented to the engineers as alternatives to Model 3.1.

1.5 Conclusions

Inspired by Wu (2014), we develop a systematic method of analysis to de-alias fully aliased 2fi's. Properties of cme's are studied. Rules of analysis are developed from the properties. The method of CME analysis is proposed based on the three rules. Three examples are given to illustrate the analysis strategy. The first two examples show dramatic improvement in model fitting and understanding of effects with the CME analysis. For the last example, though the CME analysis does not give better models based on traditional model selection criteria, the two alternative models are more parsimonious. Furthermore, the cmes identified in the two models have good engineering interpretations. Thus they can serve as alternatives to traditional models for investigators to better understand the experiment.

Effect heredity as defined in Wu and Hamada (2009, Ch.4) does not apply to the CME analysis. In Example 2, by following effect heredity, AC instead of BD should be considered because both A and C are included in the model. However, the CME analysis chooses (D|B-), which represents part of BD. In fact, by the construction definition, each cme is related to its parent effect and interaction effect, while a 2fi is related to its two parental main effects. Therefore, in the spirit of the effect heredity principle, we have a similar notion for the CME analysis: a cme is considered if both its parent effect and interaction effect are significant in the model.

CHAPTER II

FUNNEL TESTING: REPRESENTATION, DESIGN AND ANALYSIS

2.1 Introduction

In the information technology age, more business transactions are conducted on the internet. Webpage has become an importance source of revenue for many companies such as Amazon, Facebook, Walmart, eBay, etc. How to design the webpages to best serve the interest of the business owners has now become a hot research topic in e-commerce (Ash, 2009). The jobs of interest completed by the visitors to the webpages are called conversions. Typical examples of conversions are purchases, newsletter or membership subscriptions, viewing of a page, etc. The percent of visitors completed the jobs of interest is called the conversion rate. A main goal of studying the webpages design is to maximize the conversion rate. This is called *conversion rate optimization (abbreviated as CRO)*. CRO has been extremely important in large IT companies in the last decade. Through this practice, companies have seen a huge increase in their profits. Moreover, companies doing CRO for others, such as Webtrends and SiteTuners, have grown rapidly in the last ten years and are now very popular in the IT industry. Through CRO, they helped the clients achieve greater business success.

Two methods are commonly used in CRO (Ash, 2009). The first one is called A/B test. As the name indicates, this method compares two versions of a webpage: the original version and the proposed new version. A variation of this method is called A/Bn test, where multiple proposed versions of a webpage are compared with the original design in one experiment. Hypothesis testing is used to assess the difference and the best version is chosen as the design of the webpage in the future. The second

method is called multivariate test (*abbreviated as MVT*), where multiple factors, each with two or more levels, are studied in one experiment. MVT is usually implemented with fractional factorial designs, and models are fitted for the conversion rates with respect to the factors. The optimization is done by the standard method of choosing optimal level settings in design of experiment and the best combination of factors is used for future webpage design (Wu and Hamada, 2009. Montgomery, 2012).

A/Bn test is more commonly used in CRO because it is easy to understand, implement and analyze. When companies started a web campaign, they usually have different designs for their campaign page. In order to maximize the revenues from the campaign, they use the beginning part of the campaign to do an A/Bn test and use the best version as the campaign page for the rest of the campaign. On the other hand, when it comes to optimizing a product page, which consists of multiple sections such as header, banner, text and pictures, MVT can be more efficient.

The most commonly studied page with CRO is the landing page, which is the first page visitors see when directed from other sources such as search engines or directly entered web addresses (Ash, 2009). Most landing pages only have the general information of the companys products and services, and conversions usually do not take place here. For example, suppose the visitors want to buy products from the website. Before they make payments, they usually have to go through the product description, view the product pictures, check the product reviews, and enter the payment information. With so many other pages involved, studying just the landing page in order to maximize the conversion rate may be an oversimplification. The series of pages the visitors went through until a possible conversion is called the *conversion funnel* (Ozolins, 2012), or abbreviated as the *funnel* if there is no ambiguity in the context. In the last example, the conversion funnel consists of the landing page, the product description page, the product picture page, the product review page and the payment information page. Moreover, conversions may take place on different

pages. For example, the visitors can subscribe the newsletters on the landing page, the product page and even on the payment confirmation page. The pages where conversions can possibly happen are called *conversion points*. If the landing page is studied with respect to only one conversion point, the results can hardly be conclusive. The set of pages associated with the conversion of interest is defined as the *conversion system*, or abbreviated as the *system* if there is no ambiguity in the context. A conversion system consists of the landing page, all the conversion points, and all other pages that link between them. For example, suppose we have three pages: the landing page, page 1 and page 2, and both page 1 and page 2 are conversion points. Suppose visitors can go from the landing page to either page 1 or page 2, and also from page 1 to page 2. Then the conversion system consists of these three pages, and we referred this example as the *toy example*. This example will be used as the primary illustrative tool of the framework proposed in this work. In this example, the landing page and page 1 make a conversion funnel.

In the next section, we will use directed graph to represent a conversion system and use this graph representation to identify all the conversion funnels in a system. A fractional factorial design on all the pages in a system will be used to conduct the experiment. In section 3, we will propose an analysis strategy to optimize the conversions in a system. A simulated example will be given in section 4, and we conclude this work with remarks on the example and future research.

2.2 Representation and Design

The idea of using directed graph to study the internet originates from a concept called webgraph (Donato et al., 2004), where webpages are viewed as vertices and the linkage relationships between pages are expressed with directed edges in the graph. In computer science, the size of the webgraph being studied is usually very large (over millions) , and researchers are interested in the large-scale properties such as

in/out distributions, connectivity, cyclic patterns and so on. The results are used to identify communities and hubs, filter spams, rank pages and predict the growth of the internet (Donato et al., 2004). In this work, we use this idea, and represent a conversion system with directed graph. Here we are most interested in how to identify all the conversion funnels from the graphical structure. The representation is straightforward. All the webpages in a conversion system are viewed as vertices of the graph. If there is a hyperlink on page X referring to page Y, draw a directed edge from X to Y. For the toy example, the conversion system consists of three pages: the landing page, page 1 and page 2, and we denote these three vertices as v_L , v_1 and v_2 respectively. On the landing page, there are links referring to both page 1 and page 2. Therefore, there are two edges starting at v_L and directing to v_1 and v_2 respectively. We denote them as e_1 and e_2 correspondingly. Furthermore, there is a link on page 1 referring to page 2. Draw another edge from v_1 to v_2 denoted by e_3 . There are no more links in this conversion system, so the toy example is represented by the following graph in Figure 4. The conversion points are marked in solid dots in contrast to others.

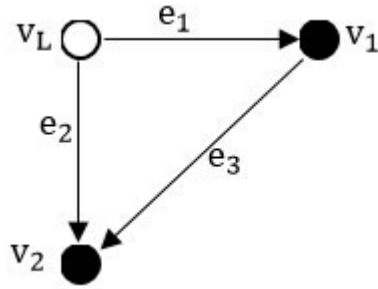


Figure 4: Directed Graph Representation, Simulated Example.

After representing the conversion system as a directed graph, the next step is to identify all the conversion funnels. By definition, a conversion funnel is a series of pages that a visitor have gone through before making a possible conversion. Since visitors always start with the landing page, and make conversions on the conversion

points, in the graph, a conversion funnel is a path from the landing page to a conversion point. To identify all the conversion funnels in the system amounts to finding all the paths connecting the landing page and all the conversion points. In the toy example, page 1 and page 2 are both conversion points. For page 2, there are two paths connecting it to the landing page: v_L via e_1 to v_1 then via e_3 to v_2 , and v_L via e_2 to v_2 . These two conversion funnels are referred as CF_1 and CF_2 . For page 1, there is only one path connecting it to the landing page: v_L via e_1 to v_1 . This conversion funnel is referred to as CF_3 . Since there are no more conversion points in this system, the toy example has three different conversion funnels.

Before this work, some researchers have realized the concept of conversion funnels and have done experiments using this concept (Qualaroo, 2014). However, they used one-page-at-a-time method, and studied the pages in a funnel sequentially. Note that, even for a small conversion system like the toy example, there are already three different conversion funnels. The old method is extremely time consuming and ignores any interactions between different pages. In this work, we study the conversion system as a whole and design one experiment for all the pages involved.

Since there are multiple pages involved in the experiment, A/Bn test will be inefficient. How about the MVT? In MVT, the first step is to identify the factors being studied. Since this experiment considers all the pages in the conversion system at one time, the set of factors consists of the factors from all the pages. For example, in the toy example, suppose each page has two factors to be studied: A and B from the landing page, C and D from page 1 and E and F from page 2. The factors being studied in this experiment are A, B, C, D, E and F. After identifying the correct set of factor, the next step is to construct a fractional factorial design for them. For details on the choice of designs, the reader may refer the book by Wu and Hamada (2009).

For ease of implementation, we assume each factor has two levels. In this case,

a 2^{k-p} design is used, where k factors are being studied, each at two levels denoted by $+$ and $-$. It is the p^{-1} fraction of the 2^k full factorial design. In some other cases, mixed level designs are also used.

The conversion data are collected as binary responses: 1 means conversion made and 0 means no conversion. The conversions for each funnel are recorded separately for ease of analysis. Take the toy example, if a conversion is made on page 2, and the visitor comes directly from the landing page, this conversion is recorded with respect to the data with CF_2 .

2.3 Analysis

The final objective of this work is to maximize the conversions in the system. Define the total conversion rate as the weighted sum of the conversion rates of all the conversion funnels in the system, where the weights reflect the importance of the funnels to the business owners. For example, in studying the sales of shoes, the weights can be the price of the shoes on different funnels. When considering the subscription rate of the newsletters, the weights can be set to be equal. The goal now is to maximize the total conversion rate.

The first step is to build a function for the total conversion rate. Since the total conversion rate is a linear function of the conversion rates of all the funnels in the system, we can build functions for the conversion rates of the funnels separately, and put them together for optimization. To build a function for the conversion rate of a funnel, one has to correctly calculate the response, i.e., the conversion rate of the funnel, and identify the right predictors. Since we have recorded the conversion data for each funnel separately, the conversion rate is the number of conversions of the funnel divided by the total visitors to the conversion system. If the conversion data are recorded only at each conversion point but not for each conversion funnel, the analysis will run into problems. Take the toy example, page 2 is a conversion

point, but it is associated with two conversion funnels CF_1 and CF_2 . If we only have the conversions on page 2, there is no way to tell whether it is from CF_1 or CF_2 . A conversion funnel consists of several pages, and the factors considered in the model for this funnel should be the unions of factors related to these pages. After identifying the right predictors and responses, the next step is to build a linear model. For information on the analysis of fractional factorial design, the readers may review the book by Wu and Hamada (2009).

The last step is to put all the models together and maximize the objective function. Since the total conversion rate is a linear function of all the factors related to the system, the optimization follows the standard method of optimal level settings in design of experiment except for one thing. In the last step, the models are built with respect to different funnels, and a page may appear in different funnels. Under certain circumstances, multiple versions are allowed for the same page. For example, if you want to design a three page conversion system to maximize the membership subscriptions. For simplicity, suppose this system is the same as in the toy example. Then, let us check which page(s) can have multiple versions and why. The landing page appears in all three conversion funnels, but the visitors have to start with the landing page, so all three funnels share the same version of the landing page. Page 1 appears in CF_1 and CF_3 , but visitors are always directed from the landing page to page 1. Therefore, page 1's in the two funnels are the same. Page 2 appears in CF_1 and CF_2 , but before page 2, visitors have viewed different pages for these two funnels: for CF_1 , visitors have viewed the landing page as well as page 1 before page 2, whereas for CF_2 , visitors have only viewed the landing page before page 2. Since page 1 is not in CF_2 , visitors in the two funnels have been exposed to different information before page 2. Therefore, page 2's in CF_1 and CF_2 can have different versions. This can be explained with the following example. Suppose page 1 shows the benefits of the membership and page 2 shows the price in the original design. Normal

visitors check the benefits and compare them with the price on page 2 and determine whether to subscribe the membership. If the visitor does not care about the price, s/he may subscribe after viewing the benefits, and if the visitor is eager to know the price, s/he may jump to page 2 directly from the landing page. For the last visitor, s/he does not have the information about the benefits, the decisions made might be biased. Suppose the membership price is high, but also offers great benefits whose monetary value may exceed the cost. For normal visitors, after comparing the cost with benefits, they are likely to pay for the membership price. But for visitors that have skipped the benefits page, just by looking at the price, they might think it is too high and decide not to go for it. These two decisions are made with different amount of information, and the second one is biased. With the concept of the conversion funnel, it is immediately noticed that these two decisions are made on two different funnels. In order to correct the second situation, we simply show the visitors another version of page 2, which has the benefits as well as the price of the membership, if they come directly from the landing page. In the analysis, we set the factors of page 2 in the model of CF_2 to be E' and F' in distinction to E and F in CF_1 . Therefore, instead of six factors, we now optimize with respect the eight factors: A , B , C , D , E , F , E' and F' .

2.4 Toy Example

In this section, we use the toy example and simulate a set of data to illustrate the analysis strategy. Recall that the conversion system in the toy example consists of three pages: the landing page, page 1 and page 2, and there are three conversion funnels in the system: CF_1 , CF_2 and CF_3 . Note that we suppose each page has two factors to be studied, so a 2^{6-2} design is used for the simulation, where each row of the design matrix represent a version of the conversion system. The design matrix is given in Table 5, whose defining relations are $\mathbf{I}=\mathbf{ABCE}=\mathbf{BCDF}=\mathbf{ADEF}$.

Table 5: Design Matrix, Toy Example.

Run	A	B	C	D	E	F
1	1	1	1	1	1	1
2	1	1	1	-1	1	-1
3	1	1	-1	1	-1	-1
4	1	1	-1	-1	-1	1
5	1	-1	1	1	-1	-1
6	1	-1	1	-1	-1	1
7	1	-1	-1	1	1	1
8	1	-1	-1	-1	1	-1
9	-1	1	1	1	-1	1
10	-1	1	1	-1	-1	-1
11	-1	1	-1	1	1	-1
12	-1	1	-1	-1	1	1
13	-1	-1	1	1	1	-1
14	-1	-1	1	-1	1	1
15	-1	-1	-1	1	-1	1
16	-1	-1	-1	-1	-1	-1

For each simulation, we first choose a version of the conversion system from the 16 candidates in the design table with equal probability, and then simulate the visitors' behavior in the chosen system. The visitor always starts with the landing page, and s/he then have three choices: go to page 1, go to page 2 or leave the system. The first step is to simulate the decision on the landing page. Suppose the visitor goes to page 1 or page 2 with probabilities t_1 or t_2 respectively, where t_1 and t_2 are functions of the factors related to the landing page, i.e., A and B. All the functions of decision probabilities used in this simulation are listed in Table 6. The decisions are made sequentially. First, we determine if the visitor goes to page 1. If s/he does not go to page 1, then we check if s/he goes to page 2. If the choice of the second one is still negative, s/he leaves the system. So to be specific, t_2 is the probability of the visitor going to page 2 given that s/he does not go to page 1. If the visitor chooses to go to page 1, then on page 1, there are still three choices: make a conversion, go to page 2 or leave. We then simulate his/her decision on page 1. Suppose s/he makes a

conversion with probability c_1 , and if s/he does not make a conversion, then goes to page 2 with probability t_3 . If the decisions are both negative, s/he leaves the system. Note that, although the choices are made on page 1, the two probabilities c_1 and t_3 are functions of factors related to both the landing page and page 1, i.e., A, B, C and D, because it is believed that the information on the landing page will affect the visitors' behavior thereafter. Finally, for visitors that land on page 2, there are two types: they come directly from the landing page, or have visited page 1. This represents the two conversion funnels CF_2 and CF_1 respectively. Since page 2 has no link referring to others, the choices are whether to make conversions. The decisions are made as follows. If the visitor comes via CF_2 , s/he converts with probability c_2 ; otherwise, convert with probability c_3 . Note that CF_2 consists of two pages. Therefore, c_2 is a function of the factors related to those two pages, i.e., A, B, E and F. Similarly, c_3 is a function of factors A, B, C, D, E and F. Each simulation terminates when the visitor either makes a conversion or leaves the system. We ran 10,000 simulations, and record the conversions of each funnel separately. The conversion rates are given below in Table 7.

Table 6: Functions for Decision Probabilities, Toy Example.

$t_1 = 0.25 - 0.1A$
$t_2 = 0.42 + 0.21B$
$t_3 = 0.1 + 0.08A - 0.3D + 0.1AD$
$c_1 = 0.38 - 0.05B + 0.12C + 0.08BD$
$c_2 = 0.07 + 0.2E + 0.07AE$
$c_3 = 0.15 + 0.06D - 0.1F + 0.02AD + 0.01CF$

For simplicity, assume all the conversion funnels have equal weights in the toy example. Denote the total conversion rate as CR_T , and the conversion rates for CF_1 , CF_2 and CF_3 as CR_1 , CR_2 and CR_3 respectively. The objective function for optimization can be written as:

Table 7: Conversion Rates for Different Funnels, Toy Example.

CF_1	CF_2	CF_3
0.00000000	0.16349206	0.08730159
0.00960000	0.16960000	0.05440000
0.00000000	0.00000000	0.05546218
0.00000000	0.00000000	0.02073365
0.00000000	0.00000000	0.06833333
0.00000000	0.00000000	0.10194175
0.00000000	0.06718750	0.03593750
0.01093750	0.06406250	0.07812500
0.00000000	0.00000000	0.14803150
0.02044025	0.00000000	0.12421384
0.00000000	0.07833333	0.09166667
0.00000000	0.08258258	0.05255255
0.00000000	0.02508361	0.16722408
0.00000000	0.03475513	0.21169036
0.00000000	0.00000000	0.07954545
0.01716069	0.00000000	0.14820593

$$CR_T = CR_1 + CR_2 + CR_3.$$

Then we build models for CR_1 , CR_2 and CR_3 separately. Recall that the first step in modeling the conversion rate of a funnel is to identify the related factors. For CR_1 , since CF_1 consists of three pages: the landing page, page 1 and page 2, the factors considered in this model are A, B, C, D, E, and F. Similarly, for CR_2 and CR_3 , the factors considered are A, B, E and F and A, B, C and D respectively.

The model building is then straightforward. For CR_1 , it is a function of all the six factors. The corresponding design matrix and responses considered in this model are shown in Table 8. The first step of modeling is to make a half-normal plot to identify significant effects. In Figure 5, it is clearly seen that main effects F and D and two-factor interaction (*abbreviated as 2fi thereafter*) AE are the most significant. Because these three terms have the same magnitude, we denote them as group 1. They are followed by main effects E and A and 2fis AF and D, whose magnitudes are

the same, and denoted as group 2. The other effects are not significant. Therefore the model for CR_1 has seven terms. The R^2 value for this model is 99.12% and the p values for the significant effects in group 1 and group 2 are 1.91e-7, and 0.133% respectively. The explicit expression of the model is:

$$\begin{aligned}
 CR_1 = & 0.00336337 - 0.00336337F - 0.00336337D - 0.0010665E - 0.0010665A \\
 & + 0.00336337AE + 0.0010665AF + 0.0010665AD.
 \end{aligned}
 \tag{26}$$

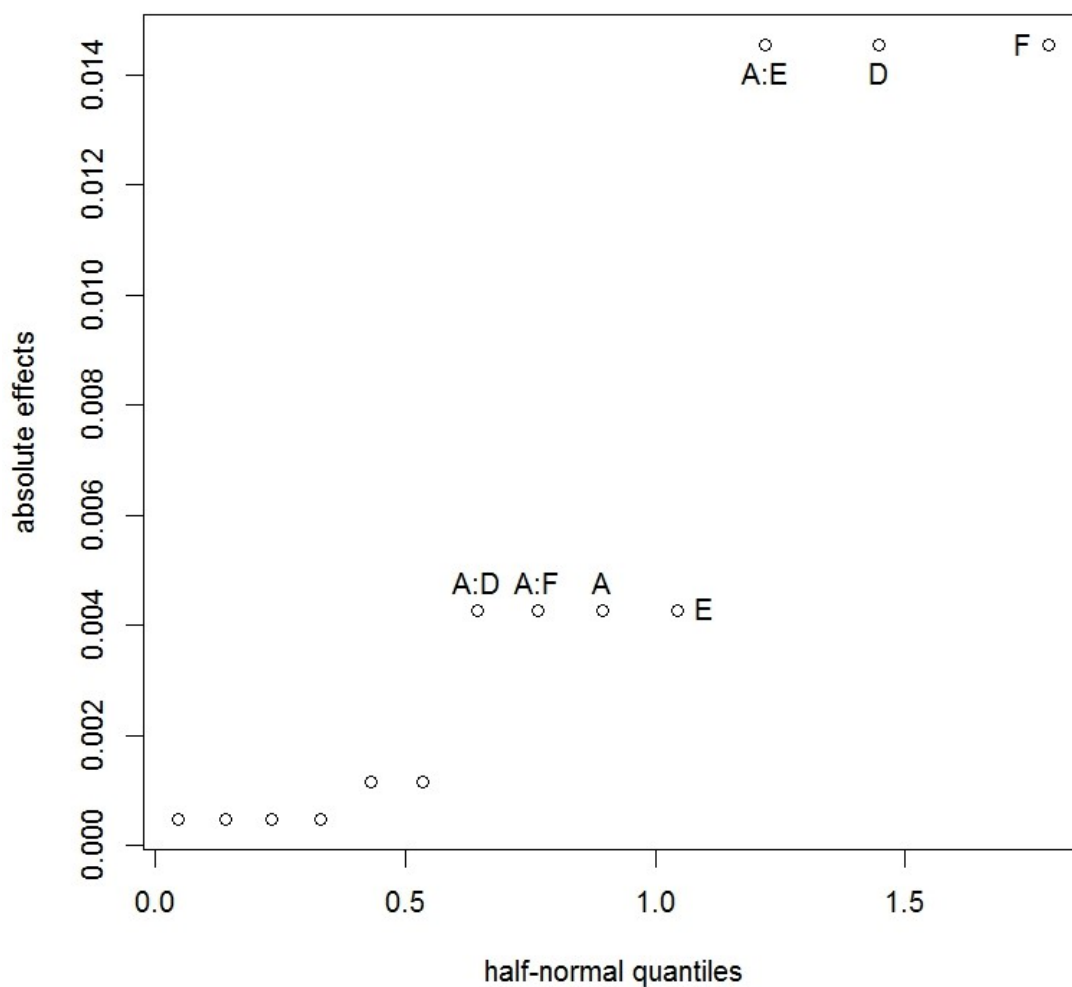


Figure 5: Half-normal Plot, Conversion Rate for CF_1 .

Table 8: Design Matrix and Response Data, Conversion Rates for CF_1 .

A	B	C	D	E	F	CR_1
1	1	1	1	1	1	0.00000000
1	1	1	-1	1	-1	0.00960000
1	1	-1	1	-1	-1	0.00000000
1	1	-1	-1	-1	1	0.00000000
1	-1	1	1	-1	-1	0.00000000
1	-1	1	-1	-1	1	0.00000000
1	-1	-1	1	1	1	0.00000000
1	-1	-1	-1	1	-1	0.01093750
-1	1	1	1	-1	1	0.00000000
-1	1	1	-1	-1	-1	0.02044025
-1	1	-1	1	1	-1	0.00000000
-1	1	-1	-1	1	1	0.00000000
-1	-1	1	1	1	-1	0.00000000
-1	-1	1	-1	1	1	0.00000000
-1	-1	-1	1	-1	1	0.00000000
-1	-1	-1	-1	-1	-1	0.01716069

For CR_2 , the corresponding funnel consists of two pages: the landing page and page 2. Therefore, only factors A, B, E and F are considered. The corresponding design matrix and responses used in this model are shown in Table 9. We start the analysis by drawing a half-normal plot to identify significant effects. In Figure 6, it is clearly seen that the main effect E is the most significant. It is followed by B and BE, and then A and AE. The other effects are not significant. Therefore the model for CR_2 has five terms. The R^2 value for this model is 97.28% and the p values for the significant effects are 4.11e-8, 6.71e-5, 6.71e-5, 0.0377% and 0.0377% respectively. The explicit model is written below:

$$\begin{aligned}
 CR_2 = & 0.042819 + 0.042819E + 0.018932B + 0.015224A \\
 & + 0.018932BE + 0.015224AE.
 \end{aligned}
 \tag{27}$$

Finally, for CR_3 , the corresponding conversion funnel CF_3 consists of two pages: the landing page and page 1. Therefore, factors A, B, C and D should be considered.

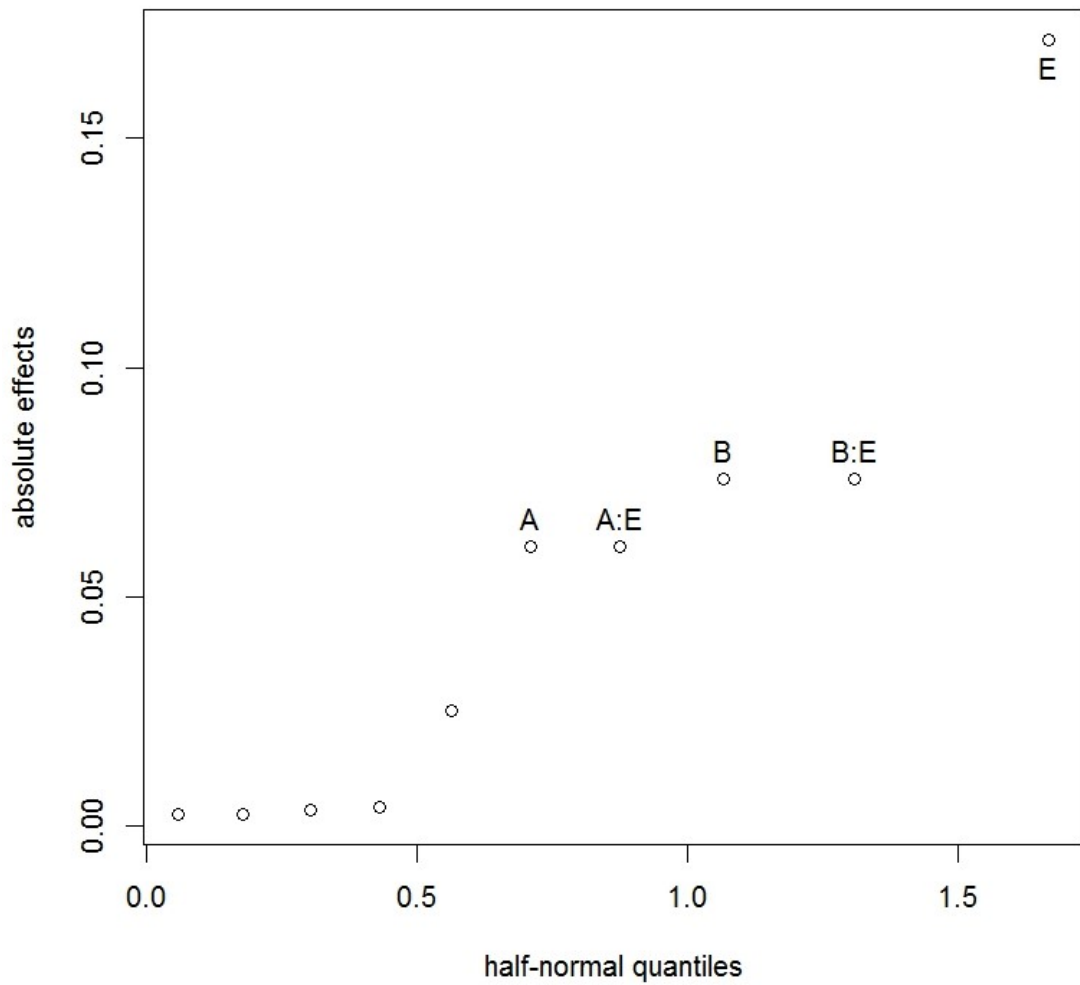


Figure 6: Half-normal Plot, Conversion Rate for CF_2 .

The corresponding design matrix and responses used in this model are shown in Table 10. We again start the analysis by drawing a half-normal plot to identify significant effects. In Figure 7, it is seen that the main effect A is the most significant. It is followed by C, BD, B, AC and AB. The other effects are not significant. Therefore the model for CR_3 has six terms. The R^2 value for this model is 98.3% and the p values for significant effects are $1.23e-7$, $1.16e-6$, $7.72e-6$, $4.47e-5$, 1.51% and 6.42%

Table 9: Design Matrix and Response Data, Conversion Rates for CF_2 .

A	B	E	F	CR_2
1	1	1	1	0.16349206
1	1	1	-1	0.16960000
1	1	-1	-1	0.00000000
1	1	-1	1	0.00000000
1	-1	-1	-1	0.00000000
1	-1	-1	1	0.00000000
1	-1	1	1	0.06718750
1	-1	1	-1	0.06406250
-1	1	-1	1	0.00000000
-1	1	-1	-1	0.00000000
-1	1	1	-1	0.07833333
-1	1	1	1	0.08258258
-1	-1	1	-1	0.02508361
-1	-1	1	1	0.03475513
-1	-1	-1	1	0.00000000
-1	-1	-1	-1	0.00000000

respectively. The explicit model expression is given below:

$$\begin{aligned}
 CR_3 = & 0.095335 - 0.032556A + 0.025057C - 0.01604B + 0.19968BD \\
 & - 0.009842AC + 0.007735AB.
 \end{aligned} \tag{28}$$

Recall that page 2 appears in both CF_1 and CF_2 and can have two different versions. Replace the factors E and F in (27) with E' and F' respectively. Now, put all three models together, we have the total conversion rate expressed as a function of all the eight factors:

$$\begin{aligned}
 CR_T = & 0.1415174 - 0.0183985A + 0.002892B + 0.025057C - 0.0036337D \\
 & - 0.0010665E - 0.0036337F + 0.042819E' + 0.007735AB \\
 & - 0.009842AC + 0.0010665AD + 0.0036337AE + 0.0010665AF \\
 & + 0.015224AE' + 0.019968BD + 0.018932BE'.
 \end{aligned}$$

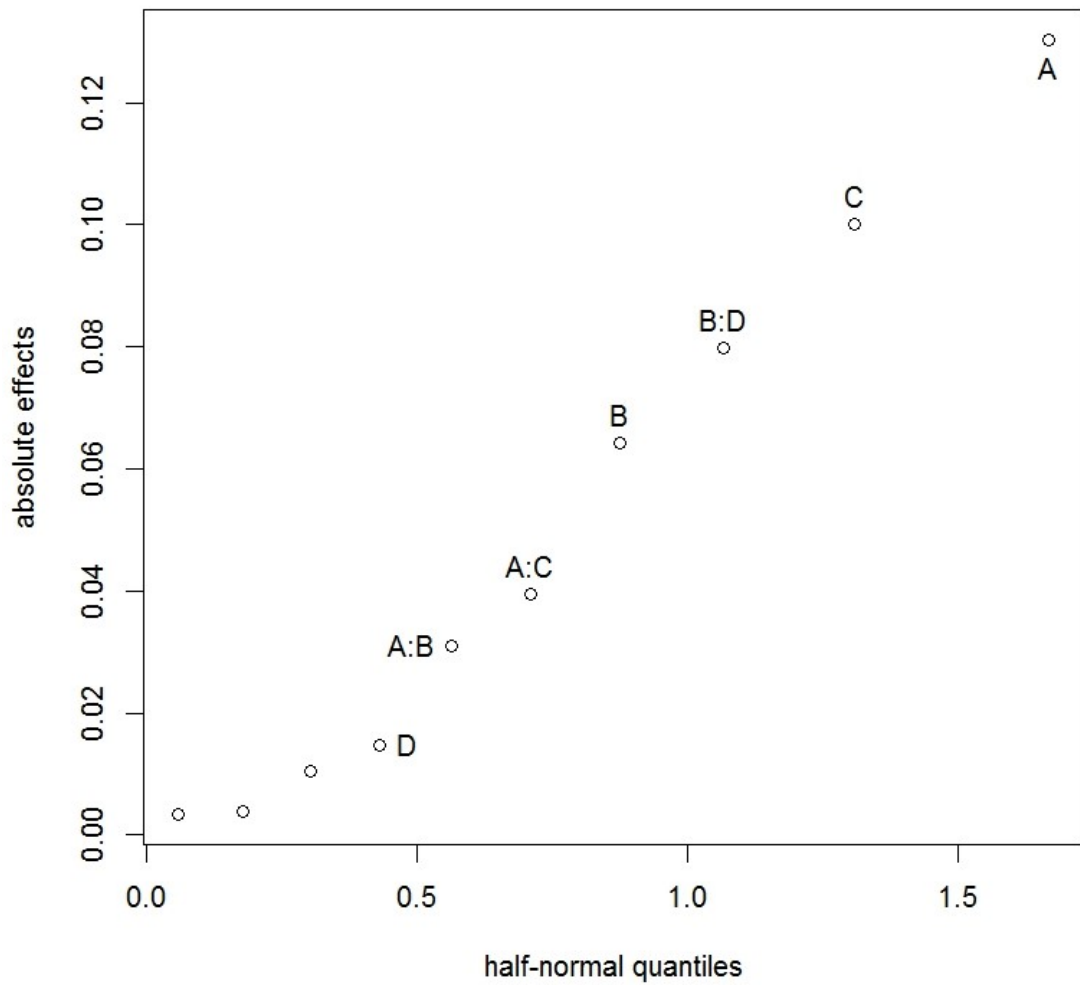


Figure 7: Half-normal Plot, Conversion Rate for CF_3 .

To maximize CR_T , we find the optimal level settings of the eight factors. By checking all the possible combinations of the eight factors, it is seen that by setting A, E and F to -, and B, C, D and E' to +, we have the maximal expected conversion rate of this system, which is 26.12%. Because F' does not appear in any significant terms, we can choose either setting according to other considerations.

Table 10: Design Matrix and Response Data, Conversion Rates for CF_3 .

A	B	C	D	CR_3
1	1	1	1	0.08730159
1	1	1	-1	0.05440000
1	1	-1	1	0.05546218
1	1	-1	-1	0.02073365
1	-1	1	1	0.06833333
1	-1	1	-1	0.10194175
1	-1	-1	1	0.03593750
1	-1	-1	-1	0.07812500
-1	1	1	1	0.14803150
-1	1	1	-1	0.12421384
-1	1	-1	1	0.09166667
-1	1	-1	-1	0.05255255
-1	-1	1	1	0.16722408
-1	-1	1	-1	0.21169036
-1	-1	-1	1	0.07954545
-1	-1	-1	-1	0.14820593

2.5 Simulated Example

In this section, we will demonstrate the idea of funnel testing with a more complicated example. Consider the conversion system shown in Figure 8. It consists of six pages: the first page is the landing page. All customers start visiting the conversion system with this page. The second page is called individual page. It is the page showing information for individual customers. The third page is called business page. It is the page showing information for business customers. There are three more pages, called Product 1 page, Product 2 page and Product 3 page respectively, for customers to make conversions for three different kinds of products. The linkage relationship between pages, as depicted in Figure 8, can be describes as follows. The customers always start with the landing page, where they have three choices: go to individual page, go to business page, or leave the system. If the customer goes to individual page, s/he then has three choices: go to Product 1 page, go to Product 2 page, or leave

the system. Similarly, if the customer goes to business page, s/he then also has three choices: go to Product 3 page, go to Product 2 page, or leave the system. Customers can convert on any of the Product pages, or leave the system. For customers on Product 1 page or Product 3 page, they have one more choice to go to Product 2 page.

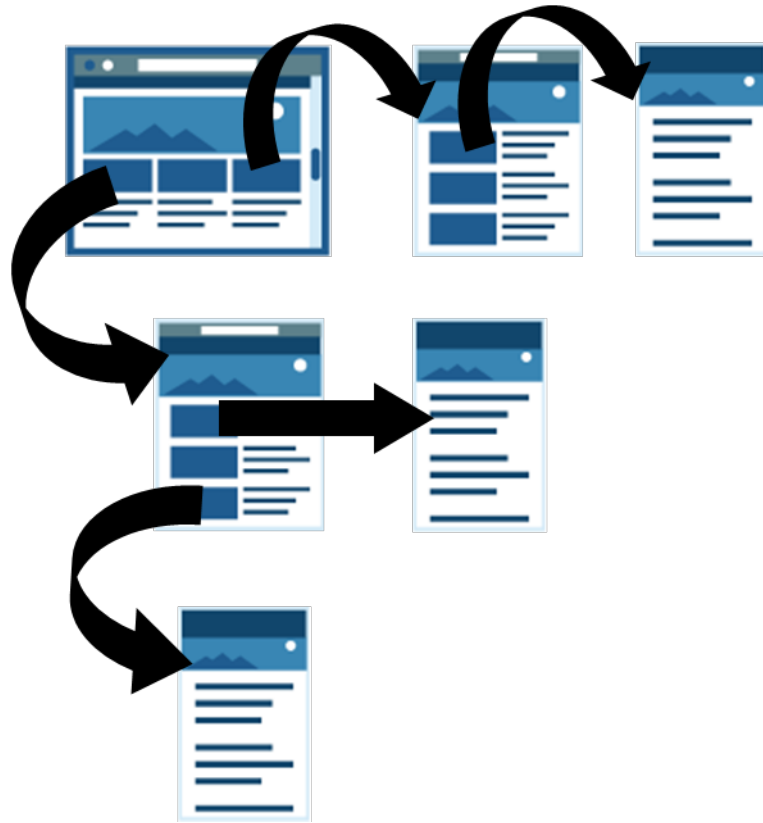


Figure 8: Simulated Example.

We start our analysis by representing this conversion system with the directed graph in Figure 9. The six pages are viewed as six vertices v_L , v_I , v_B , v_1 , v_2 and v_3 respectively, and the linkage relationships are viewed as eight directed edges denoted by e_1, \dots, e_8 . v_1 , v_2 and v_3 are marked in solid dots in contrast to others, meaning that they are conversion points.

The next step is to identify all the conversion funnels from the directed graph representation. By definition, a conversion funnel is a path from the landing page

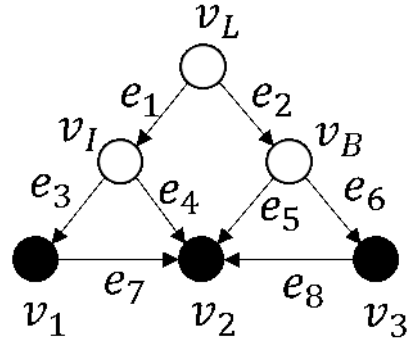


Figure 9: Directed Graph Representation, Simulated Example.

to a conversion point. For the conversion point v_1 , only the path $v_L \rightarrow v_I \rightarrow v_1$ connects v_L to it. We denote this conversion funnel by CF_1 . Similarly, only one path $v_L \rightarrow v_B \rightarrow v_3$ connects v_L to v_3 , which is denoted by CF_3 . For the last conversion point v_2 , there are four paths connect v_L to it: $v_L \rightarrow v_I \rightarrow v_2$, $v_L \rightarrow v_B \rightarrow v_2$, $v_L \rightarrow v_I \rightarrow v_1 \rightarrow v_2$, and $v_L \rightarrow v_B \rightarrow v_3 \rightarrow v_2$, which are represented by CF_{2I} , CF_{2B} , CF_{21} and CF_{23} respectively.

Suppose each page has two factors to be investigated (A and B for the landing page, C and D for individual page, E and F for business page, G and H for Product 1 page, I and J for Product 2 page, and K and L for Product 3 page), and each factor has two levels. A 2_{IV}^{12-6} design is used for the experiment. The design matrix is shown in Table 11. The defining relations of this design is $\mathbf{I}=\text{ABCG}=\text{ABDH}=\text{ACDEI}=\text{ACDFJ}=\text{ABEFK}=\text{BCEDFL}$. Each row of the matrix represents a version of the conversion system to be studied.

Table 11: Design Matrix, Simulated Example.

Run	A	B	C	D	E	F	G	H	I	J	K	L
1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	-1	1	1	1	-1	-1	-1

Continued on next page

Table 11 – continued from previous page

Run	A	B	C	D	E	F	G	H	I	J	K	L
3	1	1	1	1	-1	1	1	1	-1	1	-1	-1
4	1	1	1	1	-1	-1	1	1	-1	-1	1	1
5	1	1	1	-1	1	1	1	-1	-1	-1	1	-1
6	1	1	1	-1	1	-1	1	-1	-1	1	-1	1
7	1	1	1	-1	-1	1	1	-1	1	-1	-1	1
8	1	1	1	-1	-1	-1	1	-1	1	1	1	-1
9	1	1	-1	1	1	1	-1	1	-1	-1	1	-1
10	1	1	-1	1	1	-1	-1	1	-1	1	-1	1
11	1	1	-1	1	-1	1	-1	1	1	-1	-1	1
12	1	1	-1	1	-1	-1	-1	1	1	1	1	-1
13	1	1	-1	-1	1	1	-1	-1	1	1	1	1
14	1	1	-1	-1	1	-1	-1	-1	1	-1	-1	-1
15	1	1	-1	-1	-1	1	-1	-1	-1	1	-1	-1
16	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
17	1	-1	1	1	1	1	-1	-1	1	1	-1	-1
18	1	-1	1	1	1	-1	-1	-1	1	-1	1	1
19	1	-1	1	1	-1	1	-1	-1	-1	1	1	1
20	1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
21	1	-1	1	-1	1	1	-1	1	-1	-1	-1	1
22	1	-1	1	-1	1	-1	-1	1	-1	1	1	-1
23	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1
24	1	-1	1	-1	-1	-1	-1	1	1	1	-1	1
25	1	-1	-1	1	1	1	1	-1	-1	-1	-1	1
26	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1

Continued on next page

Table 11 – continued from previous page

Run	A	B	C	D	E	F	G	H	I	J	K	L
27	1	-1	-1	1	-1	1	1	-1	1	-1	1	-1
28	1	-1	-1	1	-1	-1	1	-1	1	1	-1	1
29	1	-1	-1	-1	1	1	1	1	1	1	-1	-1
30	1	-1	-1	-1	1	-1	1	1	1	-1	1	1
31	1	-1	-1	-1	-1	1	1	1	-1	1	1	1
32	1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1
33	-1	1	1	1	1	1	-1	-1	-1	-1	-1	1
34	-1	1	1	1	1	-1	-1	-1	-1	1	1	-1
35	-1	1	1	1	-1	1	-1	-1	1	-1	1	-1
36	-1	1	1	1	-1	-1	-1	-1	1	1	-1	1
37	-1	1	1	-1	1	1	-1	1	1	1	-1	-1
38	-1	1	1	-1	1	-1	-1	1	1	-1	1	1
39	-1	1	1	-1	-1	1	-1	1	-1	1	1	1
40	-1	1	1	-1	-1	-1	-1	1	-1	-1	-1	-1
41	-1	1	-1	1	1	1	1	-1	1	1	-1	-1
42	-1	1	-1	1	1	-1	1	-1	1	-1	1	1
43	-1	1	-1	1	-1	1	1	-1	-1	1	1	1
44	-1	1	-1	1	-1	-1	1	-1	-1	-1	-1	-1
45	-1	1	-1	-1	1	1	1	1	-1	-1	-1	1
46	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1
47	-1	1	-1	-1	-1	1	1	1	1	-1	1	-1
48	-1	1	-1	-1	-1	-1	1	1	1	1	-1	1
49	-1	-1	1	1	1	1	1	1	-1	-1	1	-1
50	-1	-1	1	1	1	-1	1	1	-1	1	-1	1

Continued on next page

Table 11 – continued from previous page

Run	A	B	C	D	E	F	G	H	I	J	K	L
51	-1	-1	1	1	-1	1	1	1	1	-1	-1	1
52	-1	-1	1	1	-1	-1	1	1	1	1	1	-1
53	-1	-1	1	-1	1	1	1	-1	1	1	1	1
54	-1	-1	1	-1	1	-1	1	-1	1	-1	-1	-1
55	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1
56	-1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1
57	-1	-1	-1	1	1	1	-1	1	1	1	1	1
58	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1
59	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	-1
60	-1	-1	-1	1	-1	-1	-1	1	-1	-1	1	1
61	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	-1
62	-1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	1
63	-1	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1
64	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	-1

Take the landing page and its corresponding factors A and B for example. Factor A may represent the choice of the header, which has two candidates with version 1 denoted by level + and version 2 denoted by -. Similarly, B can be the choice of the main picture on the landing page, with + being version 1 and - being version 2. Other factors of other pages can be interpreted similarly. Generally, each factor represents one element of its corresponding page that we want to investigate. In this example, each element has two candidate versions and we want to decide which one is better. As discussed in section 1, elements can be headers, banners, texts, pictures, etc.

The experiment is done by simulations. For each simulation, we first choose a

version of the conversion system from the 64 candidates in the design table with equal probability, and then simulate customers' behavior in the chosen system. The customers always start with the landing page. Recall that, on the landing page, they have three choices: go to individual page, go to business page, or leave the system. We simulate their decisions in the following way: they go to individual page with probability t_{LI} , where t_{LI} represents the transition probability from the landing page to individual page; they go to business page with probability t_{LB} . If the simulated decisions of the above two statements are both negative, customers leave the system. If the simulated decisions are both positive for the two statements, the customers go to either page with equal probability. For customers on individual page, we simulate their decisions among the three choices in a similar way: they go to Product 1 page with probability t_{I1} , and go to Product 2 page with probability t_{I2} . Similarly, for customers on business page, we suppose they go to Product 3 page with probability t_{B3} , and go to Product 2 page with probability t_{B2} . For customers on either Product 1 page or Product 3 page, they make conversions with probability c_1 or c_3 respectively. If they do not make conversions, they can further go to Product 2 page with probability t_{12} or t_{32} respectively. Otherwise, they leave the system. Finally, for customers on Product 2 page, they can either make conversions or leave the system. The probabilities for them to make conversions are c_{2I} , c_{2B} , c_{21} and c_{23} respectively, depending on the conversion funnels they are from. The simulation is terminated when the customers leave the system or a conversion is made.

All the decision probabilities used in the simulations are listed in Table 12. If the calculated probabilities are less than zero, we suppose they are zero in the simulation. The choice of the decision probabilities is somewhat arbitrary but with the following rationalization. Take t_{LI} for example. We have $t_{LI} = 0.5 + 0.1A$, which is a function of the main effect A of the landing page. According to the assumption, A has two levels denoted by + and -. Therefore, the transition probability from the landing page

Table 12: Functions for Decision Probabilities, Simulated Example.

$t_{LI} = 0.5 + 0.1A$
$t_{LB} = 0.2 - 0.1B$
$t_{I1} = 0.1 + 0.1A - 0.1C + 0.1BC$
$t_{I2} = 0.05D - 0.05AD$
$t_{B3} = 0.1 + 0.1E - 0.1F + 0.1C - 0.1EF$
$t_{B2} = 0.1A - 0.1AE$
$t_{12} = 0.05 - 0.05DH + 0.05C - 0.05AG$
$t_{32} = 0.05K - 0.05EL$
$c_1 = 0.1 + 0.1G - 0.1CG$
$c_3 = 0.1 + 0.1K - 0.1FL + 0.05BK - 0.05AF$
$c_{2I} = 0.1 + 0.1I - 0.1J - 0.1CJ + 0.1BI$
$c_{2B} = 0.1 + 0.1J - 0.1IJ + 0.05E - 0.05AJ$
$c_{21} = 0.05I - 0.05DJ + 0.05H + 0.05CI$
$c_{23} = 0.05J - 0.05E - 0.05FK + 0.1IJ$

to individual page is 0.6 ($=0.5+0.1$) if factor A is set to level +, and 0.4 ($=0.5-0.1$) if A is set to -. Similarly, for t_{12} , we have $t_{12} = 0.05 - 0.05DH + 0.05C - 0.05AG$. It is a function of the main effect C and 2fi's DH and AG, which involves five factors A, C, D, G and H. Among the five factors, A is from the landing page, C and D are from individual page, and G and H are from Product 1 page. To justify the choice of these five factors, note that t_{12} is the transition probability from Product 1 page to Product 2 page, which is part of CF_{21} , the path $v_L \rightarrow v_I \rightarrow v_1 \rightarrow v_2$. Before making this decision, customers have gone through the landing page, individual page and Product 1 page along the path. Therefore, it is assumed that only factors on these three pages can affect this decision. The 2fi DH in t_{12} can be interpreted as follows: factor D on individual page and factor H on Product 1 page will jointly affect the customers decision as whether to go to Product 2 page from Product 1 page. Other 2fi's in Table 12 can be interpreted similarly. By changing the level settings of the five factors, t_{12} can be as high as 0.2 or as low as 0.

We repeat the simulation for 10,000,000 times and record the conversions for each

funnel separately. The calculated conversion rates are given in Table 13. Note that CF_{21} and CF_{23} both have no conversions in the simulated results.

Table 13: Conversion Rates for Different Funnels, Simulated Example.

Run	CR_1	CR_3	CR_{2I}	CR_{2B}	CR_{21}	CR_{23}
1	0.011559	0.00073	0	0	0	0
2	0.011608	0	0	0	0	0
3	0.010784	0	0	0.002698	0	0
4	0.011201	0.00253	0	0	0	0
5	0.011635	0.002146	0	0	0	0
6	0.011559	0.003635	0	0	0	0
7	0.011191	0	0	0.001426	0	0
8	0.012039	0.001333	0	0	0	0
9	0	0	0	0	0	0
10	0	0.002268	0	0	0	0
11	0	0	0	0.001311	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0.002778	0	0
16	0	0	0	0	0	0
17	0	0.001971	0	0	0	0
18	0	0.030319	0	0	0	0
19	0	0	0	0.008195	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0

Continued on next page

Table 13 – continued from previous page

Run	CR_1	CR_3	CR_{2I}	CR_{2B}	CR_{21}	CR_{23}
22	0	0.010542	0	0	0	0
23	0	0.003757	0	0.003858	0	0
24	0	0.00412	0	0	0	0
25	0.059645	0	0	0	0	0
26	0.062159	0.006101	0	0	0	0
27	0.0621	0	0	0.003884	0	0
28	0.060896	0	0	0	0	0
29	0.061177	0	0	0	0	0
30	0.061716	0.018862	0	0	0	0
31	0.06134	0	0	0.008334	0	0
32	0.061483	0	0	0	0	0
33	0	0	0.004004	0	0	0
34	0	0.003807	0	0	0	0
35	0	0.003262	0.018645	0	0	0
36	0	0	0.003678	0	0	0
37	0	0.000699	0	0	0	0
38	0	0.012743	0	0	0	0
39	0	0.001638	0	0	0	0
40	0	0	0	0	0	0
41	0	0	0.010877	0	0	0
42	0	0.00712	0.011838	0	0	0
43	0	0	0	0	0	0
44	0	0	0	0	0	0
45	0	0	0	0	0	0

Continued on next page

Table 13 – continued from previous page

Run	CR_1	CR_3	CR_{2I}	CR_{2B}	CR_{21}	CR_{23}
46	0	0.002403	0	0	0	0
47	0	0	0	0	0	0
48	0	0	0	0	0	0
49	0	0.007058	0.010615	0	0	0
50	0	0.012452	0	0	0	0
51	0	0	0.010327	0	0	0
52	0	0	0	0	0	0
53	0	0.002347	0	0	0	0
54	0	0	0	0	0	0
55	0	0.005054	0	0	0	0
56	0	0.005172	0	0	0	0
57	0	0	0.002712	0	0	0
58	0	0	0.002895	0	0	0
59	0	0	0.00266	0	0	0
60	0	0	0.002505	0	0	0
61	0	0	0	0	0	0
62	0	0.007343	0	0	0	0
63	0	0	0	0	0	0
64	0	0	0	0	0	0

The total conversion rate is assumed to be:

$$CR_T = 2CR_1 + 2CR_3 + CR_{2I} + CR_{2B} + CR_{21} + CR_{23}.$$

We model the conversion rate for each conversion funnel separately. Recall that the

first step in modeling the conversion rate is to identify its related factors. For CR_1 , it is the conversion rate for CF_1 , which consists of three pages: the landing page, individual page and Product 1 page. Therefore, the six factors (A, B, C, D, G and H) related to these three pages are considered in modeling CR_1 . Similarly, we can find the related factors for other conversion rates, all of which are listed in Table 14.

Table 14: Conversion Rates and Related Factors, Simulated Example.

CR_1	A B C D G H
CR_3	A B E F K L
CR_{2I}	A B C D I J
CR_{2B}	A B E F I J
CR_{21}	A B C D G H I J
CR_{23}	A B E F I J K L

The model for each conversion rate is then built with respect to its related factors listed in Table 14. We consider the linear regression model with only the main effects and 2fi's and use the same method as in the toy example. The fitted model for each conversion rate is shown in Table 15. CR_{21} and CR_{23} are both zero because CF_{21} and CF_{23} have no conversions for all 64 versions of the conversion system in the simulated results.

Table 15: Fitted Models for Each Conversion Rate, Simulated Example.

CR_1	=	$0.0091 + 0.0091A - 0.0062B - 0.0062C - 0.0091G - 0.0062AB - 0.0062AC + 0.0091AG$
CR_2	=	$0.0025 + 0.0017E - 0.0016F + 0.0013K - 0.0016EF - 0.0016FL$
CR_{2I}	=	$0.0013 - 0.0013A + 0.0013D - 0.0013AD$
CR_{2B}	=	$0.0005 + 0.0005A - 0.0005E + 0.0005F - 0.0005AE + 0.0005AF - 0.0005EF$
CR_{21}	=	0
CR_{23}	=	0

In theory, Product 2 page can have four different versions, which would give us six

more factors I' , J' , I'' , J'' , I''' and J''' in the optimization procedure. This is because v_2 appears in four different funnels, and the pages before v_2 are different for each funnel. However, the factors related to v_2 do not appear in any of the models in Table 15. Therefore, such consideration become unnecessary. Finally, by putting the functions in Table 15 together, we have the model for the total conversion rate to be:

$$\begin{aligned} CR_T = & 0.025 + 0.0174A - 0.0124B - 0.0124C + 0.0013D + 0.0029E - 0.0027F \\ & - 0.0182G + 0.0026K - 0.0124AB - 0.0124AC - 0.0013AD \\ & - 0.0005AE + 0.0005AF + 0.0182AG - 0.0037EF - 0.0032FL. \end{aligned}$$

To maximize CR_T , we try all possible combinations of the nine factors involved in the above model. The maximal CR_T value is achieved by setting A, E, K and L to +, and B, C and F to -. The level settings of D and G does not affect the value of CR_T if the remaining 7 factors are chosen as above. The other 3 factors that do not appear in the above model can be set based on other considerations.

2.6 Conclusions

In this work, we propose a framework and approach to analyze a system of pages that relate to the conversion of interest based on the concept of conversion funnels. Directed graph is used to represent the system and identify all the conversion funnels. Fractional factorial design is used to conduct the experiment. An analysis strategy consists of modeling the conversion rate of each funnel separately and putting them together in the total conversion rate to do optimization.

Compared with traditional analysis strategies, funnel testing will help understand the structure of the conversion system better through the directed graph representation. For conversion funnels with many pages, traditional method can only study one page at a time. Multiple experiments have to be conducted in order to optimize such conversion. With funnel testing, users can analyze all the pages in a conversion funnel in one experiment. With a good choice of fractional factorial design, this

will save a significant amount of effort in experiment. Like the simulated example, conversion systems can become extremely complicated with multiple conversions. In traditional method, each conversion is studied separately. The overall optimality of the system can hardly be achieved. In funnel testing, we consider the total conversion of the system in the optimization procedure. The maximal total conversion rate is guaranteed with the proposed method.

So far the analysis strategy is developed in the context of specific examples. But the underlying ideas are general. It is our plan to further develop this framework into a general methodology, and address the issues of graph theory and design of experiment.

CHAPTER III

UNCERTAINTY QUANTIFICATION OF SOLAR DIFFUSE IRRADIATION ON INCLINED SURFACES FOR BUILDING ENERGY SIMULATION

3.1 Introduction

Over the last four decades, scientists and engineers have made significant progress in building energy simulation. Complex buildings now can be described with computational models that simulate realistic performance such as energy consumption. The role of simulation has been firmly established in the architecture, engineering and construction industry. It is used to inform decisions at scales as large as national energy policy measures and as small as the selection of shading devices of a residential house. Questions about the accuracy of computational models to support such decisions have given rise to attempts to quantify the uncertainty in their outcomes. Based on an appropriately quantified uncertainty, one will be able to decide what level of confidence in the simulation results is warranted.

The concept of uncertainty is not unfamiliar to building performance modelling and simulation. It has been introduced since the early 1980s when most efforts were dedicated to model development and validation. One of the best known projects was the Building Energy Simulation Test (BESTEST) (Judkoff et al. 1983) in which uncertainties of model input parameters were regarded as the major sources that complicated model validation. Although uncertainty was acknowledged in this project, BESTEST did not confront the inherent complexity of uncertainties. In stead, it emphasized the control of uncertainty up front so that uncertainty could be neglected in subsequent analysis. Uncertainties with respect to model input parameters were

addressed directly in the PASSYS project (PASsive Solar Components and SYStems Testing). PASSYS adopted sensitivity analysis (SA) techniques in the model validation process (Strachan 1993) because SA can tell the relationships between model inputs and outputs that otherwise are not apparent for a complicated engineering model. However, uncertainty quantification (UQ) of simulation results did not appear in model application contexts, such as for energy-efficient building design, before the 2000s. Some of the pioneers include Macdonald and Strachan (2001) who incorporated the UQ into ESP-r, which is the leading building simulation tool in Europe, and analysed the effect of uncertainty over building design process. In 2002, De Wit and Augenbroe (2002) initiated the integration of UQ with risk analysis in a decision-making context. This study showed how a different decision would have been made for choosing between design alternatives if the decision maker were informed about uncertainties in the predictions. More recent work by Heo et al. (2013) extended the application of UQ to the support of risk-conscious decision making in building design and retrofit when decisions are driven by return on investment expectations, or when energy savings guarantees are part of a performance contract.

A widely accepted definition of uncertainty was given by Walker et al. (2003), who defined uncertainty as “being any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system.” When models are used to predict building energy consumption at its design stage, the predictions differ from true values for a variety of reasons. First of all, key model inputs such as weather conditions, building material properties, and operation schedules are usually not known with certainty or are subject to changes in real operation conditions. Another source of discrepancy is in the process of solving the mathematical models with numerical methods, which is strictly the concern of verification. Besides, inevitable errors in the measurements and (in most cases) the uncontrollability of the experiment prohibit us from observing the true values. Furthermore, models by definition ignore to

some degree, and in almost all cases simplify the physical processes of the real world. Model discrepancy associated with ignorance and simplification is called model form uncertainty (Oberkampf and Roy 2010). Other factors that account for discrepancies between model predictions and physical measurements include human errors in preparing the inputs and processing the outputs. We typically refer to them as modeler’s bias or error.

The UQ of model input parameters is relatively straightforward because inputs by definition are observable quantities, such as thermal conductivity of a certain type of brick. Once sufficient data are collected, the uncertainty in these input parameters can be characterized by probabilistic distributions (e.g., normal distribution) with standard statistical methods (Bedford and Cooke 2001). UQ could become difficult if the parameters are time series or if correlations are apparent among parameters. One example is the uncertainty in weather conditions. A recent paper by Lee et al. (2012) proposes to characterize uncertainty in the weather variables with a vector auto-regressive process and introduces algorithms to generate stochastic weather from historical meteorological years. Methods are also well established to propagate uncertainties from model input parameters to the outcomes. For example, sample-based methods are extensively used for parameter uncertainty propagation through complex engineering models (Helton and Davis 2003). This approach first draws samples of uncertain model inputs from their distributions; each sample is then used as input into the model to obtain one realization of a model prediction; consequent uncertainties of model predictions are quantified by aggregating these realizations.

Moreover, the model itself is inherently inaccurate, i.e., a simulation result delivers a biased view of reality even if the values of the model input parameters are assigned the true values. Because the bias in the predicted quantities cannot be exactly known, its existence leads to uncertainty in the predictions. We identify this type of uncertainty as model form uncertainty, which is also referred to as “model

structural uncertainty” or “model uncertainty” in short. Model form UQ that estimates model bias (also referred to as “model discrepancy”) is a thorny problem for the modelling and simulation community (Cooke 2013, NRC 2012) despite a large methodology investment from the nuclear research sector (Oberkampf et al. 2003, Helton et al. 2012, Roy and Oberkampf 2011). According to the definition of uncertainty by Walker et al. (2003), a comprehensive UQ must assess all origins of uncertainty to give a full reflection of model prediction uncertainty. As for building energy simulation, model form UQ, has not yet received enough attention. Therefore, the authors are motivated by the assumption that model bias might be an important contributor to the overall discrepancy between the predicted and the actual use of energy by a building.

The distinction between model form uncertainty and parameter uncertainty is meaningful only if a modelling system has been specified. A computational model can be represented as a function $f(\bullet)$ that maps an input vector $X = (x_1, \dots, x_q)$ into an output $y = f(X)$. Uncertainty in X is defined as parameter uncertainty, whereas uncertainty in the function $f(\bullet)$ is defined as model form uncertainty. The difference between these two types of uncertainty is clear yet can be blurred or miscommunicated if the function and the input parameter space have not been specified. In particular, model form uncertainty in one modelling system could be the parameter uncertainty in another modelling system or vice versa. For example, air infiltration is one of many physical processes involved in modelling a full energy system of buildings. If the full building energy system accounts for the effect of air infiltration through a prescribed manner as one element of the input vector X , e.g., hourly infiltration rate, the uncertainty of the infiltration rate is purely parameter uncertainty. In contrast, air infiltration itself may form a subsystem, which is explicitly modelled as a function $g(\bullet)$ with a subset of input parameters such as leakage area of exterior walls. Regarding the second building energy system, the uncertainty of infiltration rate thus results

from not only the parameter uncertainty of the leakage area but also the model form uncertainty of the function $g(\bullet)$.

For complex systems such as buildings, model form UQ at the full system scale is intractable. This is primarily caused by the difficulties in gathering high-quality data in terms of both model input parameters and model outputs. An effective way of conducting UQ for complex systems is to develop a hierarchical structure that breaks down a complex system into subsystems and then into units. In current dynamic building simulation software, such hierarchy is indeed the structure of the models on which the tools are based. For example, EnergyPlus (EnergyPlus 2012) is a collection of modules that work together to calculate the final outcomes. Each module performs a specific function that touches only relatively few physical processes. For instance, the calculation of solar irradiation on building surfaces deploys sky models formulated by a set of algebraic equations whose outcomes affect the boundary conditions of other modules. Fortunately, at the module scale we can in many cases collect high quality physical observations. Hence, UQ of the building energy model should exploit this hierarchical composition, focusing first on the lowest-level components and moving successively to more complex levels.

The research methods for model form UQ fall in the domain of statistical inference. Recent approaches for model form UQ are categorized into two groups, i.e., a classical frequentist approach (Oberkampf and Roy 2010, Hills and Dowding 2008) and Bayesian approach (Kennedy and O'Hagan 2001, Qian and Wu 2008, Chen et al. 2008). The former draws probability distributions of model bias based on statistical data analysis, whilst the latter assumes that a modeller has prior knowledge about the model bias. The experimental data are then used to update the prior distributions and obtain posterior distributions. An extensive discussion and comparison of the two approaches can be found in Hills et al. (2008). In general, when there is sufficient data available the two approaches will converge to the same result (Bedford

and Cooke 2001). This paper uses the frequentist approach for model form UQ in subsystems or units not only because sufficient data are typically attainable at these levels, but also because the frequentist approach is firmly established in the building domain and relatively easier to use than the Bayesian approach.

In this paper, we introduce a generic framework for the quantification of model form uncertainty based on physical observations in section 3.2. In section 3.3, we demonstrate this framework in the UQ of the well-known sky model developed by Perez et al. (1990), which predicts solar diffuse irradiation on inclined surfaces. We choose the Perez model as the case study for two reasons. First, a number of studies have shown that this model performs adequately in some circumstances (Loutzenhiser et al. 2007), yet exhibits noticeable discrepancies in other circumstances (Diez-Mediavilla et al. 2005). Second, although solar irradiation is one of the most important boundary conditions for many components in the building model, we have not seen an effective approach that quantifies the model uncertainty, with the goal to improve the predictions given by the Perez 1990 model. In section 3.4, we explore the effect of model form uncertainty on the energy predictions obtained with whole building simulations. Section 3.5 presents some final observations.

3.2 Approach

Statistical inference for model form uncertainty quantification is based on the observations of model inadequacy under the conditions specified by the model input parameters. Because the estimate of model inadequacy falls in the category of model validation, model form UQ and validation are inherently interrelated. A discussion about their relationship can be found in a recent report (NRC 2012). In contrast to model validation that evaluates model validity for the experimental test conditions (referred to as samples), model form UQ is concerned with model predicative capability for samples that are not, or have yet to become, observable. This leads to the

important foundation for the development used in this study, that model validation results are used as the observations based on which model discrepancy and the associated uncertainty for new application samples is statistically inferred. This is explored in detail below.

Let $\eta(\mathbf{u})$ be the output of a computational model when the variable inputs take values $\mathbf{u} = (u_1, \dots, u_k)$. Model discrepancy refers to the difference between $\eta(\mathbf{u})$ and the true value (but unknown) T under the conditions specified by \mathbf{u} (ASME 2009). Let us consider physical observations y_{obs} that approximate the true value T . In this paper, we regard η as a deterministic model, i.e., $\eta(\mathbf{u})$ has a fixed value for a given \mathbf{u} . We can now relate model form uncertainty $diff(\mathbf{u}, \mathbf{v})$ to the simulation model outcome $\eta(\mathbf{u})$, and the physical observations y_{obs} in the following equation:

$$diff(\mathbf{u}, \mathbf{v}) = y_{obs} - \eta(\mathbf{u}). \quad (29)$$

Note that we have added a new as yet undefined variable \mathbf{v} to the expression of the model form uncertainty. This is motivated by the fact that the new variable \mathbf{v} is necessary if model discrepancy displays significant correlations with some other (undetected) variables. Typically, adding new variables needs a better understanding of the shortcoming of the model at hand but relies on an effective method for physical experimental designs. As a result, adding new variables will lead to better representation of the model inadequacy and eventually enhance model predicative capability. We take this formulation of model form uncertainty not only for computational convenience but also because the results of $diff(\mathbf{u}, \mathbf{v})$ are easy to interpret since they relate to the same physical units as the model outputs. In fact, the assessment of is only an intermediate step. As suggested by (29) the model output will be modified by $diff(\mathbf{u}, \mathbf{v})$, so that the modified results $\eta(\mathbf{u}) + diff(\mathbf{u}, \mathbf{v})$ will approach y_{obs} .

It is also worthy of mentioning the authors' earlier work in which y_{obs} is replaced by model outcomes obtained with a high fidelity model. An application of this approach is shown in (Sun et al. 2014) for the UQ of building microclimate variables. This

method presumes the existence of a high fidelity model whose model inadequacy is of secondary order in comparison with that of the low fidelity model under study. Given the features of building simulation, we found it was useful to explore the high fidelity model as an option to quantify the model form UQ of the low fidelity model. As aforementioned, a complete building energy model consists of many submodels. A submodel that has a reduced order implementation in the building energy model could also exist in a higher order implementation, e.g. to deliver the primary quantity of interest in another domain. For example, Sun et al. (2014) used a high-order meteorological model as the high fidelity model to quantify the uncertainty in a reduced order model of building microclimate. The key difference in two models was the fact that the reduced order model ignored the urban heat island (UHI) effect.

If a high fidelity model does not exist or y_{obs} is readily accessible, we turn to the approach that is the focus of this paper. Model form UQ based on physical observations can be summarised by the six steps as follows:

Step 1. Specify the computational model of interest,

Step 2. Obtain physical observations and assess measurement errors,

Step 3. Provide statistical evidence for model inadequacy,

Step 4. Develop a statistical model from the training data,

Step 5. Assess the predictive capability of this statistical model using validation data,

Step 6. Update the original computational model and quantify the uncertainties in its predictions.

We demonstrate the six steps in the uncertainty quantification of Perez model in the next section.

3.3 *Uncertainty Quantification of the Perez Model*

In building energy models, solar irradiation appears in the boundary condition of external building surfaces. In case of transparent external surfaces it also appears in the boundary conditions of internal surfaces. In rare cases, solar irradiation is directly measured for the orientation and tilt of every surface that appears in a model. In most current building energy studies this is not the case, hence it is derived from other directly measured quantities and sky condition parameters. The global irradiation on tilted surfaces is calculated by the summation of three components as follows:

$$I_g^s = I_{bn} \cos \alpha + I_d^s + I_r^s,$$

where I_g^s is the global solar irradiation on a tilted surface with tilt angle S , I_{bn} is the direct normal irradiation, α is the solar incident angle on the surface, and I_d^s is the diffuse solar irradiation on the tilted surface, and I_r^s is the ground reflected irradiation.

Different models were developed to derive solar irradiation on surfaces with any tilt angle and orientation from data for horizontal surfaces. Among the three components, calculating the direct irradiation is purely geometric and thus straightforward and identical among the models. With respect to the ground reflected irradiation, most studies adopt the isotropic assumption with which the ground reflected irradiation is estimated from the following equation:

$$I_r^s = \rho \frac{1 - \cos s}{s}, \quad (30)$$

where ρ is the ground albedo. The main difference between the models lies in the way of modelling the sky diffuse irradiation component. The model of sky diffuse irradiation started from the simple isotropic sky model assumptions (Liu and Jordan 1961), and gradually transformed into anisotropic models advanced by Gueymard (1987), Perez et al. (1990), and Muneer (2004). An extensive literature review of

such models and their comparison can be found in (Noorian et al. 2008, Gueymard 1987). Among the different models, the most notable one was developed by Perez et al. (1990), as evidenced by its wide application in solar engineering (Yang et al. 2007) and building energy simulation software such as EnergyPlus (2012). We refer to this particular model as the *Perez model* in this paper. Although the core of the Perez model pertains to the modelling of the sky diffuse component, the model performance is commonly evaluated for the global irradiation on tilted surfaces (Perez et al. 1990, Gueymard 2009). Thus, the ground reflected component calculated from (30) is regarded as an integral part of the broad Perez model. To be consistent with previous studies including Perez’s own work, we quantify the uncertainty of the Perez model in predicting the global solar irradiation on tilted surfaces.

3.3.1 Perez Model Description

This section offers a brief overview of the Perez model (Perez et al. 1990) in terms of modelling the sky diffuse irradiation. It postulates a simplified sky representation, in which the sky hemisphere is composed of a circumsolar disc and horizon band on an isotropic background. Each element has a parametric representation of solar irradiation with multiple coefficients, whose values were obtained through statistical regression analysis. Figure 10 shows its input-to-output relationship. The Perez model takes horizontal solar irradiation I_h , direct normal solar irradiation I , solar azimuth angle θ , solar altitude angle ϕ , surface tilt angle S , and surface azimuth angle ψ as input variables. It calculates diffuse solar irradiation from the sky horizon band $I_{horizon}$, the sky dome I_{dome} , the circumsolar region $I_{circumsolar}$, and total diffuse irradiation from the sky I_{sky} , the latter being the summation of the three components.

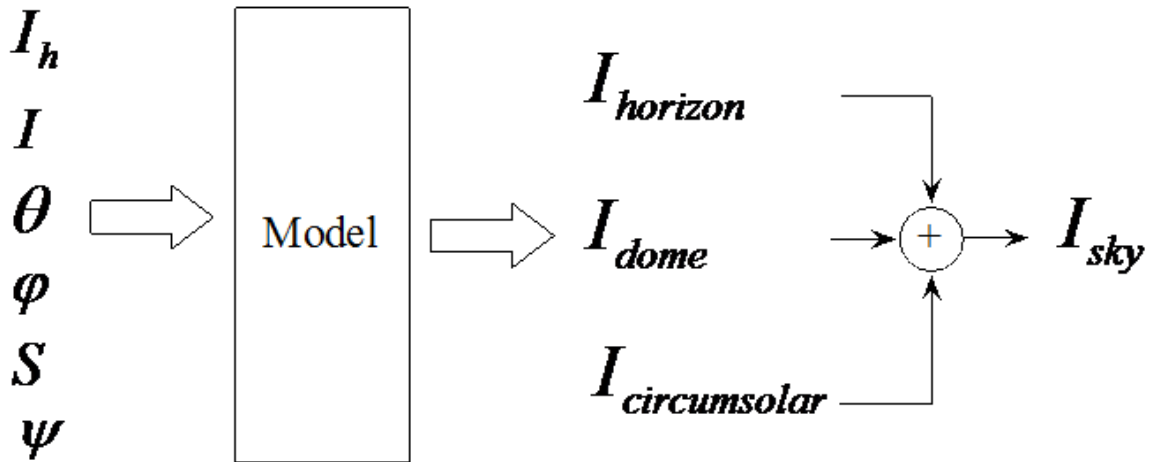


Figure 10: Input-to-output Relationship, Perez (1990) Sky Irradiation Model.

3.3.2 Experimental Data

We obtained detailed measurement data from a station of Solar Radiation Monitoring Laboratory (SRML) in Eugene, Oregon. The data include simultaneous measurements of (1) global solar irradiation on horizontal surfaces, (2) diffuse solar irradiation on horizontal surfaces, (3) direct normal solar irradiation, (4) ground reflected solar irradiation, (5) global solar irradiation on south tilted surfaces at 30° , 45° , and 90° , and (6) global solar irradiation on a north vertical surface. Global solar irradiation (i.e., the sum of direct and diffuse) is measured with the Eppley Precision Spectral Pyranometer (PSP). Diffuse solar irradiation is measured with the shaded Eppley PSE with automatic trackers. Direct normal is measured with Eppley Normal Incident Pyrheliometer (NIP). The ground reflected solar irradiation is measured with the Eppley PSP facing the ground. All devices are yearly calibrated. Table 16 provides an overview of the specifications of the instruments.

The data were collected in 2011 at 5 minute intervals. Before we conduct the analysis, we first derive hourly measurements from the raw data. The hourly aggregation reduces the short-term variation of the measurements attributable to small cloud variations and random errors from instrument measurements. Hourly data also

Table 16: Specifications of Instruments.

Instrument	Specifications
Eppley PSP	<i>Cosine Response:</i> $\pm 1\%$ from normalization 0-70° zenith angle; $\pm 3\%$ 70-80° zenith angle. <i>Accuracy:</i> The absolute accuracy of calibration is about $\pm 3-4\%$. The relative accuracy of calibration is about $\pm 2\%$.
Eppley NIP	<i>Accuracy:</i> The absolute accuracy of calibration is about $\pm 2\%$. The relative accuracy of calibration is about $\pm 1\%$.

matches with the temporal resolution of weather variables used for building simulation such as EnergyPlus (EnergyPlus 2012). We ignored measurements that are less than 50W globally on a horizontal surface because measurements of low solar irradiation are often subject to high measurement error (Reda 2011). Model inputs for solar angles θ and ϕ are computed according to the ASHRAE Handbook of Fundamentals (ASHRAE 2009). Uncertainties in the angles are very minimal and are hence ignored.

There is a need to detect systematic measurement errors before performing the UQ. Undoubtedly, undetected systematic measurement errors will contaminate the entire UQ results. In our case, we compare three independent measurements on horizontal surfaces to estimate the quality of the measurements. In principle, the following equation holds:

$$I_g^h = I_{bn} \cos \alpha + I_d^h, \quad (31)$$

where I_g^h is the global solar irradiation on horizontal surface, and I_d^h diffuse solar irradiation on horizontal surface. For horizontal surfaces, every element in (31) except α , which is computed with minimum uncertainty, is directly measured. Figure 11 compares the two sides of (31) and depicts the comparisons and linear regression analysis. It shows that direct measurements of global horizontal irradiance match well with the calculations from the beam and diffuse horizontal irradiance components. As suggested by the regression equation, the average difference between the

directly measured horizontal global irradiation and that calculated by diffuse and direct components is 5.13 W/m^2 , which equals to 1.5% of the average global irradiation. Therefore, systematic errors in the measurements are not significant. The major measurement errors are attributed to random errors.

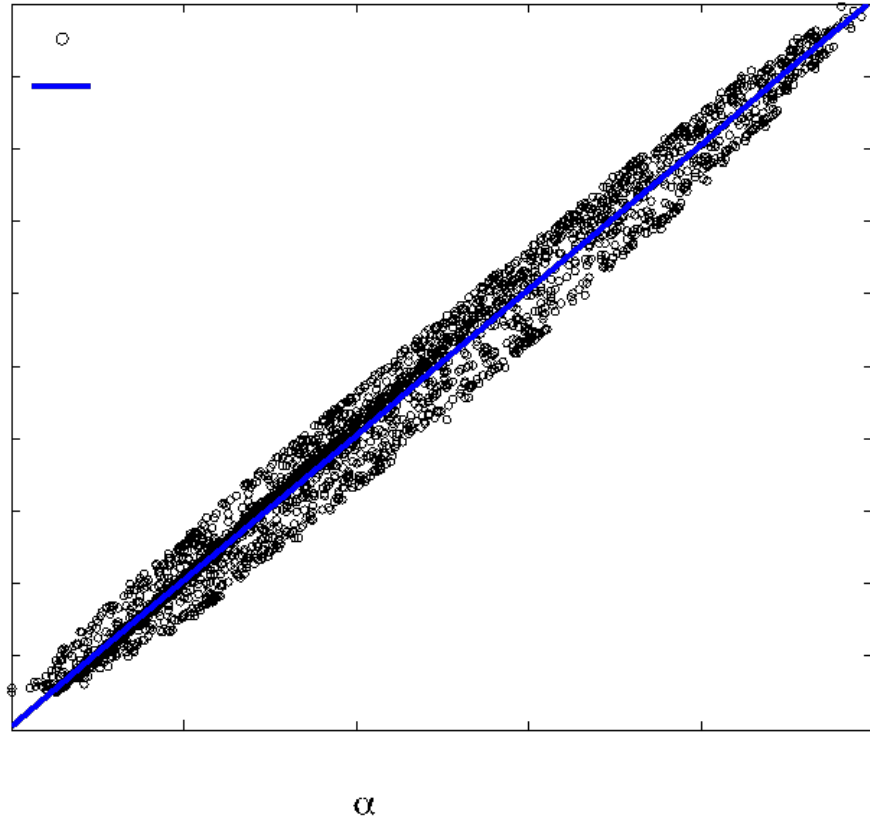


Figure 11: Direct Measured and Derived Global Horizontal Irradiation from Two Components.

3.3.3 Statistical Evidence for Model Inadequacy

We measure the discrepancies between the model predictions and measurements by the two following statistical indicators: Mean Bias Error (MBE) and Root Mean

Square Error (RMSE):

$$MBE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{N},$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2},$$

where y_i is the i th measurement, \hat{y}_i is the i th model prediction, and N is the total number of observations.

Perez (1990, p. 12) claimed that his model is location independent. However, multiple sources in the literature show that the model performance is not uniform across various locations. Many researchers have studied the Perez model, and used their own measurements to compare with the predictions from the Perez model. We collected the results from nine different researchers (Noorian et al. 2008, Utrillas et al. 1991, Diez-Mediavilla et al. 2005, Evseev and Kudish 2009, Igawa et al. 2004, Gueymard 2009, David et al. 2013, Loutzenhiser et al. 2007, Kambezidis et al. 1994), each with one or multiple studies (with different surface orientations and/or different surface tilt angles) conducted in distinct locations. We also used the Eugene data mentioned above and Perez's own results. In total we have data from eleven different sources. Among these studies, the effect of ground reflected irradiation has been considered in two ways. Noorian et al. (2008), Utrillas et al. (1991), and Igawa et al. (2004) used artificial horizons to protect the pyranometers against ground reflection, which is called *Group I* in our analysis. The rest of the researchers calculated the ground reflected irradiation from the isotropic model given by (30), which is called *Group II*. To examine the effect of ground reflected irradiation, we shall compare Perez results with Group I and II separately.

The box-plot of the percent MBE and RMSE of global solar irradiation on tilted surfaces for each location is given in Figure 12. The left three plots are from Group I; the 4th one in the middle is from Perez (1990); and the right seven plots are from

the Group II. As we compare the results from Perez (1990) with those from Group I and Group II, we notice a significant fluctuation of the mean value (the black bar in the middle of the box) across locations for both MBE and RMSE. Since the box' in the box-plot represents the middle half of the data range, the length of the box' is a measure of variation within a location. We found that, for locations with multiple data points, their within-location variations are quite large, and the magnitudes of the variations fluctuate. After omitting the plot where there is only one data point per location, the smallest variation with respect to MBE is found in the Perez data. As for the RMSE, the variation of the Perez results is comparable with location 6 (Tokyo, Japan) and 10 (Athens, Greece), yet significantly smaller than the rest. With this collection of evidence, the location independence claim of Perez (1990) is cast in doubt.

Among these data, the discrepancy values reported in Perez (1990) are, by and large, smaller than those reported by other researchers. To systematically quantify the difference, a Wilcoxon rank-sum test (Wilcoxon 1945) was used. It is the nonparametric version of the t-test. Both methods are commonly used to compare whether samples from two populations differ significantly from each other. The details about the basic statistics is referred to Chapter 5 of Wilcox (2012). The first two samples, i.e., Perez' data and Group I data, are found to be significantly different using the Wilcoxon rank-sum test with a p value of 0.001 for MBE and 0.025 for RMSE. The second two samples, i.e., Perez' data and Group II data, are also found to be significantly different in terms of MBE with a p value of 0.046. Additionally, the use of the t-test gives the same conclusion. Thus, we claim that the data used by Perez to develop the model is not representative and may contain bias.

Having observed the location dependency of the model and the possible bias in Perez' results, we now focus on the Eugene data. First, we present in Table 17 the MBE and RMSE of the global solar irradiation on four different surfaces indicated by

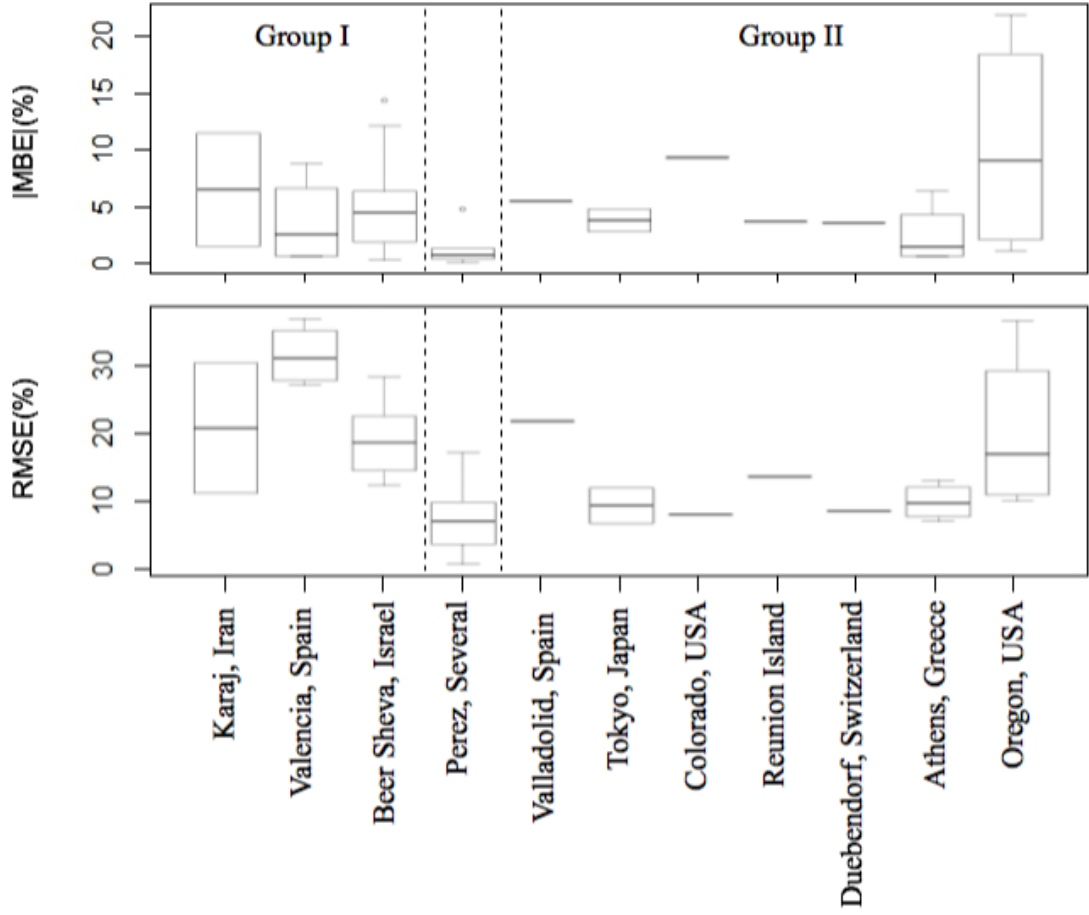


Figure 12: The Absolute MBE and RMSE of Global Solar Irradiation Predicted by the Perez Model at Different Sites.

orientation and tilt angle, in their original units as well as the percentage to the averaged measurements. It clearly shows that the Perez model over-predicts in all four cases, especially for vertical surfaces. Figure 13 compares measured global solar irradiation with Perez predictions on south vertical and north vertical surfaces. It shows that a number of model predictions are higher than measurements. MBE observed on south and north vertical surfaces are $-32 W/m^2$ and $-15 W/m^2$, respectively. The results also show considerable RMSE on the two vertical surfaces, i.e., 22% and 37% for south and north surfaces. Such discrepancies need to be quantified statistically to improve the prediction of the Perez model.

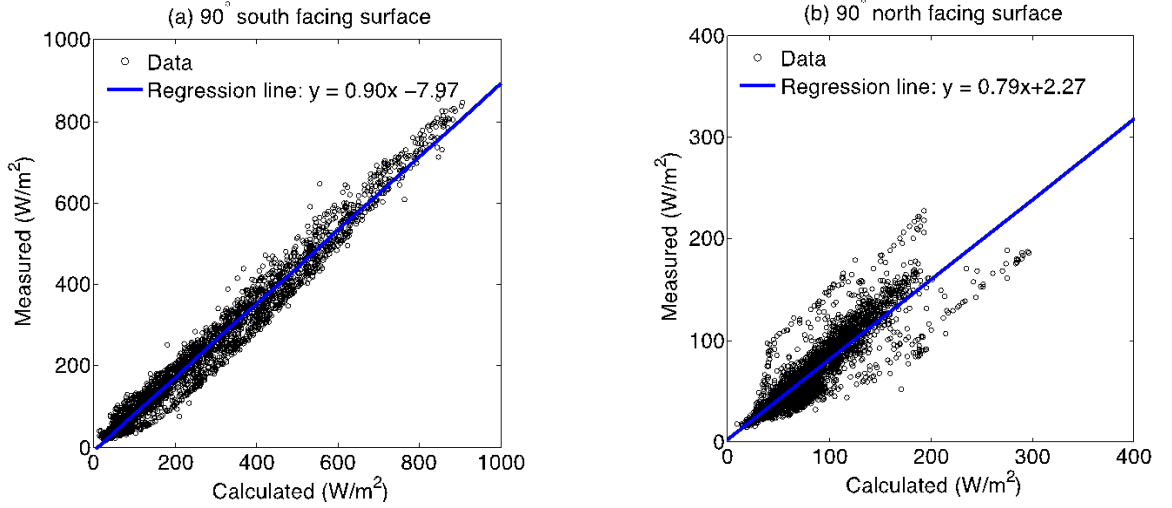


Figure 13: Measured and Predicted Global Irradiation on Two Vertical Surfaces.

Table 17: Average Annual Errors of the Perez Model in the Calculation of the Global Solar Irradiation on Four Tilted Surfaces.

	30° South	45° South	90° South	90° North
Average (W/m^2)	396	391	210	67
MBE (W/m^2)	-12	-5	-32	-15
MBE (%)	-3%	-1%	-15%	-22%
RMSE (W/m^2)	47	41	47	25
RMSE (%)	12%	10%	22%	37%

3.3.4 Model Form Uncertainty Quantification

In order to quantify the model uncertainty, we need to build a statistical adjustment to the prediction discrepancies from the Perez model using the Eugene data. The candidate parameters to be used in the adjustment model are the 10 intermediate parameters of the Perez model. They are solar zenith angle, solar azimuth angle, sky brightness factor, sky clearness factor, direct normal irradiation, diffuse horizontal irradiation, global horizontal irradiation, surface tilt angle, surface azimuth angle, and solar incident angle. They are screened by using scatter plots. We plot the prediction discrepancies against the candidate parameters, and check whether the plots display significant patterns. Among the 10 such plots, only four (solar azimuth

θ , sky brightness factor Δ , direct normal solar irradiation I , and surface tilt angle S) show some systematic patterns. The corresponding scatter plots are given in Figure 14. The plot for the solar azimuth shows a decreasing trend. For sky brightness factor and direct normal solar irradiation, we observe a funnel shape in the plots. This indicates that the variance of *diff* decreases with Δ and increases with I . The surface tilt angle S is known to be an important parameter in previous works, including Perez (1990). The trend line in this plot shows a negative slope.

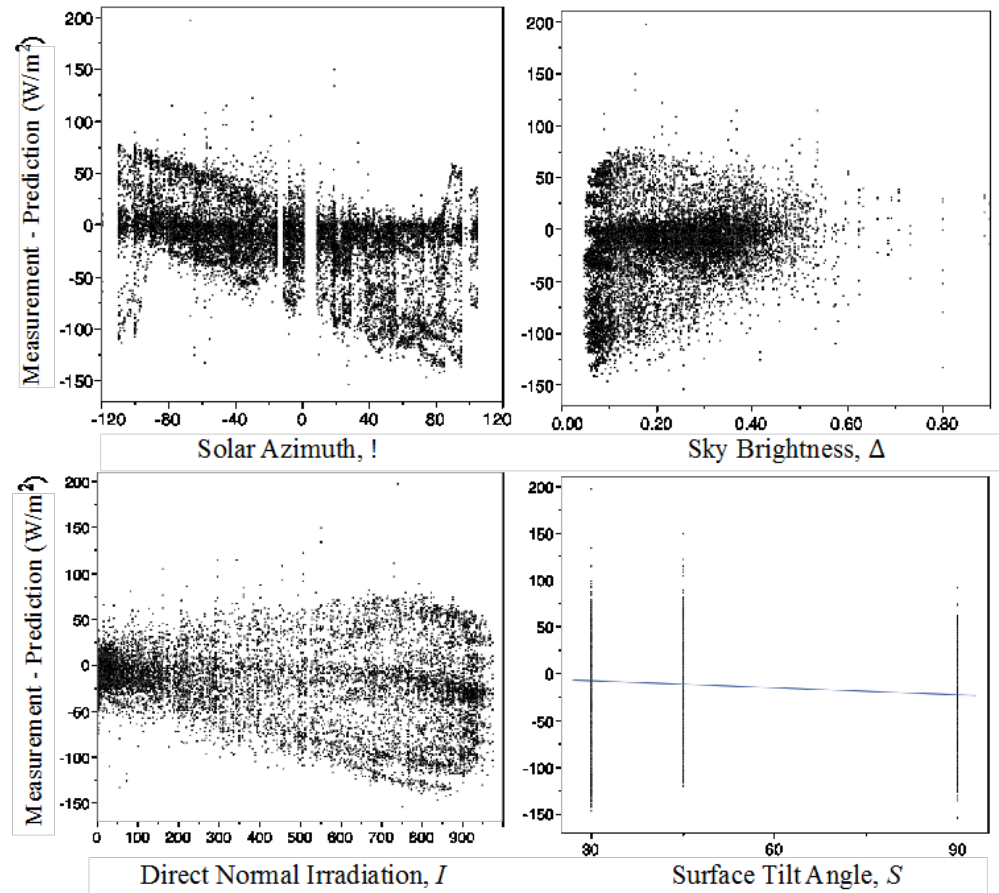


Figure 14: Model Prediction Discrepancies against Solar Azimuth, Sky brightness, Direct Normal Solar Irradiation, and Surface Tilt Angle.

The whole year's data are divided into two disjoint parts: the *training* part and the *validation* part. The latter consists of four half-months' observations in January, March, June, and September. Both datasets cover a reasonably complete range of solar angles and sky conditions, so that any potential bias or extrapolation errors can

be mitigated.

We use a two-phase regression to construct the adjustment model. Because it has to meet the zero discrepancy constraint on horizontal surfaces (when $S = 0$), we consider a polynomial regression model (Wu and Hamada 2009) in S without constant term. Only the linear and quadratic effects are found to be significant. Therefore, we obtain the following model for the first phase:

$$diff(\bullet) = f_1S + f_2S^2 + \epsilon,$$

where $diff(\bullet)$ is a function mapping inputs into an output.

In the second phase, we fit the coefficients f_1 and f_2 as linear functions of the other three parameters Δ , θ and I . By using a stepwise regression to select parameters, we obtain the following model:

$$diff(\bullet) = (c_1I + c_2\theta + c_3\Delta)S + (c_4\theta + c_5\Delta)S^2 + \epsilon. \quad (32)$$

All the coefficients in (32) are significant with p values $< 10^{-13}$. Although the R^2 value is only 0.504, it is acceptable because there are various sources of measurement errors present in the data. Ordinary residual diagnostics were performed to verify the goodness of fit of the model. The residual plot shows a random pattern and the QQ plot almost forms a 45-degree line. So the fitted linear model appears to be quite good. We have now found a reasonably good adjustment for the Perez model for the Eugene data. We call the sum of the Perez prediction and the fitted regression model in (32), the *modified Perez* model.

3.3.5 Model Validation

In this section we validate the $diff(\bullet)$ model developed above on the validation dataset, which consists of four half-months' observations in December, March, June, and September. Table 18 compares the modified Perez model to the original Perez in terms of MBE and RMSE. The results show a substantial improvement in predicting

global solar irradiation on tilted surfaces. The new model reduces the MBE by more than 50% for every test surface. For example, for south vertical surface, it decreases from -33 W/m^2 to -11 W/m^2 , and on north vertical surface, it reduces from -15 W/m^2 to 8 W/m^2 . In terms of RMSE, the modified Perez model also shows a significant improvement. Particularly for the south vertical surface, the RMSE is reduced by over 40% from 46 W/m^2 to 28 W/m^2 . Moreover, the modified model does not consistently overpredict. From the energy balance perspective, the modified model promises a more reliable prediction as it avoids significant amounts of overpredictions of solar irradiation on building envelopes stemming from Perez model inadequacy. Although the variable selection results indicate that surface azimuth is an insignificant factor of the Perez model discrepancy, we will investigate this result by collecting data on east- and west-facing surfaces in a future study.

3.3.6 Prediction of the Modified Perez Model and Uncertainty Quantification

In Table 18, the two-phase regression model $diff(\bullet)$ shows its capability of correcting the bias in predicting the mean of solar irradiation on tilted surfaces. New predictions are then simply obtained by the addition of original computational model outcomes from Perez, i.e., $\eta(\mathbf{u})$ and that from the adjustment regression model, i.e., $diff(\mathbf{u}, \mathbf{v})$. We can also construct the confidence intervals to estimate the uncertainties in new predictions from the linear regression model (Seber and Lee 2012).

We take two days (June 1 and 2, 2011) from the validation dataset to illustrate the prediction capability of $diff(\bullet)$ and the UQ for the 90° south-facing surface at each hour. Each day contains ten hours of values when the surface is exposed to the sun. We first compare the physical observations y_{obs} , with Perez model predictions y_{Perez} , and illustrate the results in Figure 15a. This figure shows that Perez model yields higher predictions at 18 out of the 20 points than the measurements whereas the maximum model discrepancy appears at around 12:00. The differences, i.e., $y_{obs} -$

y_{Perez} , will then become the physical observations on which $diff(\bullet)$ will predict. Figure 15b illustrates $y_{obs} - y_{Perez}$, predictions of $diff(\bullet)$, and a 95% confidence interval. It shows that the 95% confidence interval well covers the hourly variation of $y_{obs} - y_{Perez}$. Therefore, we will use $diff(\bullet)$ to quantify the prediction uncertainties given by the original Perez model.

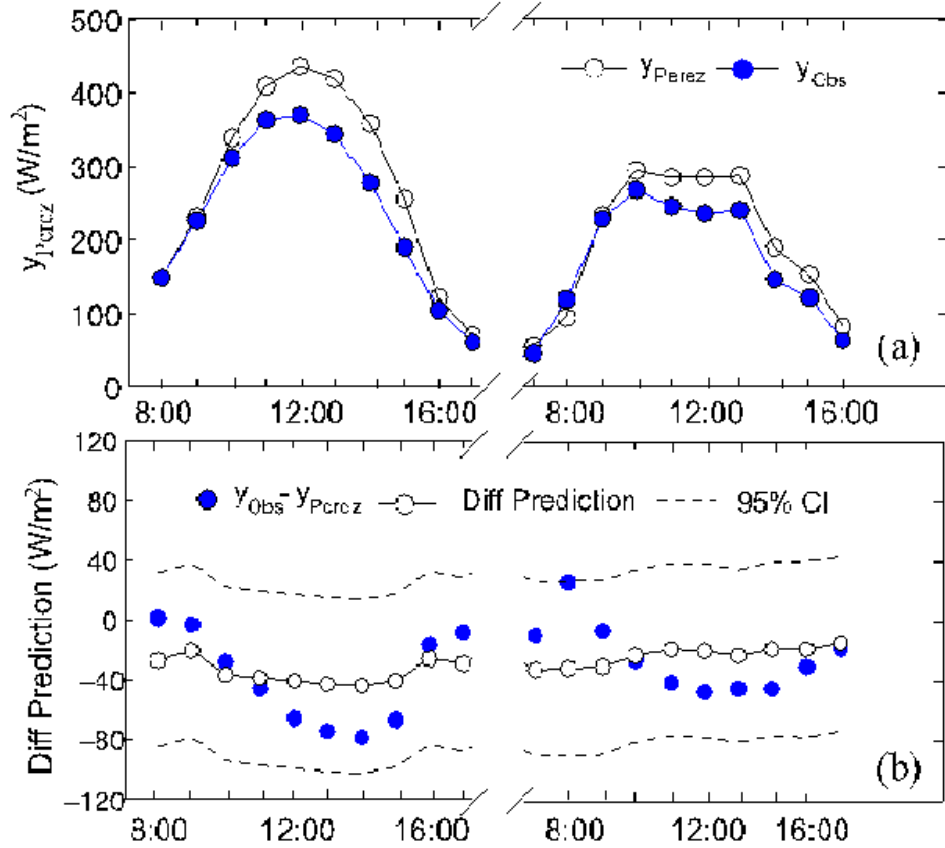


Figure 15: Results on a South-Facing Vertical Surface at Hourly Intervals over Two Days (June 1 and 2, 2011): (a) Perez Predictions, y_{Perez} , and Physical Observations, y_{obs} , and (b) Prediction of $diff(\bullet)$ and 95% Confidence Interval.

Uncertainties that are indicated by the confidence interval come from two major sources: the random errors in the measurements and the remaining model bias not captured by $diff(\bullet)$. The model form uncertainties in the predictions are thus reducible if measurement errors decrease or if more experimental data are added. The

remaining model bias can also reduce if additional variables (as explained earlier, these variables will be part of \mathbf{v}) are detected to be significant and are then added to the $diff(\bullet)$ model in a sufficient manner.

3.4 The Effect of Model Form Uncertainty on Building Energy Simulations

In this section, we study the effect of the Perez model uncertainty on the overall uncertainty of whole-building energy predictions.

3.4.1 Case Study Description

We implement the Perez model uncertainty (PMU) on the basis of EnergyPlus V7.0.0 (EnergyPlus for short) through the modification of its source code. The modified EnergyPlus tool, which is denoted with “EnergyPlus_{PMU}”, has the option of choosing between the original Perez and the modified Perez as the algorithm for the calculation of solar irradiation on tilted surfaces. Therefore, we can study the effect of the model uncertainty of Perez by running simulations in EnergyPlus and EnergyPlus_{PMU} with the same model inputs. The differences in their simulation results are thus attributed to the model form uncertainty, i.e., the addition of $diff(\bullet)$ in the solar irradiance calculation model of EnergyPlus_{PMU}.

As for the selection of buildings, we chose the small and large office reference buildings that are developed by the Department of Energy in the US (Deru et al. 2011). EnergyPlus models for these reference buildings are well scrutinized and widely used for building energy efficiency analysis. The EnergyPlus input files can be downloaded from (DoE, 2013). Details about the buildings are also found in (DoE, 2013). The reasons we chose reference buildings are twofold: (1) using reference buildings allows others to reproduce the results reported in this paper and (2) using these models not only reduce our modelling effort, but the quality of the models is guaranteed with minimum modeller’s bias. We use two buildings to test how the effect of model form

uncertainty may differ for two office buildings of different size.

To show the effect of model form uncertainty in the presence of input parameter uncertainties, we shall conduct the traditional uncertainty quantification of model input parameters. By doing so, we investigate the effect of model form uncertainty not at one fixed combination of all other input parameters but explore its effect on all possible combinations of parameter values with the uncertainty range established for them in a UQ repository (Sun et al. 2013, Lee et al. 2013) and other sources used in this study. For example, we quantify the uncertainties in microclimate variables (Sun et al. 2014), convective heat transfer coefficients (Palyvos 2008), material properties (Macdonald 2002, Domnguez-Muoz et al. 2010), infiltration rate (Wang et al. 2014) etc. Parameter uncertainties are explored using 500 samples with Latin Hypercube Sampling (LHS) technique (Wyss and Jorgensen 1998), and are propagated in the Georgia Tech Uncertainty and Risk Analysis Workbench (GURA-W) to obtain building energy predictions. This GURA-W is a software environment that automates the processes of parameter sampling, uncertainty quantification and uncertainty analysis (Lee et al. 2013). Its UQ repository characterizes the different sources of uncertainties and is thus a convenient starting point for a building energy uncertainty quantification study. Given the 500 samples, we run two sets of simulations in parallel with the Perez and the Modified Perez model in EnergyPlus and EnergyPlus_{PMU}, respectively. All simulations are conducted using the TMY data for the Oregon location.

3.4.2 Results and Discussions

This section describes the effect of Perez model form uncertainty on the prediction of annual total cooling electricity use and heating natural gas use separately. We first use the results from 500 LHS samples to construct the empirical cumulative density

functions (CDF) that describe the uncertainty in building energy prediction. We prefer CDF to probabilistic density functions (PDF) because CDFs can directly answer a question such as “how likely is it that the building energy consumption is small or smaller?” The empirical CDF obtained from using the original Perez model is compared with that obtained from using the modified Perez model. We plot the annual cooling CDFs in Figure 16. The modified Perez shifts cooling distributions from the original Perez. The results suggest that two CDFs for cooling energy consumption significantly differ from each other for both small and large office. Taking the small office as an example, using the original Perez model suggests that there is a 26% chance for this building to use 3000 kWh or less electricity for cooling. For identical uncertainties in input parameters in both models, this probability will increase to 33% if Perez model uncertainty is considered. We found that the modified Perez model does however not significantly modify the distribution of heating natural gas consumption because heating loads usually occurs at hours when there are less solar gains.

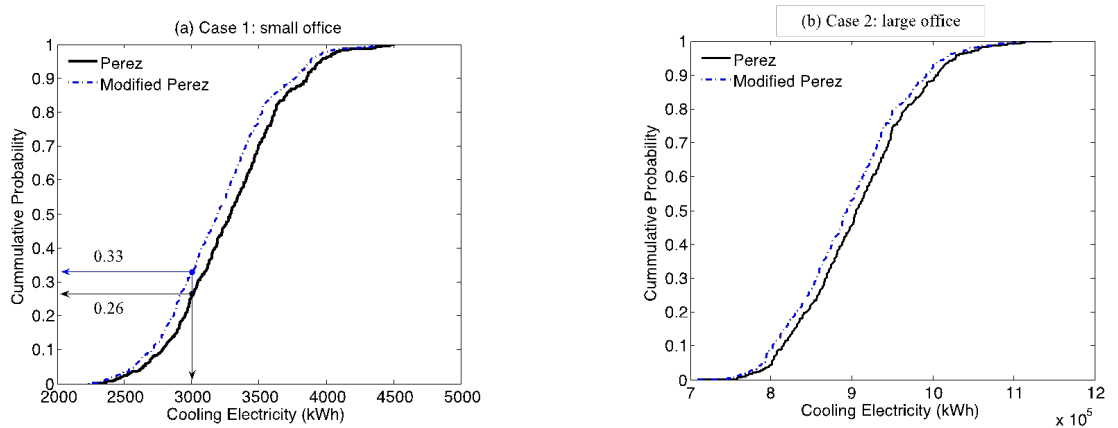


Figure 16: Annul Total Cooling Electricity Consumption Predicted from the Use of the Original and the Modified Perez Model.

If we compare the overall results we conclude that, compared with the effect of parameter uncertainties, the effect of combining them with the model form uncertainty of the Perez model is not evident from the previous analysis. We therefore

compute the sample-paired differences of 500 samples in both the cooling and heating energy consumption. Two samples are treated as a pair if the only difference between them is from the calculation of solar diffuses irradiation on tilted surfaces, i.e., one uses original Perez model and the other uses the modified Perez model. In a paired comparison design, we can calculate the effect of model form uncertainty, i.e., $diff(\bullet)$, while eliminating the source of variation due to the variability in model input parameters. The details of this type of experiment design can be found in Wu and Hamada's book (2009). We compute the effect of $diff(\bullet)$ as a percentage by the following equation:

$$E_{diff}(\%) = \frac{y_{ModifiedPerez} - y_{Perez}}{y_{Perez}} \times 100\%,$$

where E_{diff} is the effect of $diff(\bullet)$ on building energy predictions, $y_{ModifiedPerez}$ is the energy predictions using the modified Perez model in EnergyPlus_{PMU}, and y_{Perez} is the energy predictions using the original Perez model in EnergyPlus.

Because we have 500 pairs, we can plot the empirical CDFs, showing the global effect of the modified Perez model. Figure 17 illustrates E_{diff} of cooling and heating energy prediction for both the small and the large office. The mean and the corresponding cumulative probability are also shown in the figure. The effect of model uncertainty on both cooling and heating energy use is apparent when analysed with the paired comparison design. As for cooling electricity consumption, the uncertainty in the Perez model causes an average of overestimate by 2.76% for the small office and by 1.36% for the large office. Depending on the values of the other uncertain input parameters, this effect can vary by an additional $\pm 1\%$. By comparing the two building cases, we see that the degree of the effects changes from one building to another. Whereas the small office building's cooling need is dominated by solar gains, the large office building has only a moderate level of solar gains compared to the total cooling load. This explains the difference in the influence of the improved solar diffuse radiation calculation for the two buildings.

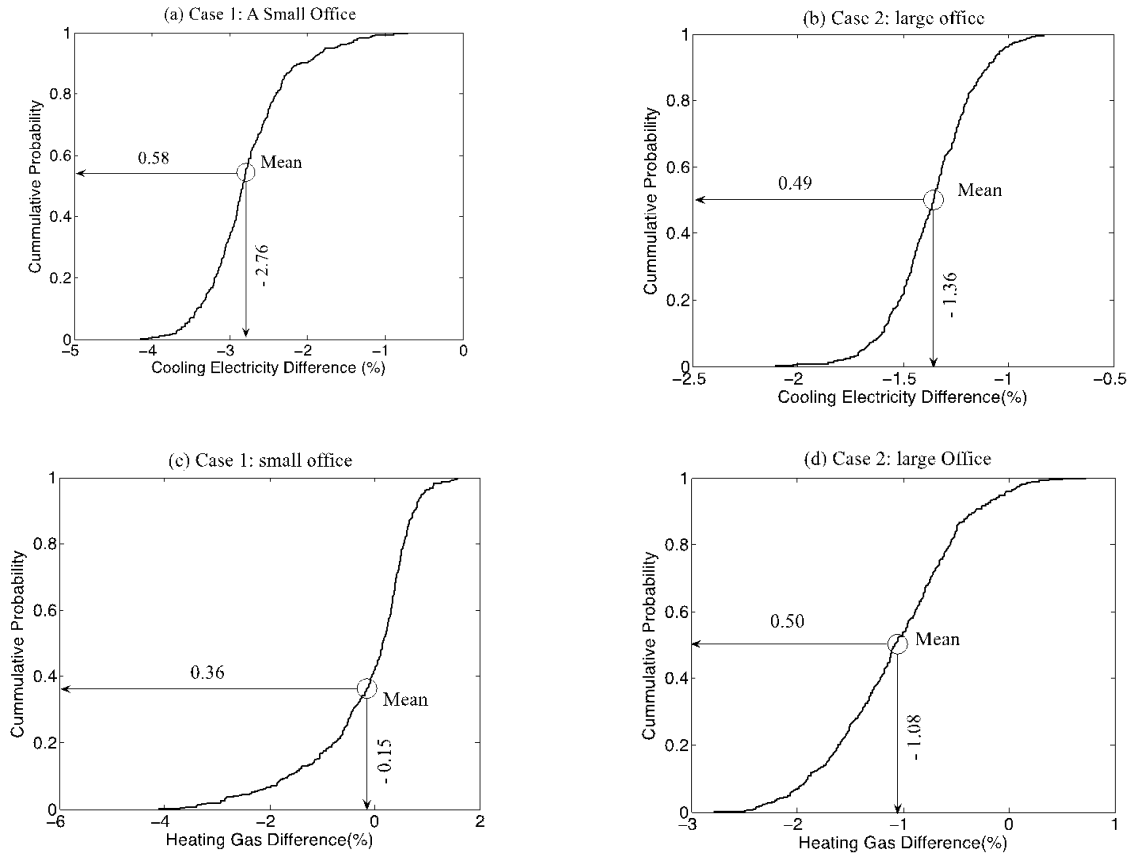


Figure 17: Sample-Paired Analysis for Cooling Electricity Use and Heating Natural Gas Use.

As for the heating natural gas consumption, the Perez model form uncertainty leads to a slight overestimate of the average by 0.15% for the small office and by 1.08% for the large office. Although the mean of the effect is less evident than that on the cooling, its variation is still obvious. For example, at certain combination of the other uncertain input parameters, the effect on heating can be as large as -3%. Additionally, the fact that the model form uncertainty leads to overestimate of heating energy use seems counterintuitive if we only stay at the level of annual total solar irradiation as shown in Table 18. Scrutinizing the bias in the Perez model prediction reveals that the Perez model tends to overestimate solar irradiation on building surfaces in the afternoon (i.e., solar azimuth θ is positive), but underestimates in the morning (i.e., solar azimuth θ is negative). This also can be seen from the plot of solar azimuth

against the Perez model bias in Figure 14. Because of internal load in the building, especially for the large office, heating demand appears more often in the morning than in the afternoon. In other words, the prediction of solar irradiation in the morning period becomes an important factor of building heating energy use. Therefore, when the modified Perez corrects this kind of dynamic bias, both heating and cooling energy predictions become lower than those obtained with the original Perez model. In fact, this phenomenon shows the significance of quantifying model form uncertainty as a function instead of through some surrogate input parameters that would attempt to capture the uncertainty through parameters that remain static through a simulation run. Model form uncertainty on the other hand varies within each simulation run, which intrinsically differs from input parameter uncertainty.

Table 18: Model Validation Statistical Results.

	30° South		45° South		90° South		90° North	
	Perez 1990	Modified Perez	Perez 1990	Modified Perez	Perez 1990	Modified Perez	Perez 1990	Modified Perez
Average (W/m^2)	385		382		210		68	
MBE (W/m^2)	-14	-7	-7	3	-33	-11	-15	8
MBE (%)	-4%	-2%	-2%	1%	-16%	-5%	-21%	11%
RMSE (W/m^2)	45	31	39	30	46	28	22	19
RMSE (%)	12%	8%	10%	8%	22%	14%	33%	28%

Lastly, both case studies show statistically significant effects of model form uncertainty in solar irradiation calculations on building energy predictions. Because the Perez model is just one out of potentially many submodels in a complete building energy model, it is reasonable to claim that model form uncertainty matters for building simulation. If the model form uncertainty is addressed at the submodel level where higher fidelity models and/or sufficient calibration data are typically available, the overall quantification of prediction uncertainty in building energy consumption can significantly improve.

3.5 Conclusions

Model form uncertainty quantification (UQ) is related to model validation but attempts to extend it to the estimation of model inadequacy in new predictions. This can be accomplished with regression analysis. We developed a general framework for model form UQ based on physical measurements. This approach has been shown to be effective at subsystem or unit level where high-quality physical measurements were attainable. The resulting expression of model form uncertainty can be included in current building energy software tools, through an “intrusive” (coding and recompilation) method.

The approach was exemplified on the Perez 1990 model, which computes solar irradiation on tilted surfaces. Model discrepancies were quantified with a linear regression model, which not only improves predictions but also estimates the uncertainties in solar diffuse irradiation on tilted surfaces. As one out of potentially many submodels, the Perez model form uncertainty showed significance in the context of uncertainty analysis of building energy simulations in two case studies.

The major conclusion of the work presented in this paper is that model form uncertainty is integral to the complete UQ of building energy simulation. Only if this type of uncertainty is effectively characterized, can one get the appropriate level of confidence in the final simulation results.

CHAPTER IV

PIG PROCESS: JOINT MODELING OF POINT AND INTEGRAL RESPONSES IN COMPUTER EXPERIMENTS

4.1 *Introduction*

In building energy simulations, estimation of the diffuse component of the solar irradiance on a building façade I_D , which appears as the boundary conditions of the simulations, remains a challenge for building technology researchers. The diffuse irradiance i_D is scattered in the atmosphere, which is viewed as a sky dome. The diffuse irradiance on a building façade I_D comes from the proportion of the diffuse irradiance over the sky dome i_D this façade is exposed to. For example, a north facing vertical façade receives the diffuse irradiance of the northern half hemisphere of the sky dome, and a horizontal surface receives all the diffuse irradiance over the sky dome. Statistically, we can view the sky dome as a two-dimensional input space \mathcal{D} . Then I_D can be regarded as an integral of the diffuse irradiance over the sky dome i_D , over the region $R \subset \mathcal{D}$ the façade is exposed to, i.e., $I_D = \int_R i_D d\mathcal{D}$. If we can model i_D as a function of the position over the sky dome $(x, y) \in \mathcal{D}$, i.e., $i_D = f(x, y)$, the diffuse irradiance on an arbitrary façade I_D can be predicted by integrating the function $f(x, y)$ over the region R of the sky dome the façade is exposed to, i.e., $I_D = \int_R f(x, y) d(x, y)$. The question is, whether we can model i_D with the integral information I_D .

In computer experiments, the computer codes or computer models, are simulations of the real systems. They usually have high-dimensional inputs and require a substantial amount of computational effort. Statistically, they are viewed as a black

box function defined on a d -dimensional input space \mathcal{D} , which is also called the design space, with one-dimensional response y , i.e., $\eta : \mathcal{D} \rightarrow \mathcal{R}^1$ and $y = \eta(x)$ for $x \in \mathcal{D}$. Theoretically, the true function can be obtained by exhaustively evaluating the computer codes in the design space \mathcal{D} , but this is not affordable. Instead, only a few samples can be obtained and an emulator or surrogate model $f(x)$ is built for the computer codes η . In traditional computer experiments, only point responses are used to build the surrogate model. That is, we have the point site input $x \in \mathcal{D}$ with the corresponding point response $y = \eta(x)$, and we want to model η by $f(x)$. Morris et al. (1993) was the first to propose a model where derivative information about η can be used to improve the emulator. That is, we have point site input $x \in \mathcal{D}$ with the corresponding point response $y = \eta(x)$ as well as the derivatives $\frac{\partial y}{\partial x_i} = \frac{\partial \eta(x)}{\partial x_i}$. They showed that these two types of information can be combined to build a better model $f(x)$ for η . Motivated by the example from building energy simulations and the work by Morris et al. (1993), we consider the case where the computer codes can produce two types of responses, i.e., for a given point site $x \in \mathcal{D}$, we will have $y = \eta(x)$, and for a given region $R \subset \mathcal{D}$, we will have $\mathcal{Y} = \int_R \eta(x) dx$. Then the question is: whether we can combine these two types of information to build a better model for η .

In the next section, a joint model for point and integral responses is proposed. In section 4.3, two examples are given to demonstrate the proposed model and its potential advantages. In section 4.4, some remarks are given and future research directions will be discussed.

4.2 PIG Process Model

4.2.1 Point GP, Integral GP and PIG Process Models

In computer experiment, the deterministic point response $y_P(x)$ is usually treated as a realization of a random process $Y_P(x)$. In this work, we adopt the commonly used Gaussian process model, which can be represented by its mean function and covariance

function. To avoid notational ambiguity, we call it the *point Gaussian process* model in contrast to the *integral Gaussian process* model and the *point-integral Gaussian process* model introduced later. A generic Gaussian process model can be written as:

$$Y_P(x) = \mu_P(x) + Z_P(x),$$

where the mean function $\mu_P(x)$ is usually fitted as a linear regression model $\mu_P(x) = \sum_j f_j(x)\beta_j = \mathbf{f}\boldsymbol{\beta}$, where $f_j(x)$'s are known regression functions and β_j 's are unknown regression coefficients. $Z_P(x)$ is assumed to be a zero-mean stationary Gaussian process with covariance function defined as:

$$K_{PP}(x, w) = \text{Cov}(Z_P(x), Z_P(w)) = \sigma^2 k_{PP}(x, w), \quad (33)$$

where σ^2 is the process variance and $k_{PP}(x, w)$ is the correlation. The form of the covariance function is usually pre-specified and characterized by a vector of unknown parameters $\boldsymbol{\tau}$. One popular choice is the power exponential correlation function defined as (Santner et al., 2003):

$$k_{PP}(x, w) = \prod_i \exp \left\{ -\frac{(x_i - w_i)^p}{2\tau_i^2} \right\} \text{ for } 0 < p \leq 2, \quad (34)$$

where x_i and w_i are the i th entries of x and w respectively, and τ_i is the unknown parameter for the i th dimension. For $p=2$, this is also called the *Gaussian correlation function*. Another popular choice is the Matrn correlation function, whose details can be found in the book by Santner et al. (2003).

For computer experiments with deterministic integral response $y_I(R)$, we can treat it as a realization of another random process $Y_I(R)$. Recall that the integral response is produced by aggregating the true function values $y_P(x)$ over the specified region R , i.e., $y_I(R) = \int_R y_P(x) dx$. If we view $y_P(x)$ as a point GP model $Y_P(x)$, the random process $Y_I(R)$, called the integral Gaussian process model, can be written as:

$$Y_I(R) = \int_R Y_P(x) dx = \int_R \mu_P(x) dx + \int_R Z_P(x) dx = \mu_I(R) + Z_I(R).$$

As in the point GP model, the integral GP model is also determined by its mean function and covariance function. Note that if we assume $\mu_P(x)$ is linear, $\mu_I(R) = \int_R \mu_P(x) dx = \int_R \sum_j f_j(x) \beta_j dx = \sum_j (\int_R f_j(x) dx) \beta_j$ is also linear with the same set of unknown regression coefficients β . $Z_I(R)$ is still a zero-mean stationary Gaussian process with covariance function defined as:

$$\begin{aligned}
K_{II}(R_1, R_2) &= Cov(Z_I(R_1), Z_I(R_2)) = Cov\left(\int_{R_1} Z_P(x) dx, \int_{R_2} Z_P(w) dx\right) \\
&= \int_{R_1} \int_{R_2} Cov(Z_P(x), Z_P(w)) dw dx = \int_{R_1} \int_{R_2} K_{PP}(x, w) dw dx \\
&= \sigma^2 \int_{R_1} \int_{R_2} k_{PP}(x, w) dw dx = \sigma^2 k_{II}(R_1, R_2), \tag{35}
\end{aligned}$$

where σ^2 is the variance of the underlying point GP model $Y_p(x)$, and $k_{II}(R_1, R_2)$ is called the *pseudo-correlation* with an abuse of terminology. The variance of $Z_I(R)$ can be calculated as:

$$Var(Z_I(R)) = Cov(Z_I(R), Z_I(R)) = \sigma^2 k_{II}(R, R).$$

This implies that the true correlation between $Z_I(R_1)$ and $Z_I(R_2)$ is:

$$\begin{aligned}
Corr(Z_I(R_1), Z_I(R_2)) &= \frac{Cov(Z_I(R_1), Z_I(R_2))}{\sqrt{Var(Z_I(R_1))Var(Z_I(R_2))}} \\
&= \frac{k_{II}(R_1, R_2)}{\sqrt{k_{II}(R_1, R_1)k_{II}(R_2, R_2)}}.
\end{aligned}$$

The covariance function $K_{II}(R_1, R_2)$ of the integral GP model is determined by the covariance function $K_{PP}(x, w)$ of the underlying point GP model. Both covariance functions share the same set of unknown parameters τ . In fact, both models share the same unknown parameters: the regression coefficients β , the variance σ^2 of the point GP model and the parameters in the covariance function. By fitting the integral responses with the integral GP model, the underlying point GP model can also be inferred. This answers the question proposed in the introduction, namely, whether we can fit the point response of the true function with the integral information.

Now suppose the computer experiment can produce both types of deterministic responses:

$$f(u) \begin{cases} y_P(u) & \text{if } u \text{ is a point site,} \\ y_I(u) & \text{if } u \text{ is a region.} \end{cases}$$

They can be viewed together as a realization of another random process $Y(u)$, denoted as the point-integral Gaussian process model, or abbreviated as the *PIG process* model. The generic expression of the PIG process model can be written as:

$$Y(u) = \mu(u) + Z(u),$$

where $\mu(u) = \sum_j F_j(u)\beta_j = \mathbf{F}\boldsymbol{\beta}$ is linear, if we assume the mean function of the underlying point GP model is linear. The known regression function $F_j(u)$ can be written as:

$$F_j(u) \begin{cases} f_j(u) & \text{if } u \text{ is a point site,} \\ \int_u f_j(x) dx & \text{if } u \text{ is a region.} \end{cases}$$

$Z(u)$ is a zero-mean stationary Gaussian process with the covariance function defined as:

$$K(u_1, u_2) \begin{cases} K_{PP}(u_1, u_2) & \text{if } u_1 \text{ and } u_2 \text{ are both point sites,} \\ K_{PI}(u_1, u_2) & \text{if } u_1 \text{ is a point site and } u_2 \text{ is a region,} \\ K_{IP}(u_1, u_2) & \text{if } u_1 \text{ is a region and } u_2 \text{ is a point site,} \\ K_{II}(u_1, u_2) & \text{if } u_1 \text{ and } u_2 \text{ are both regions.} \end{cases}$$

The covariance between two point sites and the covariance between two regions are defined in (33) and (35) respectively. For a point site x and a region R , the covariance between them is:

$$\begin{aligned} K_{PI}(x, R) &= Cov(Z_P(x), Z_I(R)) = Cov\left(Z_P(x), \int_R Z_P(w) dw\right) \\ &= \int_R Cov(Z_P(x), Z_P(w)) dw = \int_R K_{PP}(x, w) dw \\ &= \sigma^2 \int_R k_{PP}(x, w) dw = \sigma^2 k_{PI}(x, R), \end{aligned} \tag{36}$$

where σ^2 is the variance of the underlying point GP process model, and $k_{PI}(x, R)$ is called the pseudo-correlation as before. By symmetry, $K_{IP}(R, x) = K_{PI}(x, R)$. The true correlation between $Z_P(x)$ and $Z_I(R)$ is:

$$\text{Corr}(Z_P(x), Z_I(R)) = \frac{\text{Cov}(Z_P(x), Z_I(R))}{\sqrt{\text{Var}(Z_P(x))\text{Var}(Z_I(R))}} = \frac{k_{PI}(x, R)}{\sqrt{k_{II}(R, R)}}.$$

Note that $K_{PI}(x, R)$ is also determined by $K_{PP}(x, w)$. Moreover, as in the integral GP model, the unknown parameters for the PIG process model are also β , σ^2 and τ . Therefore, by jointly fitting the point and integral responses with the PIG process model, the underlying point GP model can be inferred.

In section 4.2.2, we shall discuss parameter estimation and prediction by following the frequentist approach.

4.2.2 Estimation and Prediction

Let the design space of the computer experiment be \mathcal{D} . Suppose n_1 point sites $\{x_1, \dots, x_{n_1}\}$ and n_2 regions $\{R_1, \dots, R_{n_2}\}$ have been evaluated with responses $\mathbf{y} = (y(x_1), \dots, y(x_{n_1}), \int_{R_1} y(w) dw, \dots, \int_{R_{n_2}} y(w) dw)'$ respectively. Note that $x_i \in \mathcal{D}$ for $i \in \{1, \dots, n_1\}$ and $R_j \subset \mathcal{D}$ for $j \in \{1, \dots, n_2\}$. Let $n = n_1 + n_2$. The $n \times n$ covariance matrix K is organized in 2×2 blocks:

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \sigma^2 \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} = \sigma^2 k.$$

Among the four blocks, k_{11} is the correlation matrix of the n_1 point sites:

$$k_{11} = \begin{pmatrix} 1 & k_{PP}(x_1, x_2) & \cdots & k_{PP}(x_1, x_{n_1}) \\ k_{PP}(x_2, x_1) & 1 & \cdots & k_{PP}(x_2, x_{n_1}) \\ \vdots & \vdots & \ddots & \vdots \\ k_{PP}(x_{n_1}, x_1) & k_{PP}(x_{n_1}, x_2) & \cdots & 1 \end{pmatrix},$$

where $k_{PP}(\cdot, \cdot)$ is the correlation function of the underlying point GP model; $k_{12} = k'_{21}$ is an $n_1 \times n_2$ matrix representing the pseudo-correlations between any point site and

any region:

$$k_{12} = \begin{pmatrix} k_{PI}(x_1, R_1) & k_{PI}(x_1, R_2) & \cdots & k_{PI}(x_1, R_{n_2}) \\ k_{PI}(x_2, R_1) & k_{PI}(x_2, R_2) & \cdots & k_{PI}(x_2, R_{n_2}) \\ \vdots & \vdots & \ddots & \vdots \\ k_{PI}(x_{n_1}, R_1) & k_{PI}(x_{n_1}, R_2) & \cdots & k_{PI}(x_{n_1}, R_{n_2}) \end{pmatrix},$$

where $k_{PI}(\cdot, \cdot)$ is defined in (36). Finally, k_{22} is the pseudo-correlation matrix of the n_2 regions:

$$k_{11} = \begin{pmatrix} k_{II}(R_1, R_1) & k_{II}(R_1, R_2) & \cdots & k_{II}(R_1, R_{n_2}) \\ k_{II}(R_2, R_1) & k_{II}(R_2, R_2) & \cdots & k_{II}(R_2, R_{n_2}) \\ \vdots & \vdots & \ddots & \vdots \\ k_{II}(R_{n_2}, R_1) & k_{II}(R_{n_2}, R_2) & \cdots & k_{II}(R_{n_2}, R_{n_2}) \end{pmatrix},$$

where $k_{II}(\cdot, \cdot)$ is defined in (35).

By fitting \mathbf{y} with the PIG process model, we have $\mathbf{y} \sim \mathcal{N}_n(\mathbf{F}\boldsymbol{\beta}, \sigma^2 k)$. If we assume the linear regression model consists of m terms, we have its design matrix as:

$$\mathbf{F} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_{n_1}) & f_2(x_{n_1}) & \cdots & f_m(x_{n_1}) \\ \int_{R_1} f_1(x) dx & \int_{R_1} f_2(x) dx & \cdots & \int_{R_1} f_m(x) dx \\ \vdots & \vdots & \ddots & \vdots \\ \int_{R_{n_2}} f_1(x) dx & \int_{R_{n_2}} f_2(x) dx & \cdots & \int_{R_{n_2}} f_m(x) dx \end{pmatrix}.$$

Suppose the correlation function $k_{PP}(x, w)$ of the underlying point GP model has the unknown parameters $\boldsymbol{\tau}$. The log-likelihood function, apart from additive and multiplicative constants, can be written as:

$$l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}) = -n \log - \log |k(\boldsymbol{\tau})| - \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{F}\boldsymbol{\beta})' k(\boldsymbol{\tau})^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\beta}).$$

Given $\boldsymbol{\tau}$, the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2 can be derived by solving the first order conditions of the log-likelihood function and checking its second order

conditions. After some calculations (Santer et al., 2003), we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}'k^{-1}\mathbf{F})^{-1}\mathbf{F}k^{-1}\mathbf{y}, \quad (37)$$

which is the generalized least-square estimate of $\boldsymbol{\beta}$, and

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}})'k^{-1}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}}). \quad (38)$$

With $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ defined in (37) and (38), the MLEs of $\boldsymbol{\tau}$ can be determined by maximizing the following objective function:

$$Q(\boldsymbol{\tau}) = l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \boldsymbol{\tau}) = -n \log \hat{\sigma}^2 - \log |k(\boldsymbol{\tau})|,$$

which is usually done by using standard optimization algorithms implemented in mathematical software such as Matlab. The last step described above usually requires a considerable amount of computational effort, which depends on the sample sizes n_1 and n_2 , the choice of design regions, dimension of $\boldsymbol{\tau}$ and the choice of correlation function $k_{PP}(x, w)$. With a large number of integrations involved in the calculation of k , the computational cost of the PIG process model is significantly higher than the traditional point GP model. In section 4.2.3, we shall introduce a simple specification of the correlation function $k_{PP}(x, w)$, and derive closed form expressions for $k_{PI}(x, R)$ and $k_{II}(R_1, R_2)$, both of which involve integration calculations.

The fitted PIG process model can be used to predict the point response y_0 at any untried site x_0 within the design space. The joint distribution of $y_0 = y_P(x_0)$ and $\mathbf{y} = \left(y(x_1), \dots, y(x_{n_1}), \int_{R_1} y(w) dw, \dots, \int_{R_{n_2}} y(w) dw \right)'$ is a multivariate normal:

$$\begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}_{1+n} \left(\begin{pmatrix} \hat{\mu}(x_0) \\ \hat{\boldsymbol{\mu}} \end{pmatrix}, \hat{\sigma}^2 \begin{pmatrix} 1 & k_0(x_0) \\ k_0'(x_0) & k \end{pmatrix} \right),$$

where $\hat{\mu}(x_0) = \mathbf{f}'(x_0)\hat{\boldsymbol{\beta}}$ is the estimated mean function of y_0 , $\hat{\boldsymbol{\mu}} = \mathbf{F}\hat{\boldsymbol{\beta}}$ is the estimated mean function of \mathbf{y} , $k_0(x_0) = (k_{PP}(x_0, x_1), \dots, k_{PP}(x_0, x_{n_1}), k_{PI}(x_0, R_1), \dots, k_{PI}(x_0, R_{n_2}))'$ is an $n \times 1$ vector for the pseudo-correlations between x_0 and the n

design inputs (n_1 point sites and n_2 regions), k is the pseudo-correlation matrix of the n design inputs, and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are given in (37) and (38). As derived in Santner et al. (2003), the empirical Best Linear Unbiased Predictor (BLUP) of y_0 is equal to the conditional expectation of y_0 given \mathbf{y} :

$$\hat{y}_0 = E(y_0|\mathbf{y}) = \hat{\mu}(x_0) + k'_0(x_0)k^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}). \quad (39)$$

Similarly, we can predict the integral response \mathcal{Y}_0 for any given region R_0 within the design space. The joint distribution of $\mathcal{Y}_0 = y_I(R_0) = \int_{R_0} y_P(x) dx$ and \mathbf{y} is given by:

$$\begin{pmatrix} \mathcal{Y}_0 \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}_{1+n} \left(\begin{pmatrix} \hat{\mu}(R_0) \\ \hat{\boldsymbol{\mu}} \end{pmatrix}, \hat{\sigma}^2 \begin{pmatrix} 1 & k_0(R_0) \\ k'_0(R_0) & k \end{pmatrix} \right),$$

where $\hat{\mu}(R_0) = \sum_j \left(\int_{R_0} f_j(x) dx \right) \hat{\beta}_j$ is the estimated mean function of \mathcal{Y}_0 , and $k_0(R_0) = (k_{PI}(x_1, R_0), \dots, k_{PI}(x_{n_1}, R_0), k_{II}(R_0, R_1), \dots, k_{II}(R_0, R_{n_2}))'$ is an $n \times 1$ vector for the pseudo-correlations between R_0 and the n design inputs. Then, the empirical BLUP of \mathcal{Y}_0 is the conditional expectation of \mathcal{Y}_0 given \mathbf{y} (Santner et al., 2003):

$$\hat{\mathcal{Y}}_0 = E(\mathcal{Y}_0|\mathbf{y}) = \hat{\mu}(R_0) + k'_0(R_0)k^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}). \quad (40)$$

Now go back to the traditional point GP model, where only the point responses $\mathbf{y}_0 = (y(x_1), \dots, y(x_{n_1}))$ are used to build the model. As shown in Santner et al. (2003), the empirical BLUP of the point response y_\bullet at an untried site x_\bullet is given by:

$$\hat{y}_\bullet = \mathbf{f}'(x_\bullet)\hat{\boldsymbol{\beta}} + k'_0(x_\bullet)k_{11}^{-1}(\mathbf{y}_0 - \mathbf{F}_0\hat{\boldsymbol{\beta}}), \quad (41)$$

where $\mathbf{F}_0 = (f(x_1), \dots, f(x_{n_1}))'$ is the matrix of the known regression functions of the n_1 point sites, k_{11} is the correlation matrix, and $k_0(x_\bullet) = (k_{PP}(x_\bullet, x_1), \dots, k_{PP}(x_\bullet, x_{n_1}))'$ is an $n_1 \times 1$ vector for the correlations between x_0 and the n_1 point sites. The empirical BLUPs in (39)-(41) all have interpolating properties.

4.2.3 Covariance Functions for Axis-Parallel Rectangular Regions

Let the design space \mathcal{D} be d -dimensional. For simplicity, we consider the power exponential correlation function in (34) with $p=2$ for the underlying point GP model, i.e.,

$$k_{PP}(x, w) = \prod_{i=1}^d \exp \left\{ -\frac{(x_i - w_i)^2}{2\tau_i^2} \right\} = \prod_{i=1}^d k_{PP}^i(x_i, w_i), \quad (42)$$

where x_i and w_i are the i th entries of x and w respectively, and τ_i 's are the unknown parameters.

For ease of calculation, we consider only the axis-parallel rectangular regions, i.e., $R = [a_1, b_1] \times \cdots \times [a_d, b_d] = R_1 \times \cdots \times R_d$. Then for any point site x and any region R , their pseudo-correlation $k_{PI}(x, R)$ can be written as:

$$k_{PI}(x, R) = \prod_{i=1}^d \int_{a_i}^{b_i} \exp \left\{ -\frac{(x_i - w_i)^2}{2\tau_i^2} \right\} dw_i = \prod_{i=1}^d k_{PI}^i(x_i, R_i). \quad (43)$$

Similarly, for any two regions $R_1 = [a_{11}, b_{11}] \times \cdots \times [a_{1d}, b_{1d}] = R_{11} \times \cdots \times R_{1d}$ and $R_2 = [a_{21}, b_{21}] \times \cdots \times [a_{2d}, b_{2d}] = R_{21} \times \cdots \times R_{2d}$, their pseudo-correlation $k_{II}(R_1, R_2)$ can be written as:

$$k_{II}(R_1, R_2) = \prod_{i=1}^d \int_{a_{1i}}^{b_{1i}} \int_{a_{2i}}^{b_{2i}} \exp \left\{ -\frac{(x_i - w_i)^2}{2\tau_i^2} \right\} dw_i dx_i = \prod_{i=1}^d k_{II}^i(R_{1i}, R_{2i}). \quad (44)$$

As seen above, the three pseudo-correlations can be written as products of one-dimensional pseudo-correlations. The calculation in (42)-(44) can be reduced to their corresponding one-dimensional cases:

$$k_{PP}^1(x, w) = \exp \left\{ -\frac{(x - w)^2}{2\tau^2} \right\}, \quad (45)$$

$$k_{PI}^1(x, [a, b]) = \int_a^b \exp \left\{ -\frac{(x - w)^2}{2\tau^2} \right\} dw, \quad (46)$$

$$k_{II}^1([a_1, b_1], [a_2, b_2]) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \exp \left\{ -\frac{(x - w)^2}{2\tau^2} \right\} dw dx. \quad (47)$$

The expression in (46) can be simplified as:

$$k_{PI}^1(x, [a, b]) = \int_a^b \exp \left\{ -\frac{(x-w)^2}{2\tau^2} \right\} dw = \tau\sqrt{2\pi} \left(\Phi \left(\frac{b-x}{\tau} \right) - \Phi \left(\frac{a-x}{\tau} \right) \right), \quad (48)$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left\{ -\frac{t^2}{2} \right\} dt$ is the cumulative normal distribution function.

To calculate (47), let

$$f(a, b) = \int_{-\infty}^a \int_b^{+\infty} \exp \left\{ -\frac{(x-w)^2}{2\tau^2} \right\} dw dx.$$

Because $f(a, b) = f(a-b, 0)$, we write $g(c) = f(c, 0)$, and (47) can be written as:

$$k_{II}^1([a_1, b_1], [a_2, b_2]) = g(a_1 - b_2) - g(a_1 - b_1) - g(a_2 - b_2) + g(a_2 - b_1). \quad (49)$$

To calculate $g(c)$, define the affine transformations $s = \frac{(x+w-c)}{\sqrt{2}}$, $t = \frac{(w-x+c)}{\sqrt{2}}$. Then

$$\begin{aligned} g(c) &= \int_0^{+\infty} \int_{-t}^t \exp \left\{ -\frac{\sqrt{2}t-c}{2\tau^2} \right\} ds dt = \int_0^{+\infty} t \exp \left\{ -\frac{(t-c)^2}{2\tau^2} \right\} dt \\ &= \frac{1}{2} \int_0^{+\infty} (t-c) \exp \left\{ -\frac{(t-c)^2}{2\tau^2} \right\} dt + c \int_0^{+\infty} \exp \left\{ -\frac{(t-c)^2}{2\tau^2} \right\} dt \\ &= \tau^2 \exp \left\{ -\frac{c^2}{2\tau^2} \right\} + c\tau\sqrt{2\pi}\Phi \left(\frac{c}{\tau} \right). \end{aligned} \quad (50)$$

Combining (49) and (50), we have the following closed form expression for (47):

$$\begin{aligned} k_{II}^1([a_1, b_1], [a_2, b_2]) &= \tau^2 \exp \left\{ -\frac{(a_1 - b_2)^2}{2\tau^2} \right\} + (a_1 - b_2)\tau\sqrt{2\pi}\Phi \left(\frac{a_1 - b_2}{\tau} \right) \\ &\quad - \tau^2 \exp \left\{ -\frac{(a_1 - b_1)^2}{2\tau^2} \right\} + (a_1 - b_1)\tau\sqrt{2\pi}\Phi \left(\frac{a_1 - b_1}{\tau} \right) \\ &\quad - \tau^2 \exp \left\{ -\frac{(a_2 - b_2)^2}{2\tau^2} \right\} + (a_2 - b_2)\tau\sqrt{2\pi}\Phi \left(\frac{a_2 - b_2}{\tau} \right) \\ &\quad + \tau^2 \exp \left\{ -\frac{(a_2 - b_1)^2}{2\tau^2} \right\} + (a_2 - b_1)\tau\sqrt{2\pi}\Phi \left(\frac{a_2 - b_1}{\tau} \right). \end{aligned} \quad (51)$$

Without the closed form expressions in (48) and (51), $k_{PI}^1(x, [a, b])$ and $k_{II}^1([a_1, b_1], [a_2, b_2])$ can only be calculated using numerical integration algorithms such as quadrature. As

sample size and dimension of \mathcal{D} increase, the computational time can increase significantly. For illustration, consider the Gaussian correlation function in (42) with $\mathcal{D} = [0, 1]^d$, $\tau_1 = \dots = \tau_d = 1$ and $\sigma^2 = 1$. We compare the computational time of calculating the covariance matrix K using the two methods: the closed form expressions in (48) and (51) and numerical integrations with quadrature, for different sample sizes ($n=n_1=n_2$) and values of d . For each comparison, the inputs are generated as follows. For point sites, we use Latin hypercube samples. For regions, we first fix their lowest corners with Latin hypercube samples, and then extend the intervals for each dimension with lengths independently drawn from $U[0,0.4]$. If the upper limit of the interval exceeds 1, we force it to be 1 so that the corresponding rectangular region is within the design space. For example, for an arbitrary region, suppose the lowest corner of the i th dimension is 0.7. If the sampled length is 0.2, the interval for the i th dimension is $[0.7,0.9]$. On the other hand, if the sampled length is 0.4, the interval for the i th dimension is $[0.7,1]$. The computational time is summarized in Table 19. As is easily seen, the use of the closed form expressions only takes a fraction (from 1/30 to 1/4) of the time than the use of the quadrature method. By examining the numbers further, we notice that both methods are significantly affected by the sample sizes: for fixed d , as sample sizes increase n , the computational time is nearly a quadratic function of n for both methods. However, the dimension d has little effect on the closed form calculation: for given n , as d increases, the computational time for the closed form expressions remains almost the same, which for the quadrature method, the computational time is approximately a linear function of n .

4.3 Examples

In this section, we shall demonstrate the use and performance of the proposed PIG process model with two simulated examples. We adopt the Gaussian correlation function in (42) for the underlying point GP model and apply the closed form expressions

Table 19: CPU Time Comparison.

		n=10	n=50	n=100	n=200	n=1000
d=1	Closed form	0.0468	0.6396	2.2464	9.8281	233.6427
	Quadrature	0.1560	2.6208	9.8437	41.5119	1000.3
d=2	Closed form	0.0468	0.7644	2.9952	11.6845	288.5394
	Quadrature	0.2184	5.0388	20.0461	76.6121	1944.1
d=3	Closed form	0.0312	0.6864	3.1044	12.0433	290.8483
	Quadrature	0.4368	7.3320	29.5466	117.5780	2899.3
d=5	Closed form	0.0312	0.7800	3.0108	11.7625	291.8623
	Quadrature	0.5460	11.9497	48.7815	198.8328	4814.4
d=10	Closed form	0.0312	0.7800	2.9640	12.3241	296.3239
	Quadrature	0.9672	24.1646	98.0154	389.4637	9602.8

in (48) and (51) in the computations.

4.3.1 Example 1

Let the true function be $y(x) = \exp(-1.4x) \cos(3.5\pi x)$ with the design space $\mathcal{D} = [0, 1]$. The curve of the true function is plotted in Figure 18.

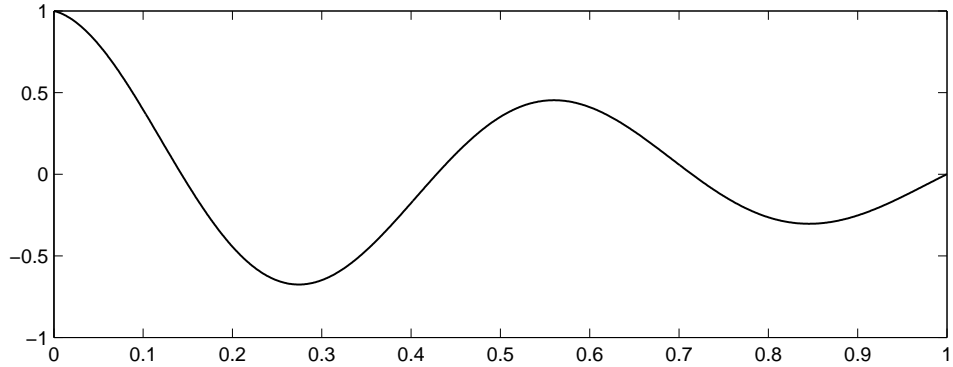


Figure 18: True Function Curve, Example 1.

Table 20 shows the training data used to build the model, which consist of five point sites and two intervals. The corresponding point and integral responses are listed in the last column of the table. We put both types of responses in one column vector denoted by $\mathbf{y} = (y_1, \dots, y_7)'$.

Table 20: Training Data, Example 1.

x_1	0	y_1	1
x_2	0.1	y_2	0.3947
x_3	0.5	y_3	0.3511
x_4	0.85	y_4	-0.3033
x_5	0.99	y_5	-0.0274
R_1	[0.2,0.35]	y_6	-0.0901
R_2	[0.6,0.7]	y_7	0.0252

The mean function of the underlying point GP model is assumed to be constant, i.e., $\mu_P(x) = \mu_0$. Then for any interval $[a, b] \subset [0, 1]$, its mean function can be written as $\mu_I([a, b]) = (b - a)\mu_0$. Let the corresponding mean function be $\mu = \mu_0 \mathbf{e}$, we have $\mathbf{e} = (1, 1, 1, 1, 1, 0.15, 0.1)'$.

The covariance matrix is organized in the 2×2 blocks:

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \sigma^2 k = \sigma^2 \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix},$$

where, for example,

$$k_{12} = \begin{pmatrix} k_{PI}(0, [0.2, 0.35]) & k_{PI}(0, [0.6, 0.7]) \\ k_{PI}(0.1, [0.2, 0.35]) & k_{PI}(0.1, [0.6, 0.7]) \\ k_{PI}(0.5, [0.2, 0.35]) & k_{PI}(0.5, [0.6, 0.7]) \\ k_{PI}(0.85, [0.2, 0.35]) & k_{PI}(0.85, [0.6, 0.7]) \\ k_{PI}(0.99, [0.2, 0.35]) & k_{PI}(0.99, [0.6, 0.7]) \end{pmatrix}.$$

Its entry $k_{PI}(x, [a, b]) = Cov(Z_P(x), Z_I([a, b]))$, as defined in (36), can be calculated by $k_{PI}(x, [a, b]) = \int_a^b k_{PP}(x, w) dw$. Similarly, we have

$$k_{22} = \begin{pmatrix} k_{II}([0.2, 0.35], [0.2, 0.35]) & k_{II}([0.2, 0.35], [0.6, 0.7]) \\ k_{II}([0.6, 0.7], [0.2, 0.35]) & k_{II}([0.6, 0.7], [0.6, 0.7]) \end{pmatrix}.$$

Its entry $k_{II}([a, b], [c, d])$, as defined in (35), can be calculated by $k_{II}([a, b], [c, d]) = \int_a^b \int_c^d k_{PP}(x, w) dw dx$.

The log-likelihood function, apart from additive and multiplicative constants, can be written as:

$$l(\mu_0, \sigma^2, \tau) = -7 \log \sigma^2 - \log |k(\tau)| - \frac{1}{\sigma^2} (\mathbf{y} - \mu_0 \mathbf{e})' k^{-1}(\tau) (\mathbf{y} - \mu_0 \mathbf{e}),$$

where the unknown parameters consist of the regression coefficient μ_0 , the variance σ^2 of the underlying point GP model, and the parameter τ in the covariance function.

The maximum likelihood estimates of the unknown parameters can be obtained through derivative-free unconstrained optimization algorithm in conjunction with $\hat{\mu}_0 = \frac{\mathbf{e}' k^{-1} \mathbf{y}}{\mathbf{e}' k^{-1} \mathbf{e}}$ and $\hat{\sigma}^2 = \frac{1}{7} (\mathbf{y} - \hat{\mu}_0 \mathbf{e})' k^{-1} (\mathbf{y} - \hat{\mu}_0 \mathbf{e})$. The maximum likelihood is achieved at $\hat{\tau} = \pm 0.1138$. The corresponding MLE values of μ_0 and σ^2 are $\hat{\mu}_0 = 0.1263$ and $\hat{\sigma}^2 = 0.2617$.

The predictive curve is drawn by applying the empirical BLUP for untried sites in (39). As can be seen Figure 19, the predictive curve passes through the point responses of the training data, and is very close to the true function. For the first trough around $x = 0.27$, though we do not have any point response in that area, the integral information helps recover the trend successfully.

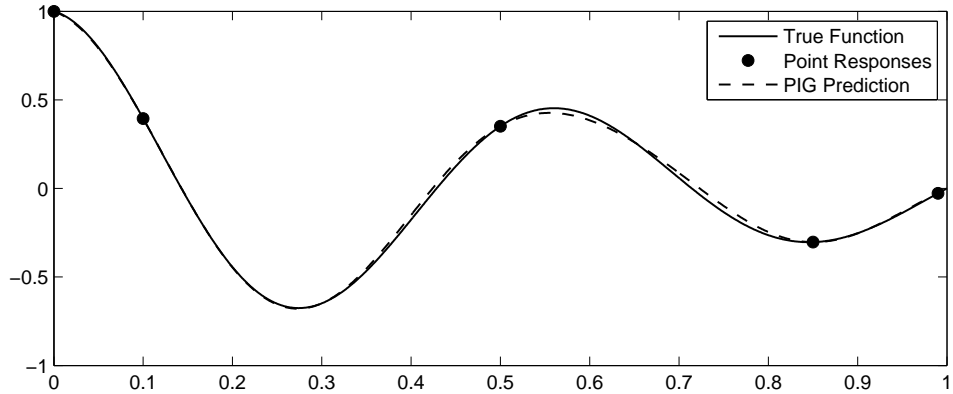


Figure 19: True and PIG Prediction Curves, Example 1.

We compare the proposed PIG process model with point GP models. Because point GP models cannot handle integral information, one can either fit the model

without the integral data, or transfer the integral responses into point responses. Three possibilities are considered below.

The first point GP model is built with only the five point responses in the training data, denoted by GP(5). By assuming the constant mean function, the MLEs for the three unknown parameters are $\hat{\tau} = \pm 0.0783$, $\hat{\mu}_0 = 0.2575$ and $\hat{\sigma}^2 = 0.1902$. The predicative curve, drawn by applying (41), is shown in Figure 20. The predicative curve still passes through the five training point data, but the first trough of the true curve is not captured due to the lack of information in that area.

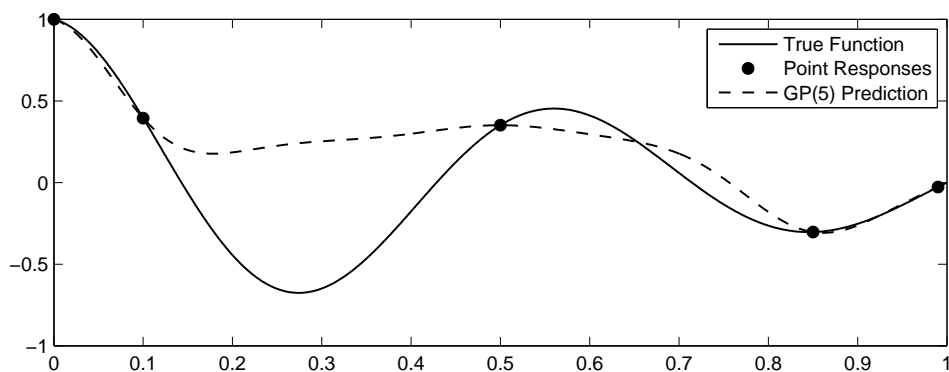


Figure 20: True and GP(5) Prediction Curves, Example 1.

The second point GP model utilizes the integral information by considering them as the average values of the true function within the specified intervals. That is, transfer the input interval into its center point, and transfer the integral response into the average value of the true function within the interval, which is the integral response divide by the length of the interval. The two integral responses in the training data are now considered as two point responses (0.275,-0.6007) and (0.65,0.252). We denote this model by GP(5+2). By assuming the constant mean function, with the seven point responses, the MLEs for the three unknown parameters are $\hat{\tau} = \pm 0.1170$, $\hat{\mu}_0 = 0.1464$ and $\hat{\sigma}^2 = 0.2465$. The predicative curve, drawn by applying (41), is shown in Figure 21. The original point responses are marked in solid circles and the point responses transferred from integral responses are marked in solid triangles. The

predicative curve passes through all seven points responses, but the first trough of the true function is not fully captured because the true value -0.6757 at $x = 0.275$ differs from the transferred point response value -0.6007 .

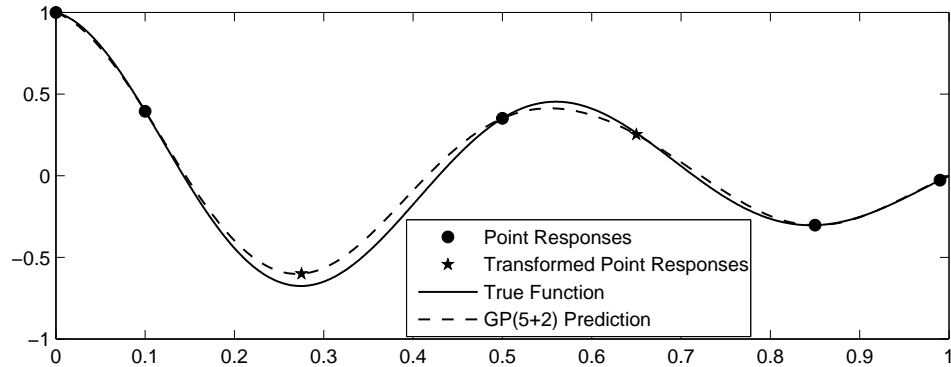


Figure 21: True and GP(5+2) Prediction Curves, Example 1.

In the last point GP model, we are assumed to have the true point responses at the middle of the two intervals, i.e., we have two more point responses $(0.275, -0.6757)$ and $(0.65, .)$. We denoted this model by GP(7). By assuming the constant mean function, with these seven point responses, the MLEs are $\hat{\tau} = \pm 0.1144$, $\hat{\mu}_0 = 0.1306$ and $\hat{\sigma}^2 = 0.2612$. The predicative curve, drawn by applying (41), is shown in Figure 22. With the help of the true point response at $x = 0.275$, this model captures the first trough of the true function successfully. Overall, the predictive curve of GP(7) is very close to the true function. By comparing it with Figure 19, it is noticed that the PIG process has also achieved the same level of predication accuracy by using *only* the integral information. In this example, it is assumed that only integral responses are available for the two intervals. The additional two point responses used in GP(7) are purely hypothetical. From the comparison, we can see that, by using integral responses in the PIG process model, we can achieve the same level of prediction accuracy as GP(7), which requires additional point responses from the two designed regions.

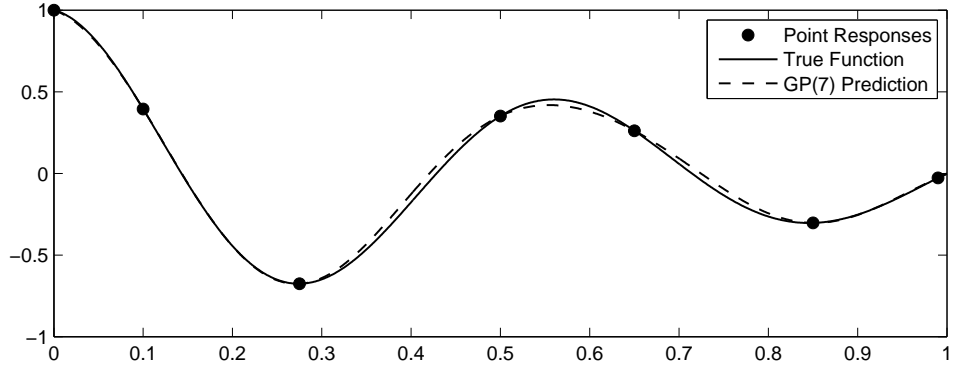


Figure 22: True and GP(7) Prediction Curves, Example 1.

4.3.2 Example 2

We now consider a two-dimensional case with the true function

$$y(x_1, x_2) = \left[4 - 2(2x_1 - 1)^2 + \frac{1}{3}(2x_2 - 1)^3 \right] x_1 + [-4 + 4(2x_2 - 1)^2] (2x_2 - 1)^2.$$

Let the design space be $\mathcal{D} = [0, 1] \times [0, 1]$. The 3D plot and the contour plot of the true function are shown in Figure 23.

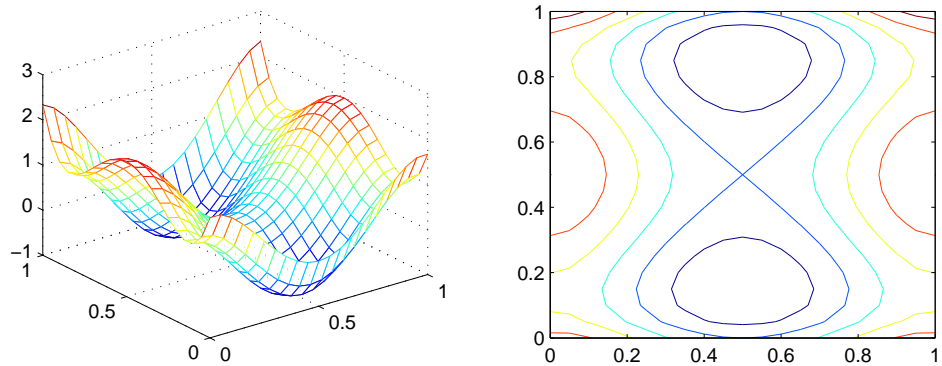


Figure 23: 3D Plot and Contour Plot of the True Function. Example 2.

Table 21 shows the training data used to build the model, which consists of 10 Latin hypercube samples (point sites) and four rectangular regions. The design sites and design regions are plotted in Figure 24. The corresponding point and integral response values are listed in the last column of the table and denoted by one column vector $\mathbf{y} = (y_1, \dots, y_{14})'$.

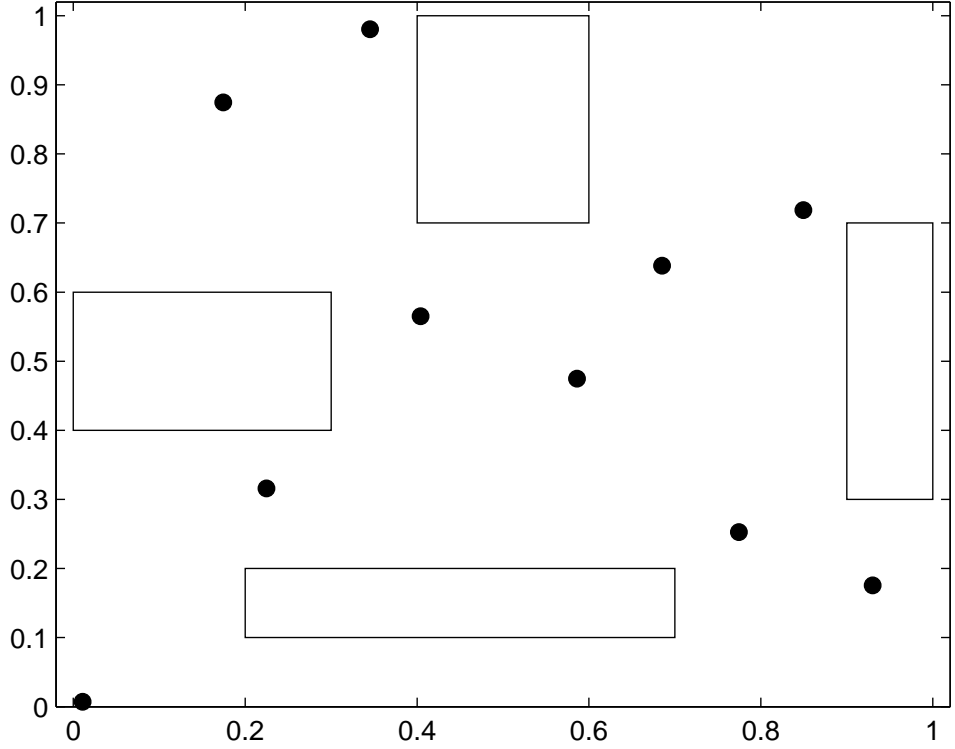


Figure 24: Plot of Training Data. Example 2.

The mean function of the underlying point GP model is assumed to be constant, i.e., $\mu_P(x) = \mu_0$. Then for any rectangular region $R = [a, b] \times [c, d] \subset \mathcal{D}$, its mean function can be written as $\mu_I([a, b] \times [c, d]) = (b - a)(d - c)\mu_0$. Let the corresponding mean function be $\mu = \mu_0 \mathbf{e}$, we have $\mathbf{e} = (1, , 1, 0.05, 0.06, 0.04, 0.06)'$.

The covariance matrix is organized in the 2×2 blocks:

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \sigma^2 k = \sigma^2 \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix},$$

where k_{12} is a 10×4 matrix representing the pseudo-correlations between any design site and any design region. For example, the first entry of k_{12} is the pseudo-correlation between $x_1 = (0.2249, 0.3159)$ and $R_1 = [0.2, 0.7] \times [0.1, 0.2]$. By (36). It can be

Table 21: Training Data, Example 2.

x_1	(0.2249,0.3159)	y_1	0.5541
x_2	(0.1745,0.8746)	y_2	0.4104
x_3	(0.8494,0.7186)	y_3	0.8717
x_4	(0.9301,0.1756)	y_4	0.8225
x_5	(0.0109,0.0071)	y_5	1.5815
x_6	(0.4041,0.5650)	y_6	0.0781
x_7	(0.3453,0.9803)	y_7	0.1084
x_8	(0.5861,0.4749)	y_8	0.1069
x_9	(0.6851,0.6384)	y_9	0.2285
x_{10}	(0.7742,0.2527)	y_{10}	0.2710
R_1	$[0.2,0.7] \times [0.1,0.2]$	y_{11}	-0.0323
R_2	$[0.4,0.6] \times [0.7,1]$	y_{12}	-0.0423
R_3	$[0.9,1] \times [0.3,0.7]$	y_{13}	0.0686
R_4	$[0,0.3] \times [0.4,0.6]$	y_{14}	0.0821

calculated as:

$$\begin{aligned}
 & k_{PI}((0.2249, 0.3159), [0.2, 0.7] \times [0.1, 0.2]) \\
 &= \int_{0.2}^{0.7} \int_{0.1}^{0.2} k_{PP}((0.2249, 0.3159), (x_1, x_2)) dx_2 dx_1. \quad (52)
 \end{aligned}$$

By separability of $k_{PP}(x, w)$, (52) can be written as:

$$\left(\int_{0.2}^{0.7} k_{PP}^1(0.2249, x_1) dx_1 \right) \left(\int_{0.1}^{0.2} k_{PP}^2(0.3159, x_2) dx_2 \right).$$

Additionally, k_{22} is the 4×4 pseudo-correlation matrix of the five design regions. For example, its entry at row 1 and column 2 is the pseudo-correlation between $R_1 = [0.2, 0.7] \times [0.1, 0.2]$ and $R_2 = [0.4, 0.6] \times [0.7, 1]$. By (35). It can be calculated as:

$$\begin{aligned}
 & k_{II}([0.2, 0.7] \times [0.1, 0.2], [0.4, 0.6] \times [0.7, 1]) \\
 &= \int_{0.2}^{0.7} \int_{0.1}^{0.2} \int_{0.4}^{0.6} \int_{0.7}^1 k_{PP}((x_1, x_2), (w_1, w_2)) dw_2 dw_1 dx_2 dx_1. \quad (53)
 \end{aligned}$$

By separability of $k_{PP}(x, w)$, (53) can be written as:

$$\left(\int_{0.2}^{0.7} \int_{0.4}^{0.6} k_{PP}^1(x_1, w_1) dw_1 dx_1 \right) \left(\int_{0.1}^{0.2} \int_{0.7}^1 k_{PP}^2(x_2, w_2) dw_2 dx_2 \right).$$

Other entries of k_{12} and k_{22} can be similarly calculated.

The log-likelihood function, apart from additive and multiplicative constants, can be written as:

$$l(\mu_0, \sigma^2, \boldsymbol{\tau}) = -14 \log \sigma^2 - \log |k(\boldsymbol{\tau})| - \frac{1}{\sigma^2} (\mathbf{y} - \mu_0 \mathbf{e})' k^{-1}(\boldsymbol{\tau}) (\mathbf{y} - \mu_0 \mathbf{e}),$$

where the unknown parameters are μ_0 , σ^2 and $\boldsymbol{\tau} = (\tau_1, \tau_2)'$. The MLEs of the unknown parameters can be obtained through derivative-free unconstrained optimization algorithm in conjunction with $\hat{\mu}_0 = \frac{\mathbf{e}' k^{-1} \mathbf{y}}{\mathbf{e}' k^{-1} \mathbf{e}}$ and $\hat{\sigma}^2 = \frac{1}{14} (\mathbf{y} - \hat{\mu}_0 \mathbf{e})' k^{-1} (\mathbf{y} - \hat{\mu}_0 \mathbf{e})$. The maximal likelihood is achieved at $\hat{\tau}_1 = \pm 0.2321$ and $\hat{\tau}_2 = \pm 0.2185$. The corresponding MLE values of μ_0 and σ^2 are $\hat{\mu}_0 = 0.6998$ and $\hat{\sigma}^2 = 0.7887$.

The predictive contour of the PIG process model, shown in Figure 25b, is drawn by applying (39). As we compare it with the contour of the true function in Figure 25a, it is immediately clear that, this predictive contour captures the major trends of the true function. The overall predictive accuracy is quite satisfactory, considering that we only have 10 point and 4 integral responses.

As in Example 1, we compare the PIG process model with two point GP models. The first point GP model is fitted only with the 10 point responses in the training data. We denote this model by GP(10). The 10 Latin Hypercube samples are plotted with the contour of the true function in Figure 26.

It is noticed that, there are no point sites around the two local minima of the true function. Therefore, the contour changes in these regions may not be captured by the corresponding point GP model. By assuming the constant mean function, with the 10 point responses, the MLEs are $\hat{\tau}_1 = \pm 0.2006$, $\hat{\tau}_2 = \pm 0.8177$, $\hat{\mu}_0 = 0.8242$ and $\hat{\sigma}^2 = 0.2591$. The predictive contour of GP(10), shown in Figure 25c, is drawn by applying (41). From the contour, we can see that the predictive function is nearly

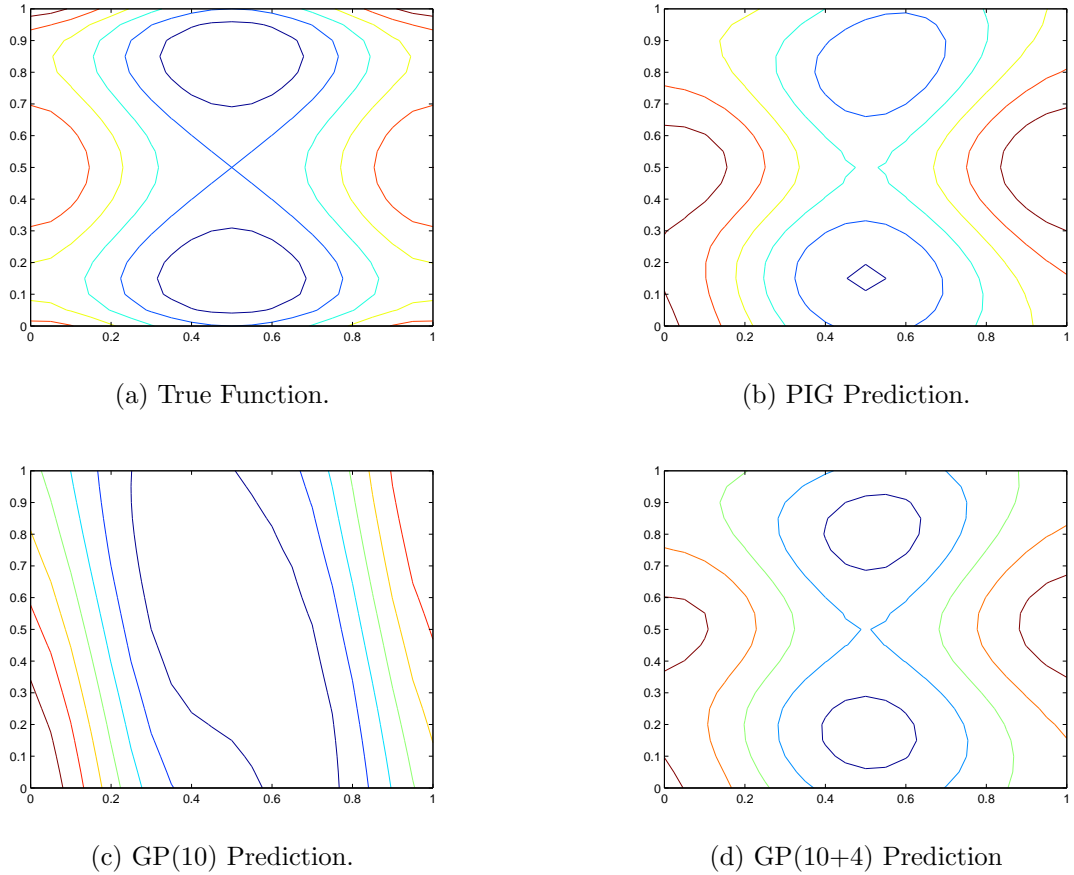


Figure 25: True and Predictive Contour Plots, Example 2.

constant over x_2 . Thus, it cannot capture the changes over x_2 . By comparing it with Figure 25a, it is clear that the overall predictive accuracy is very low.

In the second point GP model, we transfer the integral data into four point responses in the following way: for the input, transfer the rectangular region into its center point; for the response, transfer the integral value into the average value of the true function within the region, i.e., use the integral value divide by the area of the rectangular region. Now we have four more point responses listed in Table 22. We denote this model by GP(10+4).

The scatter plot of these 14 point sites is shown in Figure 27 with the contour of the true function. The original 10 point sites are marked in solid circles and the four transferred point sites are marked in stars. It is clear that these 14 point sites have

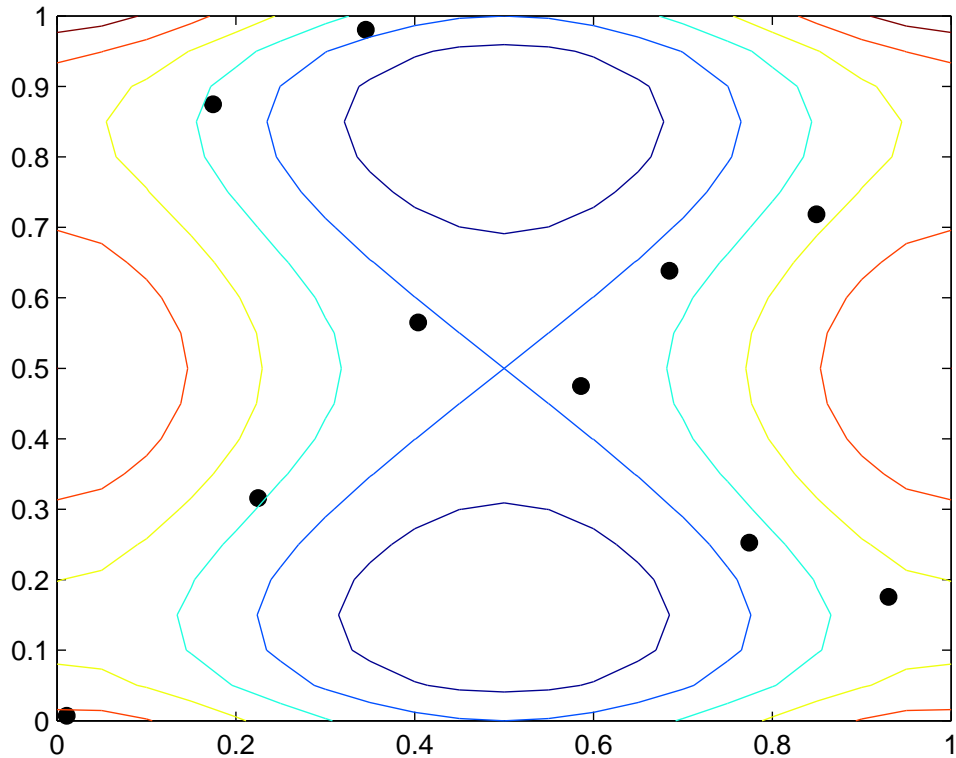


Figure 26: Scatter Plot of Ten Point Sites and Contour Plot of the True Function, Example 2.

a good coverage of the design space. In particular, the regions around the two local minima are now covered by two transferred point responses.

By assuming the constant mean function, with these 14 point responses, the MLEs are $\hat{\tau}_1 = \pm 0.241$, $\hat{\tau}_2 = \pm 0.224$, $\hat{m}u_0 = 0.758$ and $\hat{\sigma}^2 = 0.587$. The predictive contour of GP(10+4), drawn by applying (41), is shown in Figure 25d. By comparing it with Figure 25a, we can see that, this contour also captures the major trends of the true function. By comparing the predictive contours in Figure 25b and Figure 25d, it is hard to decide whether the PIG process model is better than GP(10+4). We can use the predictive errors for the point sites x_{11}, \dots, x_{14} for further comparison, which is given in Table 23. It is clear that all predictive errors are much larger for GP(10+4) than the PIG process model, This is because the true point response values

Table 22: Additional Four Point Data, Example 2.

x_{11}	(0.45,0.15)	y'_{11}	-0.6464	y''_{11}	-0.9609
x_{12}	(0.5,0.85)	y'_{12}	-0.7058	y''_{12}	-0.9996
x_{13}	(0.95,0.5)	y'_{13}	1.7158	y''_{13}	1.9278
x_{14}	(0.15,0.5)	y'_{14}	1.3681	y''_{14}	1.4798

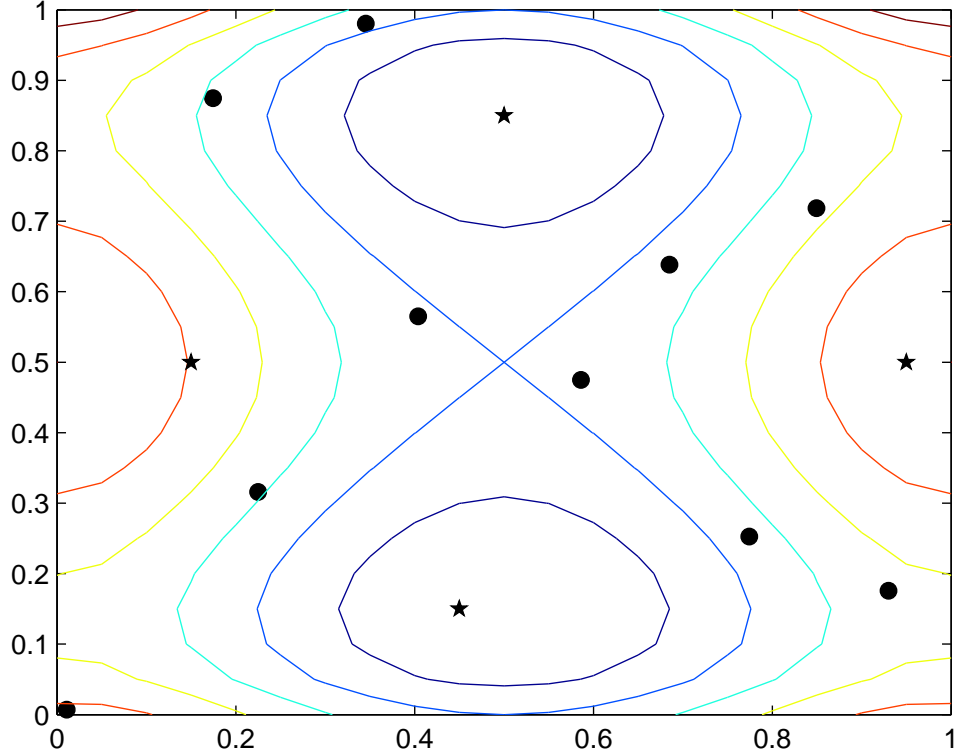


Figure 27: Scatter Plot of Fourteen Point Sites and Contour Plot of the True Function, Example 2.

at these four point sites (column 6 of Table 22), are different from their corresponding transferred point response values (column 4 of Table 22). The interpolating property of (9) forces the predicative contour of GP(10+4) to pass through the four wrong point responses, which makes the corresponding predicative errors large. In this example, the two regions R_1 and R_2 , as well as their corresponding center points x_{11} and x_{12} are around the two minima of the true function. If the objective of the study is to

Table 23: Comparison of Prediction Errors for the Four Point Data, Example 2.

Point Input	Error for PIG	Error for GP(10+5)	True Response
x_{11}	-0.0356	0.3145	-0.9609
x_{12}	0.0914	0.2938	-0.9996
x_{13}	0.0022	-0.2120	1.9278
x_{14}	0.0475	-0.1117	1.4798

find the minima of the true function, the PIG process model is clearly a better choice.

4.4 Conclusions

In this paper, we have proposed a new Gaussian process model, called the PIG process, to jointly model the point and integral responses. Parameter estimation and prediction based on maximum likelihood estimation are developed. In the important case of axis-parallel rectangular regions, closed form expressions for the covariance functions are derived, which can substantially reduce the computational time. Two simulated examples are used to demonstrate the use and performance of the PIG process model. Two point GP models, one ignoring the integral information and the other replacing it with point responses, are compared with the proposed model. It is shown that the integral information can be very helpful in fitting the point response model for the true function. Furthermore, it performs better than treating it as point responses, and almost as good as having extra point responses in the design regions. The computation of the PIG process model requires inverse calculations of the pseudo-correlation matrix k , which can be time consuming and unstable. Computational tricks are needed to solve this problem. Moreover, the MLEs of $\boldsymbol{\tau}$ require optimizations of a complicated objective function. A more efficient algorithm is needed to handle high dimensional inputs and large sample size.

For simplicity, in this work, we have only considered the Gaussian correlation function and axis-parallel rectangular regions, and followed the frequentist approach.

Similar ideas can be generalized to other choices of correlation functions and regions, and the computational issues with the covariance matrix K can be discussed in future research. A Bayesian version of the PIG process model can be extended as well. In the point GP model, space filling designs such as Latin hypercube samples (Santner et al., 2003) are chosen for computer experiments. In the PIG model, how to simultaneously choose the point sites and regions are more challenging. A generalized notion of space filling can be defined and criteria be proposed by employing new definitions of distance.

REFERENCES

- [1] ASH, T., *Landing Page Optimization: the Definitive Guide to Testing and Tuning for Conversions*. Indianapolis: Wiley, 2009.
- [2] ASHRAE, *ASHRAE Handbook Fundamentals*. American Society of Heating, Refrigeration and Air-Conditioning Engineers, SI ed., 2009.
- [3] ASME, *Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer*. American Society of Mechanical Engineers, 2009.
- [4] BEDFORD, T. and COOKE, R., *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, 2001.
- [5] BOX, G. E. P., HUNTER, J. S., and HUNTER, W. G., *Statistics for Experimenters: Design, Innovation, and Discovery*. New York: Wiley, 2 ed., 2005.
- [6] BRAJAC, M. and MOREY, C., “GM10 chassis body marriage simulation: a designed experiment approach,” *Journal of Designed Experiment Case Studies*, vol. 1, no. 1, pp. 4–31, 1987. GM of Canada.
- [7] CHEN, W., XIONG, Y., TSUI, K.-L., and WANG, S., “A design-driven validation approach using bayesian prediction models,” *Journal of Mechanical Design*, vol. 130, no. 2, 2008.
- [8] COOKE, R. M., “Uncertainty analysis comes to integrated assessment models for climate change and conversely,” *Climatic Change*, vol. 117, no. 3, pp. 467–479, 2013.
- [9] DAVID, M., LAURET, P., and BOLAND, J., “Evaluating tilted plane models for solar radiation using comprehensive testing procedures, at a southern hemisphere location,” *Renewable Energy*, vol. 51, pp. 124–131, 2013.
- [10] DE WIT, S. and AUGENBROE, G., “Analysis of uncertainty in building design evaluations and its implications,” *Energy and Buildings*, vol. 34, no. 9, pp. 951–958, 2002.
- [11] DERU, M., FIELD, K., STUDER, D., BENNE, K., GRIFFITH, B., TORCELLINI, P., LIU, B., HALVERSON, M., WINIARSKI, D., ROSENBERG, M., and OTHERS, *U.S. Department of Energy Commercial Reference Building Models of the National Building Stock*. National Renewable Energy Laboratory, 2011.
- [12] DIEZ-MEDIAVILLA, M., DE MIGUEL, A., and BILBAO, J., “Measurement and comparison of diffuse solar irradiance models on inclined surfaces in valladolid (spain),” *Energy conversion and management*, vol. 46, no. 13, pp. 2075–2092, 2005.

- [13] DOMÍNGUEZ-MUÑOZ, F., ANDERSON, B., CEJUDO-LÓPEZ, J. M., and CARRILLO-ANDRÉS, A., “Uncertainty in the thermal conductivity of insulation materials,” *Energy and buildings*, vol. 42, no. 11, pp. 2159–2168, 2010.
- [14] DONATO, D., LAURA, L., LEONARDI, S., and MILLOZZI, S., “Large scale properties of the webgraph,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 239–243, 2004.
- [15] ENERGYPLUS, “Energyplus engineering reference,” 2012. Retrieved 2012-11-26. http://apps1.eere.energy.gov/buildings/energyplus/energyplus_documentation.cfm.
- [16] EVSEEV, E. G. and KUDISH, A. I., “The assessment of different models to predict the global solar radiation on a surface tilted to the south,” *Solar Energy*, vol. 83, no. 3, pp. 377–388, 2009.
- [17] FINNEY, D. J., “The fractional replication of factorial arrangements,” *Annals of Eugenics*, vol. 12, no. 1, pp. 291–301, 1943.
- [18] GUEYMARD, C., “An anisotropic solar irradiance model for tilted surfaces and its comparison with selected engineering algorithms,” *Solar Energy*, vol. 38, no. 5, pp. 367–386, 1987.
- [19] GUEYMARD, C. A., “Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications,” *Solar Energy*, vol. 83, no. 3, pp. 432–444, 2009.
- [20] HAMADA, M. S. and WU, C. F. J., “Analysis of designed experiments with complex aliasing,” *Journal of Quality Technology*, vol. 24, no. 3, pp. 130–137, 1992.
- [21] HELTON, J. C. and DAVIS, F. J., “Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems,” *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003.
- [22] HELTON, J. C., HANSEN, C. W., and SALLABERRY, C. J., “Uncertainty and sensitivity analysis in performance assessment for the proposed high-level radioactive waste repository at yucca mountain, nevada,” *Reliability Engineering & System Safety*, vol. 107, pp. 44–63, 2012.
- [23] HEO, Y., AUGENBROE, G., and CHOUDHARY, R., “Quantitative risk management for energy retrofit projects,” *Journal of Building Performance Simulation*, vol. 6, no. 4, pp. 257–268, 2013.
- [24] HILLS, R. G., DOWDING, K. J., and SWILER, L., “Thermal challenge problem: summary,” *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29, pp. 2490–2495, 2008.

- [25] HILLS, R. G. and DOWDING, K. J., “Multivariate approach to the thermal challenge problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29, pp. 2442–2456, 2008.
- [26] IGAWA, N., KOGA, Y., MATSUZAWA, T., and NAKAMURA, H., “Models of sky radiance distribution and sky luminance distribution,” *Solar Energy*, vol. 77, no. 2, pp. 137–157, 2004.
- [27] JUDKOFF, R., WORTMAN, D. N., and BURCH, J., “Empirical validation using data from the seri class-a validation house,” tech. rep., Solar Energy Research Institute, Golden, Colorado, 1983.
- [28] KAMBEZIDIS, H. D., PSILOGLOU, B. E., and GUEYMARD, C., “Measurements and models for total solar irradiance on inclined surface in athens, greece,” *Solar Energy*, vol. 53, no. 2, pp. 177–185, 1994.
- [29] KENNEDY, M. C. and O’HAGAN, A., “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.
- [30] LEE, B. D., SUN, Y., HU, H., AUGENBROE, G., and PAREDIS, C. J., “A framework for generating stochastic meteorological years for risk-conscious design of buildings,” in *Proceedings of IBPSA-USA 2012*, 2012.
- [31] LIU, B. and JORDAN, R., “Daily insolation on surfaces tilted towards equator,” *Trans ASHRAE*, vol. 56, pp. 526–541, 1961.
- [32] LOUTZENHISER, P. G., MANZ, H., FELSMANN, C., STRACHAN, P. A., FRANK, T., and MAXWELL, G. M., “Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation,” *Solar Energy*, vol. 81, no. 2, pp. 254–267, 2007.
- [33] MACDONALD, I. and STRACHAN, P., “Practical application of uncertainty analysis,” *Energy and Buildings*, vol. 33, no. 3, pp. 219–227, 2001.
- [34] MACDONALD, I. A., *Quantifying the effects of uncertainty in building simulation*. PhD thesis, University of Strathclyde, 2002.
- [35] MONTGOMERY, D. C., *Design and Analysis of Experiments*. New York: Wiley, 3 ed., 1991.
- [36] MONTGOMERY, D. C., *Design and Analysis of Experiments*. New York: Wiley, 8 ed., 2012.
- [37] MORRIS, M. D., MITCHELL, T. J., and YLVISAKER, D., “Bayesian design and analysis of computer experiments: use of derivatives in surface prediction,” *Technometrics*, vol. 35, no. 3, pp. 243–255, 1993.
- [38] MUNEEER, T., *Solar Radiation and Daylight Models*. Routledge, 2007.

- [39] NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., and WASSERMAN, W., *Applied Linear Statistical Models*. Chicago: Irwin, 4 ed., 1996.
- [40] NOORIAN, A. M., MORADI, I., and KAMALI, G. A., “Evaluation of 12 models to estimate hourly diffuse irradiation on inclined surfaces,” *Renewable Energy*, vol. 33, no. 6, pp. 1406–1412, 2008.
- [41] NRC, *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. National Academies Press, 2012.
- [42] OBERKAMPF, W. L. and ROY, C. J., *Verification and Validation in Scientific Computing*. New York: Cambridge University Press, 2010.
- [43] OBERKAMPF, W. L., TRUCANO, T. G., and HIRSCH, C., “Verification, validation, and predictive capability in computational engineering and physics,” *Applied Mechanics Reviews*, vol. 57, no. 5, pp. 345–384, 2004.
- [44] OZOLINS, M., “What is a conversion funnel,” August 2012. Retrieved 2014-6-6. <http://www.webics.com.au/blog/conversion-tracking/conversion-funnel/>.
- [45] PALYVOS, J. A., “A survey of wind convection coefficient correlations for building envelope energy systems modeling,” *Applied Thermal Engineering*, vol. 28, no. 8, pp. 801–808, 2008.
- [46] PEREZ, R., INEICHEN, P., SEALS, R., MICHALSKY, J., and STEWART, R., “Modeling daylight availability and irradiance components from direct and global irradiance,” *Solar Energy*, vol. 44, no. 5, pp. 271–289, 1990.
- [47] QIAN, P. Z. G. and WU, C. F. J., “Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments,” *Technometrics*, vol. 50, no. 2, pp. 192–204, 2008.
- [48] QUALAROO, “*The Beginner’s Guide to Conversion Rate Optimization*.” Retrieved 2014-6-6. <https://qualaroo.com/beginners-guide-to-cro/>.
- [49] REDA, I., *Method to Calculate Uncertainties in Measuring Shortwave Solar Irradiance Using Thermopile and Semiconductor Solar Radiometers*. NREL, 2011.
- [50] ROY, C. J. and OBERKAMPF, W. L., “A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing,” *Computer Methods in Applied Mechanics and Engineering*, vol. 200, no. 25, pp. 2131–2144, 2011.
- [51] SANTNER, T. J., WILLIAMS, B. J., and NOTZ, W., *The Design and Analysis of Computer Experiments*. New York: Springer, 2003.
- [52] SEBER, G. A. and LEE, A. J., *Linear Regression Analysis*. John Wiley & Sons, 2 ed., 2012.

- [53] SMRL. <http://solardat.uoregon.edu/SolarData.html>.
- [54] STRACHAN, P., “Model validation using the passys test cells,” *Building and Environment*, vol. 28, no. 2, pp. 153–165, 1993.
- [55] SUN, Y., HEO, Y., TAN, M., XIE, H., JEFF WU, C., and AUGENBROE, G., “Uncertainty quantification of microclimate variables in building energy models,” *Journal of Building Performance Simulation*, vol. 7, no. 1, pp. 17–32, 2014.
- [56] SUN, Y., SU, H., WU, C. J., and AUGENBROE, G., “Uncertainty quantification of solar diffuse irradiation on inclined surfaces for building energy simulation,” in *Proceedings of Building Simulation 2013*, pp. 71–78, 2013.
- [57] UTRILLAS, M. P., MARTINEZ-LOZANO, J. A., and CASANOVAS, A. J., “Evaluation of models for estimating solar irradiation on vertical surfaces at valencia, spain,” *Solar Energy*, vol. 47, no. 3, pp. 223–229, 1991.
- [58] WALKER, W. E., HARREMOËS, P., ROTMANS, J., VAN DER SLUIJS, J. P., VAN ASSELT, M. B., JANSSEN, P., and KRAYER VON KRAUSS, M. P., “Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support,” *Integrated assessment*, vol. 4, no. 1, pp. 5–17, 2003.
- [59] WANG, Q., AUGENBROE, G., and SUN, Y., “The role of construction detailing and workmanship in achieving energy efficient buildings,” in *Proceedings of Construction Research Congress 2014*, pp. 19–24, 2014.
- [60] WILCOX, R. R., *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, 2012.
- [61] WILCOXON, F., “Individual comparisons by ranking methods,” *Biometrics Bulletin*, pp. 80–83, 1945.
- [62] WU, C. F. J., “Post-fisherian experimentation: from physical to virtual,” *Journal of the American Statistical Association*, 2014. accepted.
- [63] WU, C. F. J. and HAMADA, M. S., *Experiments: Planning, Analysis, and Optimization*. New York: Wiley, 2 ed., 2009.
- [64] WYSS, G. D. and JORGENSEN, K. H., *A Users Guide to LHS: Sandia’s Latin Hypercube Sampling Software*. Sandia National Laboratories.
- [65] YANG, H., LU, L., and ZHOU, W., “A novel optimization sizing model for hybrid solar-wind power generation system,” *Solar Energy*, vol. 81, no. 1, pp. 76–84, 2007.