

©2016

Mark David Rodgers

ALL RIGHTS RESERVED

SIMULATION-BASED OPTIMIZATION MODELS FOR ELECTRICITY GENERATION EXPANSION
PLANNING PROBLEMS CONSIDERING HUMAN HEALTH EXTERNALITIES

By

MARK DAVID RODGERS

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Industrial and Systems Engineering

Written under the direction of

Dr. David W. Coit

and approved by

New Brunswick, New Jersey

October, 2016

ABSTRACT OF THE DISSERTATION

Simulation-Based Optimization Models for Electricity Generation Expansion Planning

Problems Considering Human Health Externalities

By Mark David Rodgers

Dissertation Director:

Prof. David W. Coit

This dissertation is focused on the development of mathematical models to solve electricity generation expansion planning problems that minimize total system-wide costs, including human health externalities. A generation expansion planning model is a mathematical optimization framework employed to determine the type of generation technology to invest in, and when and where these investments should be made in order to minimize market costs such as investment costs, fixed and variable operating & maintenance costs, and fuel costs over a long term planning horizon. Fossil fuels (such as coal, oil, and natural gas), which are the primary sources of energy for electricity, are among the most economical sources of electricity. However, burning fossil fuels creates by-products that contribute to ground-level ozone, particulates, and acid rain, which have harmful health effects. Based on EPA research, exposure to these elements causes various respiratory-related illnesses leading to lost days at school or work on a daily basis. In this research, a simulation-based approach is employed to quantify human health externalities by linking the outputs of expansion planning simulations with an EPA

screening tool that determines the human health externalities from the electricity sector. From this data set, a statistical prediction model is employed to approximate health costs as a function of electricity generation. This explicit representation of the relationship between electricity generation and health externalities is then incorporated in the objective function of a generation expansion planning problem as a metamodel or surrogate curve. This research is the first comprehensive attempt to dynamically quantify human health externalities in the context of generation expansion planning. Additionally, this research leads contributions for developing generation expansion planning models considering human health externalities as costs in the objective function. This research also leads contributions for developing large scale simulation-based optimization models, by applying a rigorous search algorithm to determine candidate solution points to enhance prediction capabilities of the metamodel, and thus yield more accurate and realistic optimization solutions.

Acknowledgements

First, I'd like to thank my parents for their unconditional love and unwavering support. My parents always taught me to have confidence in my abilities, to believe in myself, and to accept and embrace life's challenges. Without your guidance, support, and the tough sacrifices you've made for me, I wouldn't have had the access and opportunity to grow into the man I am today. Thank you for buying me Lego sets as a young boy, and for inspiring my interest in math, science, and engineering. Because of you both, I found my passion as a child and had the tools to blaze my own trail in life as an engineer. Most importantly, thank you for encouraging me to be myself. Words cannot express the immense amount of gratitude I owe you both.

I thank my incredible wife, Whitney, for sticking by my side throughout this process, and for embarking on this journey with me. Thank you for being my best friend and for listening to my nerdy dissertation research jargon. And of course, thank you for your encouragement and support. This is just as much your achievement as it is mine.

I owe tremendous gratitude to my co-advisors, Professor David Coit and Professor Frank Felder. Thank you both for your flexibility, guidance, encouragement, and consistent support. Professor Coit, I've always admired your ability to communicate technical information concisely and succinctly, and will emulate that in my professional career. Professor Felder, you've been an immense wealth of knowledge for me, and truly piqued my interest in energy systems. I couldn't have asked for better advisors and mentors.

Many thanks to the professors on my committee for their time, advice, and valuable suggestions. Special thanks to Professor Annmarie Carlton for generously sharing your knowledge on air quality modeling, and to Professor Albin who was the first person who

reached out to me when I enrolled at Rutgers, and has continued to be a positive influence in my academic and professional career. I also want to thank Professor Ying Hung for introducing me to kriging models. I also thank Dr. Saltuk Selcuklu and Dr. Caroline Farkas for collaborating with me on various projects during my time as a graduate student at Rutgers.

Table of Contents

Abstract of the Dissertation	ii
Acknowledgements	iv
List of Tables	ix
List of Figures.....	xi
1. Introduction.....	1
<i>1.1. Electric Power Systems Overview.....</i>	<i>4</i>
<i>1.2. Motivation of Research</i>	<i>6</i>
<i>1.3. Objectives and Research Contributions.....</i>	<i>10</i>
2. Background and Literature Review	13
<i>2.1. Electrical Power Systems Planning</i>	<i>13</i>
<i>2.1.1 Economic Dispatch.....</i>	<i>16</i>
<i>2.1.2 Unit Commitment.....</i>	<i>18</i>
<i>2.1.3 Unit Commitment Dispatch</i>	<i>19</i>
<i>2.2. Generation Expansion Planning Overview.....</i>	<i>21</i>
<i>2.3. Literature Survey.....</i>	<i>24</i>
<i>2.3.1. Health and Environmental Implications of Electricity Dispatch</i>	<i>25</i>
<i>2.3.2. Stochastic GEP Models</i>	<i>36</i>
<i>2.3.3. GEP Models Using Metaheurstics</i>	<i>39</i>
<i>2.4. Metamodeling Techniques.....</i>	<i>41</i>
<i>2.4.1. Experimental Design</i>	<i>41</i>
<i>2.4.2. Metamodels.....</i>	<i>43</i>
<i>2.4.3. Evaluation of Metamodels</i>	<i>49</i>
<i>2.4.4. Metamodels for Optimization.....</i>	<i>50</i>
<i>2.4.5. Applications</i>	<i>54</i>

3. Analytical Framework.....	59
3.1 <i>Establishing Model Relationships.....</i>	59
3.2 <i>Seed Generation via Generation Expansion Planning Simulations and Experimentation</i>	61
3.2.1. <i>Generation Expansion Planning Mathematical Model for Simulation</i>	62
3.2.2. <i>COBRA Overview</i>	71
3.2.3. <i>Simulation and Experimentation Approach for Seed Generation</i>	75
3.3. <i>Constructing a Metamodel of Human Health Externalities.....</i>	78
3.3.1. <i>Kriging Metamodel.....</i>	79
3.4. <i>Implementing a Metamodel in the Objective Function</i>	81
3.4.1. <i>Design Criteria.....</i>	82
3.4.2. <i>Termination Criteria.....</i>	83
4. Reduced-Form Model Evaluation	85
4.1. <i>Reduced-Form Model Formulation</i>	87
4.2. <i>Reduced-Form Test Case Definitions and Assumptions</i>	89
4.3. <i>Dispatching Results.....</i>	98
4.4. <i>Investment Results</i>	102
4.5. <i>Emissions and Human Health Results</i>	104
4.6. <i>Cost and Pricing Results</i>	107
5. Full-Form Model Evaluation	116
5.1 <i>Dispatching Results.....</i>	122
5.2 <i>Investment Results</i>	126
5.3 <i>Emissions and Human Health Results</i>	129
5.4 <i>Cost and Pricing Results</i>	132
5.5 <i>Regional Details – Simulated Health Damages Case.....</i>	141
5.5.1 <i>Maryland/Delaware (MDDE) Regional Summary.....</i>	147

5.5.2 <i>New England (NE) Regional Summary</i>	149
5.5.3 <i>New Jersey (NJ) Regional Summary</i>	152
5.5.4 <i>New York (NY) Regional Summary</i>	154
5.5.5 <i>New York City (NYC) Regional Summary</i>	157
5.5.6 <i>Rest of PJM (RoPJM) Regional Summary</i>	159
6. Conclusions and Research Extensions	162
References	168
Appendix A: Additional Model Assumptions	183

List of Tables

Table 1.1 – Net Electricity Generation by Energy Source (U.S. Energy Information Administration, 2015)	2
Table 2.1 – Related Research Summary	29
Table 2.2 – Basis Functions for RBF Metamodels (Han & Zhang, 2012)	46
Table 3.1 – GEP Model Set Definitions	68
Table 3.2 – GEP Model Parameter Definitions	69
Table 3.3 – COBRA Tools and Data Sources (United States Environmental Protection Agency, 2015).....	73
Table 3.4 – Canned Response Functions (United States Environmental Protection Agency, 2015).....	74
Table 4.1 – Reduced-Form Model Test Cases.....	91
Table 4.2 – Generating Units and Their Costs, Capacities, and Emissions Rates	92
Table 4.3 – Mean Absolute Prediction Error Values from Cross Validation Procedure ..	95
Table 4.4 – Regional Dispatch Summary	100
Table 4.5 – Regional Investment Summary.....	102
Table 4.6 – CO ₂ Emissions Summary.....	104
Table 4.7 – NO _x Emissions Summary.....	104
Table 4.8 – SO ₂ Emissions Summary	104
Table 4.9 – Expected Negative Health Outcomes Summary.....	106
Table 4.10 – Expected Human Health Externalities Summary	112
Table 5.1 – Full Model Test Cases	117
Table 5.2 – Mean Absolute Prediction Error Values from Cross Validation Procedure	119
Table 5.3 – Regional Dispatch Summary	124
Table 5.4 – Regional Summary of New Investments	126
Table 5.5 – CO ₂ Emissions Summary.....	129

Table 5.6 – NO _x Emissions Summary.....	129
Table 5.7 – SO ₂ Emissions Summary	129
Table 5.8 – Expected Negative Health Outcomes Summary.....	131
Table 5.9 – Expected Human Health Externalities Summary	137
Table 5.10 – Expected Health Outcomes.....	145
Table 5.11 – Expected Human Health Externalities	146
Table A.1 – Initial Capacity by Region (in MW) (Cambridge Energy Solutions, 2011)	183
Table A.2 – Transmission Limits by Region (in MWh) (Cambridge Energy Solutions, 2011)	184
Table A.3 – Unit Derating and Capacity Value Percentages (Cambridge Energy Solutions, 2011; U.S. Energy Information Administration, 2015)	184
Table A.4 – Unit Capacity Factors by Region (Cambridge Energy Solutions, 2011; U.S. Energy Information Administration, 2015)	185
Table A.5 – Reserve Margin Percentages.....	185
Table A.6 – RGGI CO ₂ Emissions Limits	185
Table A.7 – Regional Emissions Limits (U.S. Environmental Protection Agency, 2013)	185
Table A.8 – Annual Emissions Limits (U.S. Environmental Protection Agency, 2013)	186
Table A.9 – Transmission Losses	186
Table A.10 – Renewables Trading Network (General Electric International, 2014)	186
Table A.11 – Available Renewables by Region (General Electric International, 2014)	187
Table A.12 – Minimum Percentage of Total Annual Dispatch from Renewables by Region (General Electric International, 2014).....	188
Table A.13 – Minimum Percentage of Total Regional Dispatch from Renewable Energy Sources (General Electric International, 2014).....	188
Table A.14 – Annual Construction Limits by Unit (in MW per Region).....	189
Table A.15 –Overall Construction Limits by Unit	190

List of Figures

Figure 1.1 – Net Electricity Generation in the United States from 2001 through 2014 (U.S. Energy Information Administration, 2015)	2
Figure 1.2 – Schematic Diagram of a Typical Electric Power System (U.S.-Canada Power System Outage Task Force, 2004)	5
Figure 1.3 – North American Power Grid (Scientific American, 2016).....	5
Figure 1.4 – NAAQS Ozone Violations in the Eastern U.S. (U.S. Environmental Protection Agency, 2013)	8
Figure 1.5 – Air Quality in the United States (Goffman, 2010)	9
Figure 2.1 – Chronological Load Graph	15
Figure 2.2 – Typical Load Duration Curve.....	16
Figure 2.3 – Typical GEP Problem Formulation	22
Figure 2.4 – Typical GEP Inputs and Outputs	23
Figure 2.5 – Framework for Building Metamodels (Han & Zhang, 2012).....	50
Figure 2.6 – Simple Framework for Simulation-Based Optimization (Han & Zhang, 2012)	51
Figure 3.1 – GEP Simulation-Based Optimization Framework Considering Human Health Externalities.....	60
Figure 3.2 – GEP and COBRA Simulation Output Data Structure with Illustrative Data	62
Figure 3.3 – Schematic Flow Diagram of GEP & COBRA Experiments	76
Figure 3.4 – Qualitative Comparison of Metamodeling Methods	79
Figure 3.5 – Gaussian Semivariogram Example.....	80
Figure 4.1 – Transmission Network of the Northeastern United States	85
Figure 4.2 – Mean Absolute Prediction Error Values from Cross Validation Procedure.	96
Figure 4.3 – Annual Dispatch Summary Dashboard	98
Figure 4.4 – Total Dispatch Percentage Dashboard.....	99
Figure 4.5 – Total Cost Summary (Including Simulated Health Damages)	107

Figure 4.6 – Total Cost Summary (Including Deterministic Health Damages)	108
Figure 4.7 – Market Cost Summary.....	109
Figure 4.8 – Simulated Health Damages Summary	110
Figure 4.9 – Deterministic Health Damages Summary	110
Figure 4.10 – Energy Comparison (U.S. Energy Information Administration, 2015) ...	113
Figure 4.11 – Capacity Shadow Prices	114
Figure 4.12 – Simulation-Based Optimization Iteration Summary	115
Figure 5.1 – Mean Absolute Prediction Error Values from Cross Validation Procedure	120
Figure 5.2 – Annual Dispatch Summary Dashboard	122
Figure 5.3 – Total Dispatch Percentage Dashboard.....	123
Figure 5.4 – Total Cost Summary (Including Simulated Health Damages)	132
Figure 5.5 – Total Cost Summary (Including Deterministic Health Damages)	133
Figure 5.6 – Market Cost Summary.....	134
Figure 5.7 – Simulated Health Damages Summary.....	135
Figure 5.8 – Deterministic Health Damages Summary	135
Figure 5.9 – Energy Comparison (U.S. Energy Information Administration, 2015)	138
Figure 5.10 – Capacity Shadow Prices	139
Figure 5.11 – Simulation-Based Optimization Iteration Summary	140
Figure 5.12 – Total Costs by Region	141
Figure 5.13 – Total Market Costs by Region.....	141
Figure 5.14 – Health & Other Damages by Region.....	142
Figure 5.15 – Energy Shadow Prices by Region	143
Figure 5.16 – Capacity Shadow Prices by Region.....	144
Figure 5.17 – Annual Dispatch Summary – MDDE Region	147
Figure 5.18 – Summary of New Investments – MDDE Region	147

Figure 5.19 – Annual Capacity Summary – MDDE Region	148
Figure 5.20 – Annual Emissions Summary – MDDE Region	148
Figure 5.21 – Annual Dispatch Summary – NE Region.....	149
Figure 5.22 – Summary of New Investments – NE Region	150
Figure 5.23 – Annual Capacity Summary – NE Region.....	150
Figure 5.24 – Annual Emissions Summary – NE Region	151
Figure 5.25 – Annual Dispatch Summary – NJ Region.....	152
Figure 5.26 – Summary of New Investments – NJ Region	152
Figure 5.27 – Annual Capacity Summary – NJ Region.....	153
Figure 5.28 – Annual Emissions Summary – NJ Region	153
Figure 5.29 – Annual Dispatch Summary – NY Region	154
Figure 5.30 – Summary of New Investments – NY Region	155
Figure 5.31 – Annual Capacity Summary – NY Region	155
Figure 5.32 – Annual Emissions Summary – NY Region	156
Figure 5.33 – Annual Dispatch Summary – NYC Region.....	157
Figure 5.34 – Summary of New Investments – NYC Region	157
Figure 5.35 – Annual Capacity Summary – NYC Region.....	158
Figure 5.36 – Annual Emissions Summary – NYC Region	158
Figure 5.37 – Annual Dispatch Summary – RoPJM Region.....	159
Figure 5.38 – Summary of New Investments – RoPJM Region.....	160
Figure 5.39 – Annual Capacity Summary – RoPJM Region	160
Figure 5.40 – Annual Emissions Summary – RoPJM Region.....	161
Figure A.1 – Load Profile by Region (U.S. Energy Information Administration, 2015)	183

1. Introduction

Generation expansion planning (or GEP) is the process of analyzing, evaluating, and recommending which technologies should be added to the power grid in order to satisfy changing demand for electricity. As part of this process, the typical objective-function costs considered are investment (or capital costs), fixed operating and maintenance (O&M) costs, and variable O&M costs (including fuel costs). However, this research presents the first comprehensive approach to solving the generation expansion planning problem that minimizes total system-wide costs including investment costs, fixed and variable O&M costs, fuel costs, the social cost of carbon and methane leakage, and human health externalities associated with air emissions from electricity generation.

Per the U.S. Energy Information Administration, electricity generation has increased in since 2001, as shown in Figure 1.1. Furthermore, uninterrupted access to electricity is critical for various global industries that are vital to economic growth. As a result, decision-makers must decide on the best approach to expand the power grid. Such decisions are made systematically and analyzed in depth to address demand and reliability concerns. Robust mathematical models play an instrumental role in allowing stakeholders to make informed decisions in the expansion planning process.

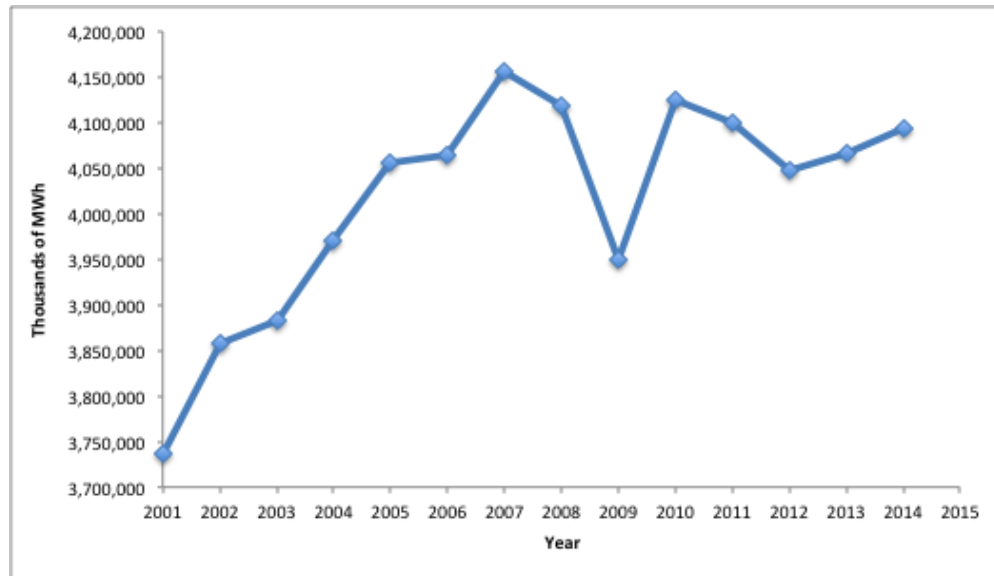


Figure 1.1 – Net Electricity Generation in the United States from 2001 through 2014
(U.S. Energy Information Administration, 2015)

Table 1.1 – Net Electricity Generation by Energy Source (U.S. Energy Information Administration, 2015)

Energy Source	2014 U.S. Net Electricity Generation (GWh)	% of 2014 U.S. Net Generation
Coal	1,581,710	39%
Natural Gas	1,126,609	28%
Nuclear	797,166	19%
Renewables	279,213	7%
Conventional Hydroelectric	259,367	6%
Petroleum Liquids	18,276	0%
All Others	13,461	0%
Other Gases	12,022	0%
Petroleum Coke	11,955	0%

As evidenced by Table 1.1, in 2014, nearly 4,100 million GWh of energy were generated in the United States. Fossil fuels, mainly coal and natural gas, are the main sources of energy for electricity, and account for 67% of electricity generation in the United States. Additionally, per the U.S. Environmental Protection Agency, the electricity sector generates the largest share of greenhouse gas emissions. The most prevalent greenhouse gases are carbon dioxide (CO₂) and methane (CH₄) which are the result of burning fossil fuels (particularly coal and natural gas) (U.S. Environmental Protection Agency, 2016). Greenhouse gases trap heat in the atmosphere, which ultimately leads to increasing temperatures across the globe. This leads to the increased frequency and severity extreme weather events such as hurricanes and heat waves (U.S. Environmental Protection Agency, 2016). Furthermore, burning fossil fuels creates pollutants and by-products that contribute to ground-level ozone, particulates, and acid rain, which have harmful health effects. Exposure to these elements causes thousands of respiratory-related illnesses leading to lost days at school or work on a daily basis and death (U.S. Environmental Protection Agency, 2013). In previous research, generation expansion planning models have attempted to account for these health effects by including emissions limits and renewable portfolio standards constraints in an optimization model. Although these models are effective in obtaining reasonable solutions, they do not fully address the issue of human health externalities absorbed by consumers as a result of pollutants from electricity generation.

In this work, we introduce a simulation-based method, which combines a generation expansion planning model with an EPA screening tool in order to develop an electricity generation expansion plan that minimizes total system-wide costs, including

human health externalities from the electricity sector. Key parameters in the generation expansion planning model are modified based on a structured simulation approach, and the model is thus re-solved multiple times to obtain a diverse sample space of expansion plans as a function of these parameter changes. The changes from a baseline expansion plan are then assessed in the EPA screening tool in order to quantify the human health externalities associated with a given expansion plan. Based on the simulation output, a statistical interpolation method, kriging, is introduced and applied to predict and quantify human health externalities, as a function of power grid expansion decisions. This metamodel is used as a surrogate objective function in a GEP optimization model to determine an expansion plan. We then apply a sampling method to select points in the feasible region that maximize the fit of the metamodel. We continue this iterative procedure until termination criteria have been met.

1.1. Electric Power Systems Overview

The National Academy of Engineering refers to the U.S. electric power grid as the “supreme engineering achievement of the 20th century”, as it serves over 143 million residential, commercial, and industrial customers through over 6 million miles of transmission and distribution lines (Massachusetts Institute of Technology, 2011). The electric power grid is a real-time energy system, meaning that power is generated, transmitted, and delivered at the instant that demand (or load) occurs. The distribution network is the means by which electrical energy is delivered from substations to the consumer. Load is the amount of energy consumption by consumers. The term “grid” refers to generation and transmission. The relationship between the elements of the electric power grid is shown in Figure 1.2.

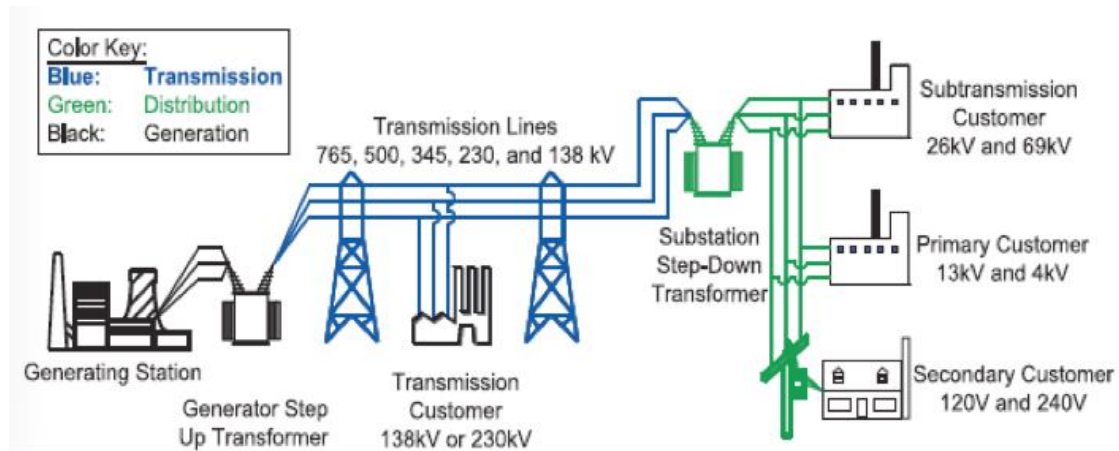


Figure 1.2 – Schematic Diagram of a Typical Electric Power System (U.S.-Canada Power System Outage Task Force, 2004)

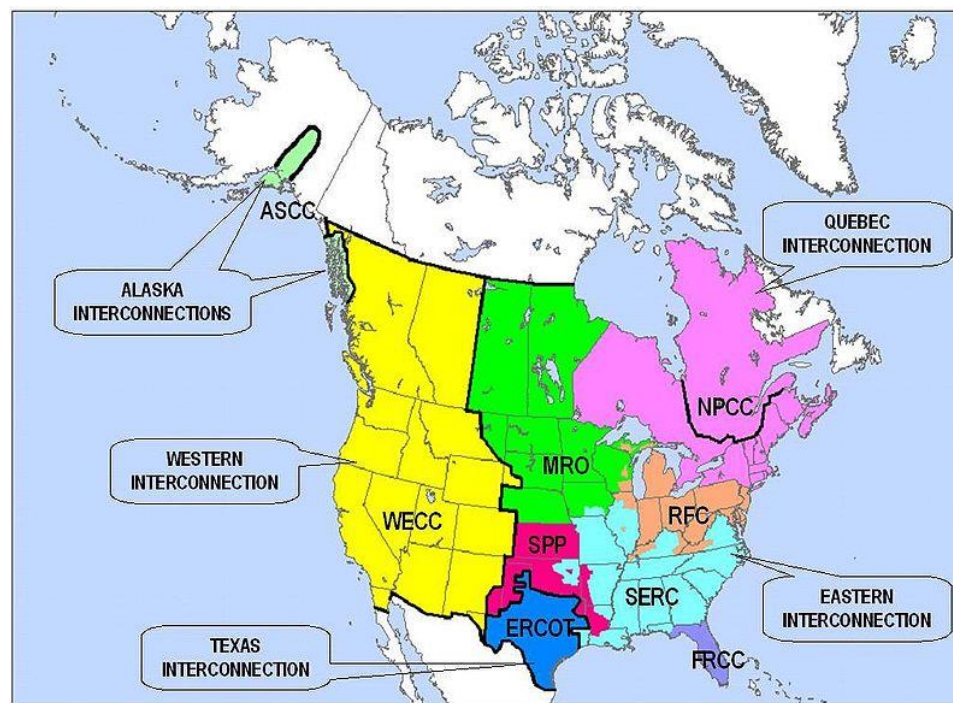


Figure 1.3 – North American Power Grid (Scientific American, 2016)

Figure 1.3 shows a diagram, which breaks down the regions and interconnections within the North American grid. The Eastern grid covers the eastern two-thirds of the U.S. and Canada, the Western grid covers the rest of the U.S. and Mexico, and the Electric Reliability Council of Texas (ERCOT) covers most of Texas. The North

American grid operates on AC lines within each interconnect, and interconnects are joined together by DC links. The North American Electric Reliability Corporation (NERC), a not-for-profit international organization, ensures the reliability of the bulk power system in North America. NERC develops and enforces reliability standards; monitors the bulk power system; assesses future adequacy; audits owners, operators, and users for preparedness and educates and trains industry personnel (North American Electric Reliability Corporation, 2016).

1.2. Motivation of Research

Electricity expansion decisions are informed by solving the generation expansion planning (GEP) problem, which is typically formulated as a least-cost stochastic optimization problem (Tekiner, 2010). The most commonly considered costs in least-cost GEP models are market costs such as investment costs, fixed O&M costs, and variable O&M costs (including fuel costs). With the primary objective being cost minimization, the goal is to ensure that expansion decisions are made economically from the perspective of a utility company (Wood & Wollenberg, 1996). However, traditional least-cost GEP models do not completely include the total societal costs associated with electricity generation. Thus, it is necessary to investigate the human health externalities of electricity generation, as well as the societal health damages of greenhouse gases such as carbon dioxide air emissions and methane leakage, and include these costs in the objective function of the GEP.

Since fossil fuels are well developed, cost-effective, and reliable sources of energy, the resulting expansion plans of traditional least-cost GEP models typically have significant investments in these fuel sources (Wood & Wollenberg, 1996). However, as

mentioned previously, fossil fuel combustion produces harmful by-products and pollutants that ultimately lead to a wide range of health damages. Upon experiencing any of these damages, additional costs are passed onto consumers or to society at large in the form of externalities (National Research Council of the National Academies, 2010). Human health externalities are difficult to quantify explicitly as a function of generation expansion planning decisions. To compensate for this, researchers have extended traditional GEP models to include emissions limits and renewable portfolio standards as model constraints, and have even applied deterministic penalty functions for generating units emitting pollutants. Additionally, to account for human health externalities, researchers have post-optimally assessed the outputs of traditional GEP models in EPA screening models. Furthermore, researchers also employ deterministic estimates of human health externalities of electricity generation to be included in the objective function of GEP models. However, since the electricity sector only makes up a portion all pollutant emissions, and due to the fact that pollutant emissions have a highly non-linear relationship with air quality and thus human health effects, using deterministic estimates yields potentially misleading and inaccurate results. To mitigate the health implications of fossil fuels moving forward, it is necessary to develop a method to quantify these externalities, so that they can be used as part of the expansion planning process. Furthermore, quantifying these health externalities will allow the GEP problem to be extended so that the optimal expansion plan actually minimizes total system-wide costs, inclusive of societal health damages.

Electricity demand has a strong correlation with NO_x emissions from electric generating units, as well as atmospheric ozone concentration and National Ambient Air

Quality Standards (NAAQS) ozone exceedances in the Eastern U.S. This is exemplified in Figure 1.4, which shows the number of violations of either the 8-hour or 1-hour ozone NAAQS. NO_x emissions react with atmospheric oxygen to produce elevated levels of ozone near the earth's surface, which has various harmful health implications (Machol & Rizk, 2013). Ozone exposure and the associated pollutants can irritate the respiratory system, and has been linked to heart attacks, asthma, bronchitis, and even potential risk of death from long-term exposure (Delucchi, Murphy, & McCubbin, 2002).

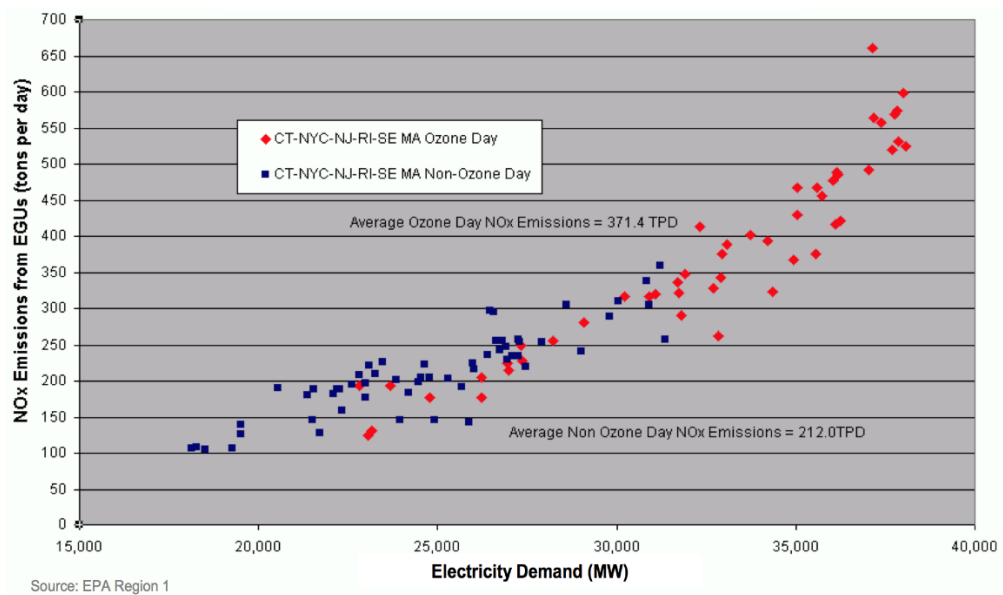


Figure 1.4 – NAAQS Ozone Violations in the Eastern U.S. (U.S. Environmental Protection Agency, 2013)

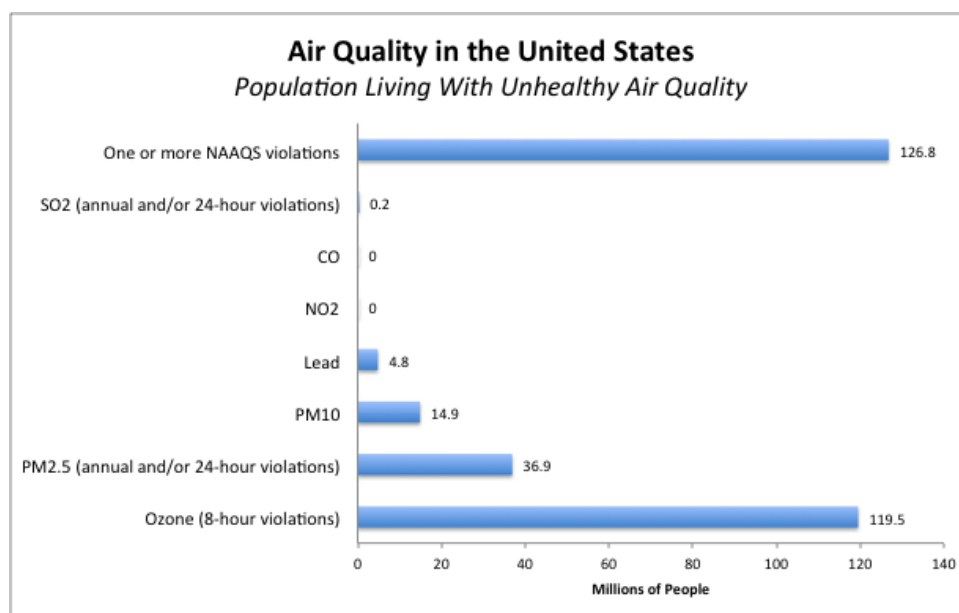


Figure 1.5 – Air Quality in the United States (Goffman, 2010)

The increasing number of Eastern US ozone violations shown in Figure 1.4 is indicative of the poor air quality in the US. Based on the Figure 1.5, nearly 130 million people in the US are living in areas with at least one or more NAAQS violation per year and therefore are at risk of experience health damages.

Mitigating human health externalities has large, quantifiable economic benefits. For example, the EPA's Clean Power Plan, proposed in 2014, presents a plan to reduce carbon emissions from power plants by 30 percent by 2030 from 2005 levels. Per their analyses, since power plants are the largest concentrated source of carbon dioxide emissions in the United States (accounting for approximately one-third of all domestic greenhouse gas emissions), this yields indirect climate and health benefits that outweigh the cost of the plan. The EPA projects that the total costs of the EPA's Clean Power Plan are between \$7.3 billion and \$8.8 billion in 2030, whereas the projected climate and health benefits range from \$55 billion to \$93 billion per year in 2030 (U.S. Environmental Protection Agency, 2016).

1.3. Objectives and Research Contributions

In this research, a simulation-based optimization framework is applied to the GEP problem. The numerical examples and test cases are based on the Northeastern US, which includes New York State, New York City, New England, and the PJM Interconnection. The PJM Interconnection is the regional transmission organization (RTO) that manages wholesale electricity markets in Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia. The scope of this work is to build generation expansion planning models that minimize investment costs, operational costs, variable costs, the social cost of carbon, and human health externalities associated with pollutant emissions from electricity generation. There are two major objectives of this research, which are as follows:

1. Establish a mathematical relationship to predict human health externalities associated with electricity generation and emissions;
2. Develop a generation expansion planning (GEP) model that minimizes investment costs, fixed O&M costs, variable O&M costs (including fuel costs), the social cost of carbon and methane leakage, and human health externalities associated with air emissions from electricity generation.

In order to achieve these objectives, we establish a direct link between electricity use with emissions, the resulting impact on air quality, and ultimately an estimate of human health externalities. From both a problem domain and industrial engineering/operations research perspective, there are 3 major research contributions, which are as follows:

1. Statistically linking generation expansion planning decisions and human health externalities via the utilization of metamodeling;
2. Solving a large-scale, multi-period optimization model over an extended time horizon with a metamodel as a surrogate objective function;
3. Developing a new iterative method to search the feasible region of solutions, resample, and refit the metamodel for human health externalities, to ensure a relationship between decision variables and the metamodel is accurate, and thus, the optimization method yields robust candidate solutions;

The first research contribution stems from the linear linkages of inputs and outputs between a simulated GEP model and an EPA screening tool to estimate human health externalities. In this research, we apply a kriging interpolation model to express the relationship between electricity expansion plans and human health externalities. Kriging is a geospatial estimation method that predicts the values of a random field at unobserved locations based on an interpolated function of observed samples. This metamodel serves as a surrogate objective function and/or constraints in an optimization problem.

The second research contribution arises from the nature the GEP problem. In order to incorporate human health externalities into the total cost objective function, the kriging metamodel will be used as an approximation. This allows us to solve a large-scale, multi-period GEP problem over an extended time horizon with a metamodel as a surrogate objective function for human health externalities. Typically, the incorporation of metamodels into optimization problems is used for smaller scale applications with few decision variables, such as aircraft design problems. However, since a typical GEP

problem has thousands of decision variables, the application of metamodels in such problems is novel.

The third research contribution applies analytical and statistical methods to ensure our metamodel closely mimics the relationship between generation expansion planning decisions and human health externalities. Following our proposed iterative approach allows us to find optimal or near optimal solutions, given that a portion of the objective function is a surrogate curve serving as an approximation of human health externalities. Furthermore, applying such analytical methods in large-scale optimization problems, like the GEP problem, is a contribution to the operations research and industrial engineering fields.

2. Background and Literature Review

This section of the dissertation is divided into four parts. In the first part, key operational definitions and optimization formulations are presented for short-term planning problems for electric power systems. The second part discusses the generation expansion planning (GEP) problem. In this part, GEP problems are broken down into the most common formulations. The third section presents a literature survey of the research related to GEP models including various GEP research extensions such as variations of the objective function and constraints, various solution methods used to solve GEP models, approaches to quantify health damages associated with air pollution from electricity dispatch, and the implications of methane leakage. In the last part of this section, simulation based optimization approaches are discussed, and a literature survey of the research in this field is presented.

2.1. Electrical Power Systems Planning

There are multiple types of generating units that are used to satisfy the demand for electricity. These units can be categorized into three major categories: baseload units, peaking units, and intermediate load units (Wood & Wollenberg, 1996). Baseload units are used to meet most of a region's continuous energy demand, or base load. These units produce energy at a constant rate at relatively lower direct costs in comparison to other available plants in the network. Typically, nuclear, coal, natural gas, combined cycle gas turbines (CCGT), and geothermal are among the generation technologies to meet base load (Wood & Wollenberg, 1996). Peaking units are dispatched only when demand is high, or at its peak. Since these units are dispatched once less economical options have been exhausted, these are more expensive than baseload units. Oil, diesel, natural gas,

and pumped storage are among the technologies used to satisfy peak demand. Intermediate units operate between the extremes of base and peak load. Oil and combined-cycle gas turbine (CCGT) units are examples of technologies used to satisfy intermediate load. Renewable units, such as wind, biomass, and solar units are also used to satisfy load within the network. While these units are able to generate electricity without producing air emissions at the levels of fossil fuel powered units, the intermittency of these units, particularly wind and solar generation, hinders the availability of these units, and thus impacts the ability to balance supply and demand instantaneously (Marneris, Biskas, & Bakirtzis, 2016)

Generating unit efficiency is a major driver of the tradeoff between low operating and capital costs versus increased operational flexibility (Wood & Wollenberg, 1996). Operating costs and flexibility are directly related to the startup time and the ramp rate. The startup time is the amount of time it takes to warm up a unit so that it is operational. The ramp rate is the rate at which a unit can change its output. Furthermore, startup time and ramp rate are impacted by the heat rate (usually measured in BTU/kWh), which is an inverse measure of efficiency of a generating unit. Units with higher heat rates have longer startup times and small ramp rates, thus making them less efficient. For example, coal units generally have heat rates between 9,000 and 11,000 BTU/kWh, while natural gas units have between 6,000 and 8,000 BTU/kWh (Massachusetts Institute of Technology, 2011). Peaking units have heat rates of over 12,000 BTU/kWh, thus making these units the least efficient, and the most expensive to operate on a daily basis (Massachusetts Institute of Technology, 2011).

Demand for electricity is described by a load duration curve (LDC). In these diagrams, the chronological sequence of load is sorted in descending order. Each location on the horizontal axis corresponds to the amount of time in the given time period where the electricity demand is greater than or equal to the corresponding load value on the vertical axis. The area under the LDC represents the total electricity power demand in the period. In general, peak demand in each day or peak demand in each hour are used to construct the LDC. Figure 2.1 shows the chronological load as a function of time, and Figure 2.2 illustrates a typical LDC. L_{max} and L_{min} are the respective maximum and minimum load in the given time period involved, and T is the total number of time units in the given time period.

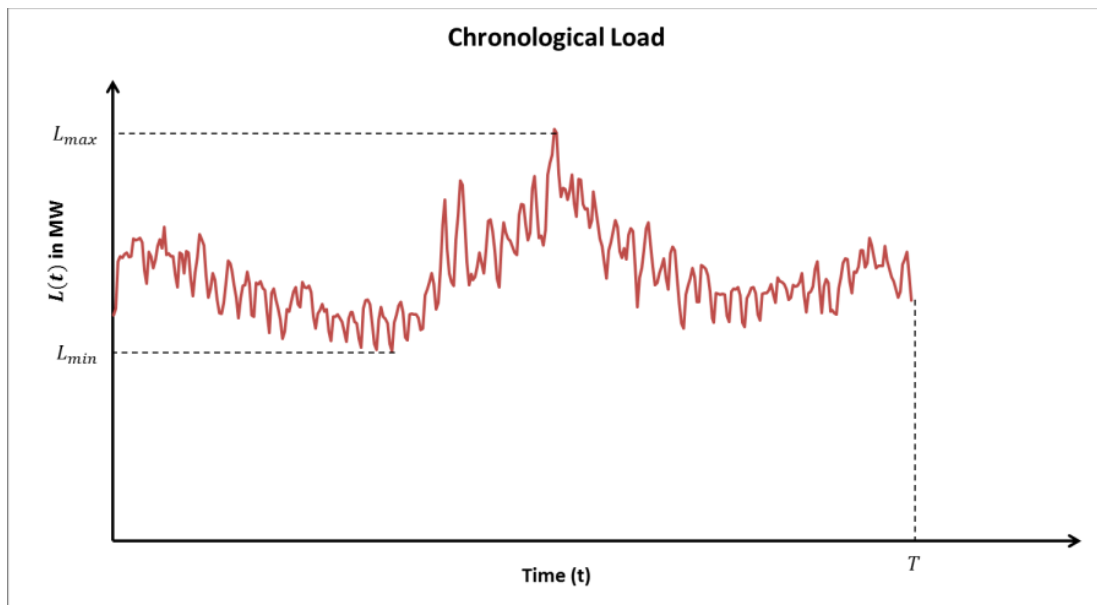


Figure 2.1 – Chronological Load Graph

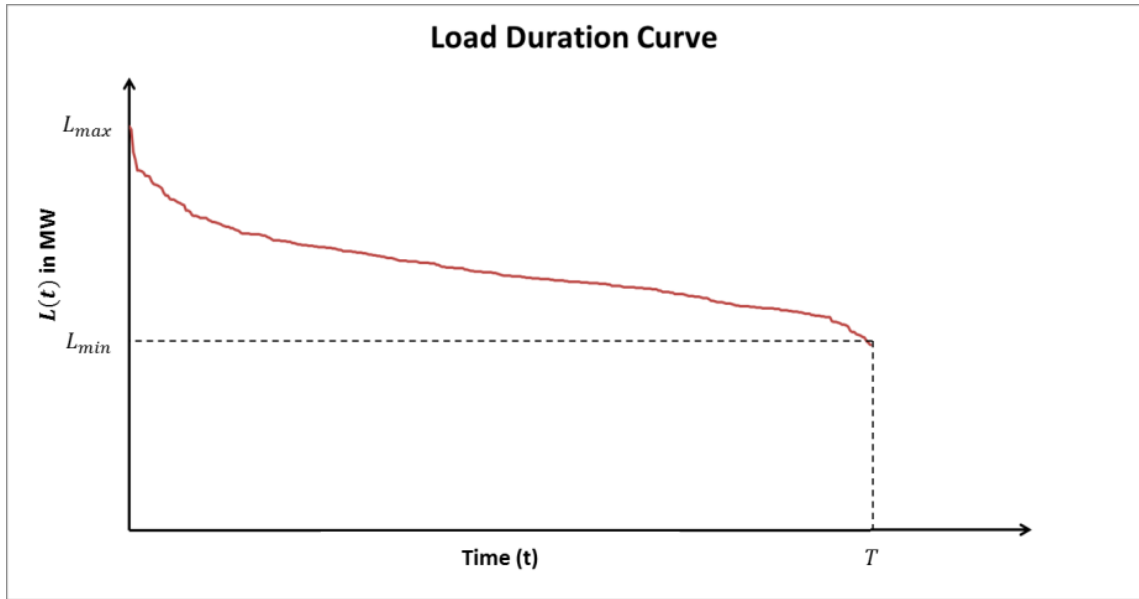


Figure 2.2 – Typical Load Duration Curve

This LDC is used in power systems planning and optimization models for long-term expansion planning and network reliability evaluation. It can also be used to represent capacity utilization requirements for each increment of load. The horizontal axis can be interpreted as a capacity utilization rate, and the vertical axis can be interpreted as the capacity requirement in MW. Using this representation, a vertical section of the LDC can be extracted from the curve to compute the capacity utilization requirements for a specific load increment (Wood & Wollenberg, 1996).

2.1.1 Economic Dispatch

The economic dispatch model is a short-term planning model, which determines the optimal output of a number of electricity generation facilities, to meet the system load, at the lowest possible cost, subject to system constraints. In such markets, dispatching cost functions are approximated by polynomial functions of power generation in each hour, as evidenced by Equation 2.1.

$$c_i = \alpha_i + \beta_i p_i + \gamma_i p_i^2 \quad (2.1)$$

In Equation 2.1, c_i represents the cost of generating unit i , α_i , β_i , and γ_i are polynomial constants, and p_i is the power generated by unit i . In a network with m generating units, this be formulated as an optimization problem. The formulation is given as follows (Wood & Wollenberg, 1996).

Minimize

$$c_T = \sum_{i=1}^m c_i = \sum_{i=1}^m \alpha_i + \beta_i p_i + \gamma_i p_i^2 \quad (2.2)$$

Subject to

$$\sum_{i=1}^m p_i = p_D \quad (2.3)$$

$$p_i > 0, \forall i \quad (2.4)$$

In the above optimization problem, c_i and p_i are defined in Equation 2.1, c_T is the total cost of all generators, m is the total number of generators in the system, and p_D is the total demand on the system. Equation 2.3 ensures that the total amount of generation from all units in the system meets total system load. The Lagrangian Relaxation Method can solve simple economic dispatching problems, like the one above, to optimality.

There are many variations of this optimization problem, which are based on simple extensions. The most common extensions include linear approximations of cost curves, indexed decision variables and model parameters for hourly dispatch, maximum and minimum generation limits, and forecasted transmission losses. These extensions are presented in the formulation below (Wood & Wollenberg, 1996).

Minimize

$$c_T = \sum_{j=1}^{24} \sum_{i=1}^m c_{ij} = \sum_{j=1}^{24} \sum_{i=1}^m \alpha_{ij} + \beta_{ij} p_{ij} \quad (2.5)$$

Subject to

$$\sum_{i=1}^m p_{ij} = p_{Dj} + p_{Lj}, \forall j \quad (2.6)$$

$$p_{ij} \leq p_{max,ij}, \forall i, \forall j \quad (2.7)$$

$$p_{ij} \geq p_{min,ij}, \forall i, \forall j \quad (2.8)$$

The objective function shown in Equation 2.5 of the extended formulation now has an additional index j , which indicates the hour. The cost curves submitted by the generators, denoted by c_{ij} , are now linear functions of the power, denoted by p_{ij} . Equation 2.6 is analogous to Equation 2.3 in that it ensures total demand in the system is met. However, Equation 2.6 includes a loss term, p_{Lj} , to ensure that the total amount of generation accounts for the demand plus losses. Equations 2.7 and 2.8 specify maximum and minimum values for each of the generating units in a given hour. Simplex-derived algorithms can solve this problem since the constraints and the objective function are linear.

2.1.2 Unit Commitment

Given a load profile and a set of available generating units in the network, the unit commitment model determines which units should be started-up and shutdown in each period at least cost, in order to satisfy the load on the system. In comparison to the economic dispatch model, the unit commitment model only considers only start-up and shutdown decisions, which are binary decision variables. Dispatching decisions are not determined these models. The formulation for the unit commitment model is given as follows (Wood & Wollenberg, 1996).

Minimize

$$z = \sum_{j=1}^{24} \sum_{i=1}^m (s_{ij}x_{ij} + q_{ij}u_{ij}) \quad (2.9)$$

Subject to

$$\sum_{i=1}^m p_{ij,max}x_{ij} = p_{Dj} + p_{Lj}, \forall j \quad (2.10)$$

$$u_{ij} + x_{ij} = 1, \forall i, \forall j \quad (2.11)$$

$$x_{ij} \in \{0,1\}, \forall i, \forall j \quad (2.12)$$

$$u_{ij} \in \{0,1\}, \forall i, \forall j \quad (2.13)$$

The objective function of this model, described in Equation 2.9, is the total start-up and shutdown costs for the system. The start-up costs and shutdown costs are given by s_{ij} and q_{ij} , respectively. Start-up and shutdown decisions are given by x_{ij} and u_{ij} , respectively. Equation 2.10 is analogous to Equations 2.3 and 2.6 from the economic dispatch models. This equation states that total supply, must be equivalent to total demand plus losses. The main difference in this equation is the inclusion of maximum generation. The sum of the maximum generation values for the committed units must satisfy demand plus losses. Equation 2.11 states that a unit cannot be started-up and shutdown in the same time period. Equations 2.12 and 2.13 ensure that the start-up and shutdown decisions are binary.

Since the given formulation is an integer program, there are various metaheuristics that can be used to solve this problem to optimality (or near optimality). The model can be further extended to multiple areas. Additional constraints can be added as well, including transmission limits, emissions caps, and spinning reserve requirements (Wood & Wollenberg, 1996). Furthermore, stochasticity can be incorporated into these problems by considering uncertainty in demand, costs, and other system parameters.

2.1.3 Unit Commitment Dispatch

The unit commitment dispatch model is a combination of the economic dispatch model and the unit commitment model. This model is a mixed integer nonlinear optimization problem that minimizes start-up costs, shutdown costs, and variable costs

from energy generation. The decision variables in this problem include dispatching decisions as well as start-up and shutdown decisions. The formulation is given as follows (Wood & Wollenberg, 1996).

Minimize

$$z = \sum_{j=1}^{24} \sum_{i=1}^m (c_{ij}(p_{ij}) + s_{ij}x_{ij} + q_{ij}u_{ij}) \quad (2.14)$$

Subject to

$$\sum_{i=1}^{n(j)} x_{ij}p_{ij} + \sum_{i=1}^{a(j)} (1 - u_{ij})p_{ij} = p_{Dj} + p_{Lj}, \forall j \quad (2.15)$$

$$p_{ij} \leq \varphi_i p_{i,j-1} + \delta_i, \forall i, \forall j \quad (2.16)$$

$$u_{ij} + x_{ij} = 1, \forall i, \forall j \quad (2.17)$$

$$p_{ij} \leq p_{max,ij}, \forall i, \forall j \quad (2.18)$$

$$p_{ij} \geq p_{min,ij}, \forall i, \forall j \quad (2.19)$$

$$x_{ij} \in \{0,1\}, \forall i, \forall j \quad (2.20)$$

$$u_{ij} \in \{0,1\}, \forall i, \forall j \quad (2.21)$$

The objective function in this model, given by Equation 2.14, is the total costs including start-up costs, shutdown costs, and variable costs. Equation 2.15 is the supply and demand constraint, which states that the total generation from all units started-up in period j and all units operating in period j must satisfy demand plus losses in that period. Equation 2.16 is the ramp-up constraint, which states that the generation by a given unit in a given period cannot exceed some function of the generation by that unit in the previous period. Equation 2.17 states that a unit cannot be started-up and shutdown in the same time period. Equations 2.18 and 2.19 specify maximum and minimum values for each of the generating units in a given hour. Equations 2.20 and 2.21 ensure that the start-up and shutdown decisions are binary.

2.2. Generation Expansion Planning Overview

The generation expansion planning (GEP) problem is used to determine the optimal selection of power generation technologies to be added to the existing electrical grid, as well as determining when and where these generating units should be built in order to satisfy the increasing energy demand over a specified planning horizon. The most commonly studied GEP problems are typically cost minimizing problems (Hemmati, Hooshmand, & Khodabakhshian, 2013). However, since GEP is a complicated, multi-faceted problem, solutions that only consider cost are not necessarily indicative of the expansion planning decisions that are made. Because of this, researchers have studied various mathematical formulations of GEP models to enhance decision making capabilities (Hemmati, Hooshmand, & Khodabakhshian, 2013).

Furthermore, expansion plans rely on assumptions about the future. These assumptions have an impact on forecasted energy demand, fuel prices, and other system-related parameters. Since the GEP problem is formulated based on these assumptions, there are a considerable amount of uncertainties in the system, which has an impact on the resulting expansion plan. Thus, researchers have investigated the stochastic nature of GEP problems.

In addition to considering different formulations of the GEP problem, researchers have also investigated the application of different solution techniques to finding the optimal expansion plan. These solution methods include mathematical programming, meta-heuristic search algorithms, machine learning, and dynamic programming (Hemmati, Hooshmand, & Khodabakhshian, 2013). Although many variations of the generation expansion planning problem have been studied in the academic community, a

GEP model that determines the optimal expansion plan that minimizes the human health externalities of pollutants from electricity generation using simulation-based methods has not been studied.

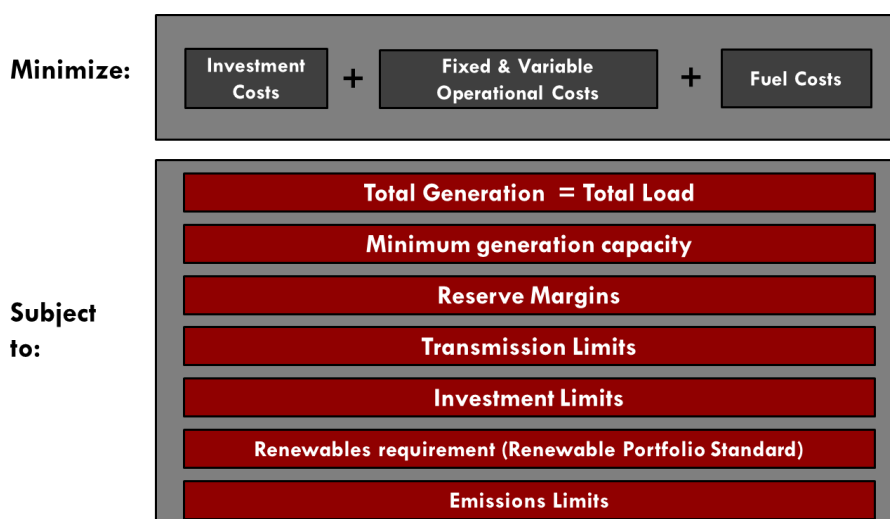


Figure 2.3 – Typical GEP Problem Formulation

Figure 2.3 outlines a typical GEP problem formulation. The objective function often includes market costs such as investment costs, fixed and variable operating & maintenance costs, and fuel costs. Investment costs are the capital costs required to build a new unit in a particular location. In linear models, investment and maintenance costs are given in terms of \$/MW (Hemmati, Hooshmand, & Khodabakhshian, 2013). However, for integer models with binary investment constraints, investment costs are given in \$, as the investment decisions are associated with fixed capacities. Similarly, in linear models, fixed maintenance costs are considered in terms of \$/MW since they are assumed to be functions of capacity, inclusive of new investments. These costs are incurred by all plants in the network on an annual basis. Variable operating & maintenance costs, as well as fuel costs, are given in \$/MWh as they are the cost per unit generation of electricity (Hemmati, Hooshmand, & Khodabakhshian, 2013).

As evidenced by Figure 2.3, all GEP problems have a constraint that states that total generation must be equivalent to the total load. Often times generating units must operate at a minimum level, which can be considered as a constraint in the model as well. Reserve margins act as reliability constraints in GEP models, to ensure that there is sufficient capacity in the network to satisfy peak demand levels. The transmission limit constraint imposes an upper limit on the amount of electricity that can be sent from one bus to another bus within the network. Investment limits can be specified in terms of MW of capacity invested in a given unit or in terms of monetary values, which would function as a budget constraint. Renewables requirements, also referred to renewable portfolio standards, can also be implemented in the GEP as a constraint. This constraint specifies a minimum amount of generation from renewable energy sources every year. Another common constraint in GEP problems is emissions limits (or maximums).

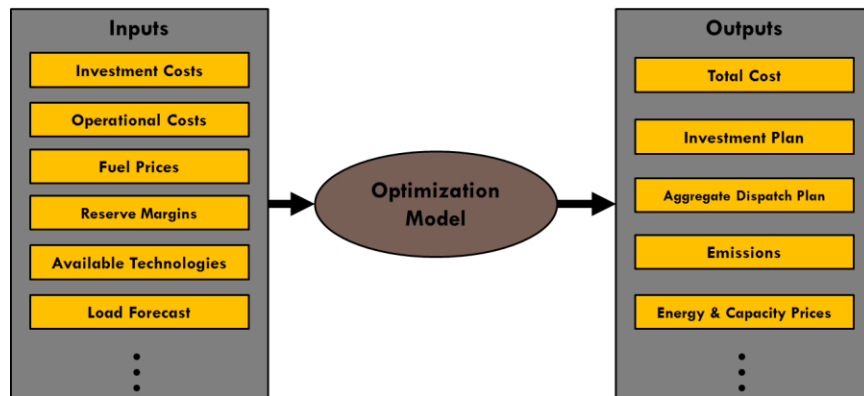


Figure 2.4 – Typical GEP Inputs and Outputs

Figure 2.4 provides the key inputs and outputs to a typical GEP model. Market costs, such as investment costs, fixed and variable operating & maintenance costs, and fuel prices are used in the objective function of the optimization problem. Generation characteristics include heat rates and emissions rates, as well as existing generating unit

capacities, which are used in formulating the constraints in the optimization problem. The available technologies are a pool of generating units that are available for investment and dispatching. The load forecast is used in the supply and demand constraint.

The main outputs are the total cost of the expansion plan at optimality, the investment plan, and the aggregate dispatch plan. Dispatching decisions are aggregated in GEP models, because the hour-by-hour, unit-by-unit level of granularity is not necessary for long-term planning. In addition to dispatching decisions, the resulting emissions are outputs as well. Additionally, energy and capacity shadow prices are outputs to GEP problems, provided that the associated constraints in the GEP model do not have integer-constrained decision variables. In general, a typical expansion plan will include large centralized generating units, including coal, nuclear, and natural gas. This is due to the fact that most GEP models minimize market costs only.

2.3. Literature Survey

In this section, we provide an overview of the current research as it relates to generation expansion planning and human health externalities and damages in the context of this research. Specifically, we discuss various GEP model formulations that aim to reduce human health externalities associated with air pollution from the electricity sector, current research methods to assess health damages associated with air pollution from electricity generation, various GEP solution approaches and metaheuristics, as well approaches on how uncertainty or stochasticity is treated in the context of generation expansion planning models.

2.3.1. Health and Environmental Implications of Electricity Dispatch

As mentioned previously, a large amount of our electricity is generated from fossil fuels – mainly coal and natural gas. Fossil fuels are burned to produce steam, which is then used to power a steam turbine that drives an electrical generator. From the perspective of generating companies, fossil fuels are the least expensive options to satisfy demand, since their market costs are often less expensive than alternative sources of energy. However, burning fossil fuels produces emissions such as CO₂, SO_x, NO₂, and particulate matter, which are potential health hazards. These emissions contribute to smog, which causes heart conditions and chronic lung conditions such as asthma, bronchitis, and emphysema (National Research Council of the National Academies, 2010). Moreover, smog exposure can cause inflammation in breathing passages, decreased lung capacity, and shortness of breath (Union of Concerned Scientists, 2013). Smog can form in most climates where there is a large amount of air pollution, and it tends to be worse when the weather is warmer and sunnier when electricity demand is highest (Union of Concerned Scientists, 2013).

Fossil fuel emissions also react with atmospheric water molecules to produce acid rain. Acid rain refers to various forms of acidic precipitation and dry acidic deposition. It arises from the combustion of fossil fuels that produce sulfur dioxide and nitric oxides, which is converted to sulfuric acid and nitric acid upon reacting with atmospheric water (Union of Concerned Scientists, 2013). While acid rain does not have any direct human health effects, fine acidic particulates in air contribute to heart and lung conditions, which results in health externalities or costs absorbed by consumers (Union of Concerned Scientists, 2013).

Nuclear energy, which is another major energy source for electricity, also has negative human health effects that can be quantified as externalities. One of the major sources of nuclear health risks arises from radiation. Nuclear power plants produce radioactive materials that may come into contact with people during routine operations, accidents, transportation, and waste management (National Research Council of the National Academies, 2010). These radioactive materials are subatomic particles that can penetrate the human body, and can potentially damage cells. This can ultimately lead to cancerous growths and genetic diseases (National Research Council of the National Academies, 2010).

During peak periods of electricity demand, we rely on the most economical generating units available. As mentioned previously, during such periods, the cheapest options are exhausted, and thus more expensive fossil fuels with higher emissions rates are dispatched (Farkas, et al., 2015). During these periods, since we consume more energy from sources that produce emissions that are potential health hazards, and thus, the risk of the associated negative health implications is elevated.

Per the National Research Council of National Academies, NO_x compounds and volatile organic compounds (VOCs) can result in particulate matter and ozone. Both particulate matter and ozone, can lead to serious cardiovascular and respiratory issues, birth defects, as well as premature death (National Research Council of the National Academies, 2010). Quantifying the human health externalities of such pollutants allows decision makers to fully understand the impact of electricity generation from fossil fuels on human health. However, these costs absorbed by consumers are not able to be quantified effectively through deterministic methods. This is mainly because human

health externalities have a highly non-linear relationship with electricity generation, emissions, and location, which cannot be expressed in a closed-form. In order to capture this relationship appropriately, rigorous analytical methods are necessary for approximation purposes. In the context of generation expansion planning models, due to the computational rigor associated with determining the location-based health effects of electricity generation, simple approximations, such as objective function penalties for emissions or deterministic cost multipliers, yields inaccurate results.

Greenhouse gases (often abbreviated as GHGs) are gases in atmosphere that absorb and emit radiation within the thermal infrared range (U.S. Environmental Protection Agency, 2016). This process is the fundamental cause of the greenhouse effect, which refers to the exchange of incoming and outgoing radiation that increases global temperatures, leading to climate change (U.S. Environmental Protection Agency, 2016). According to the EPA, the most prominent greenhouse gases in are CO₂, which accounts for approximately 81% of the total greenhouse gas emissions in the United States as of 2014, and methane (CH₄), accounting for roughly 11% of U.S. greenhouse gas emissions in 2014 (U.S. Environmental Protection Agency, 2016). Furthermore, electricity generation accounted for nearly 30% of all U.S. greenhouse gas emissions in 2014 – approximately two-thirds of which are from burning fossil fuels (U.S. Environmental Protection Agency, 2016).

The health implications of greenhouse gas emissions are a result of increased global temperatures, which leads to changes in precipitation, increased frequencies and intensities of extreme weather events, as well as rising sea levels (U.S. Environmental Protection Agency, 2016). These consequences impact human health by affecting our

sources of food and water, air quality, and exposure to extreme weather events (U.S. Environmental Protection Agency, 2016).

In order to quantify the health damages associated with greenhouse gas emissions, the EPA targets the most prominent of the greenhouse gas, CO₂, and conducted a comprehensive study to quantify the estimated economic health damages associated with its air emissions (U.S. Environmental Protection Agency, 2016). These health damages are referred to as the social costs of carbon, and are represented in terms of \$/metric ton of CO₂ (U.S. Environmental Protection Agency, 2016).

Methane (CH₄), is the second most prevalent greenhouse gas emitted in the United States, which is mainly due to leakage from the production, processing, storage, transmission, and distribution of natural gas, which occurs at a rate between 1% and 8% per MWh of electricity generated (U.S. Environmental Protection Agency, 2016), (Brandt, et al., 2014). Not only does methane contribute to the greenhouse effect, which leads to the human health effects mentioned previously, according to the Intergovernmental Panel on Climate Change (IPCC), the 20 year global warming potential (or GWP) of methane is approximately 72 times that of CO₂ (Intergovernmental Panel on Climate Change, 2016). To elaborate, the GWP is a measure of how much heat a greenhouse gas traps in the atmosphere, relative to CO₂. Aside from the global warming potential of methane and its approximated leakage rate, the associated societal health damages have not been studied in sufficient detail.

Table 2.1 – Related Research Summary

Authors	Generation Expansion Planning	Unit Commitment	Market Costs	Health Damages					Climate Change Impacts	Emissions Limits	Renewable Portfolio Standards	Response Surfaces
				Deterministic Cost Multipliers (or Penalties)	Predictive Models	Included Within Optimization Framework	Post-Optimal Assessment	Application of EPA Models				
Schenk & Chan – 1981	✓	-	✓	✓	-	-	-	-	-	✓	✓	-
Nordlund <i>et al.</i> – 1987	✓	-	✓	-	-	-	-	-	-	✓	✓	-
Rowe <i>et al.</i> – 1996	-	-	-	-	✓	-	-	✓	-	-	-	-
Farghal & Aziz – 1998	✓	-	✓	-	-	-	-	-	-	✓	✓	-
Thanh & Lefevre – 2000	-	-	-	-	✓	-	-	-	-	-	-	-
Kenfack <i>et al.</i> – 2001	✓	-	✓	-	-	-	-	-	-	✓	✓	✓
Karaki <i>et al.</i> – 2002	✓	-	✓	✓	✓	✓	-	-	-	-	-	-
Burtaw <i>et al.</i> - 2003	-	-	-	-	✓	-	-	✓	-	-	-	-
Chen & Hobbs – 2005	✓	-	✓	✓	-	-	-	-	-	-	✓	-
Nualhong <i>et al.</i> – 2005	✓	-	✓	✓	-	-	-	-	-	-	✓	-
Alnatheer – 2006	✓	-	✓	✓	-	-	-	-	-	-	-	-
Alves & Uturbey – 2010	✓	-	✓	✓	-	-	-	-	-	-	-	-
Becker <i>et al.</i> – 2011	✓	-	✓	✓	-	-	-	-	-	-	-	-
Selcuklu <i>et al.</i> – 2013	✓	-	✓	-	-	-	-	-	✓	-	-	-
Farkas <i>et al.</i> – 2015	-	✓	-	-	-	-	✓	✓	✓	-	-	-
Kerl <i>et al.</i> – 2015	-	✓	✓	-	✓	✓	-	-	-	-	-	-

Table 2.1 summarizes the research that is closely related to our framework to solve the generation expansion planning problem considering human health externalities in the objective function by simulation-based optimization. Per the literature presented in this section, the most popular approach to assessing and mitigating the human health externalities associated with electricity dispatch is by applying deterministic cost estimation methods. As it relates to power systems planning and optimization models, other popular methods include implementing renewables requirements and emissions limits as constraints, and using multiple objective methods.

Kerl *et al.* introduce the Air Pollutant Optimization Model (APOM), which solves the unit commitment model in the state of Georgia with the assistance of a reduced form air quality model (Kerl, et al., 2015). The objective function of this model is to minimize electricity production costs and monetized health impact cost estimates. Similar to our research, these researchers recognize that health impact costs have a highly non-linear relationship with electricity generation and the associated air emissions, thus making these costs difficult to quantify through deterministic methods. The pollutants under investigation in this research are ozone, PM_{2.5}, SO₂ and NO_x. To estimate their health impact costs, a baseline unit commitment model along with the inputs and outputs of two EPA models are synchronized linearly to develop response surfaces, or regression curves, that predict health impact costs as a function of decision variables in the unit commitment model. The first EPA model, the Community Multiscale Air Quality model (CMAQ), geospatially and temporally models air pollutant concentrations in the atmosphere. The second EPA model, the Environmental Benefits Mapping and Analysis Program (BenMAP), determines the economic value of the subsequent air pollution-related

illnesses and deaths. To assess the stochasticity of health impact costs, various sensitivity analyses are executed by varying key parameters in the unit commitment model, and assessing the results post-optimally in CMAQ and BenMAP.

In order to address health and environmental concerns in the context of generation expansion planning, the most popular approach is to consider these costs as deterministic penalties in the objective function. In the single objective case, Alves and Uturbey applied a GEP model to study expansion planning decisions in Brazil. However, in efforts to address environmental concerns, the researchers included aggregate damage estimates of environmental impact costs in the objective function along with the electricity generation costs (Alves & Uturbey, 2010). Nguyen examines the impacts of including external costs such as environmental and health damages from power production on power generation expansion planning in Vietnam. They consider these costs as deterministic penalties in the objective function of the model (Nguyen, 2008). Using a simple linear programming model over a 20-year period to 2025, the study shows that there are substantial changes in the resulting expansion plan in favor of renewable energy technologies and other low emitting technologies. These changes lead to a reduction in fossil fuel requirements, and consequently, a reduction of CO₂, NO_x, SO₂, and PM emissions, which could be expected to also reduce the associated environmental and human health impacts. Similar research was done by Alnatheer in Saudi Arabia's electricity sector; however, this work considered both generation and transmission expansion decisions, as well as the land and water impacts (Alnatheer, 2006).

In the multiple objective case, Karaki *et al.* use dynamic programming to solve a weighted, bi-objective GEP problem, where the objectives are to minimize electricity

generation costs and minimize deterministic environmental impact costs (Karaki, Chaaban, Al-Nakhl, & Tarhini, 2002). They utilize a method to simulate the cost of dispatching electricity, which is called the probabilistic production costing (PPC) simulation. Additionally, a 2-state Markov model is used to model generating unit availability. Because of the 2-state Markov model, the authors use dynamic programming to solve the GEP. Becker *et al.* also apply multiple objective optimization methods to solve the GEP problem by minimizing the competing objectives of electricity generation costs and minimizing emissions costs, which are derived deterministically from literature (Becker, Soloveitchik, & Olshansky, 2011).

A key extension of generation expansion planning considering human health and environmental implications is the consideration of climate change. Selcuklu *et al.* solve a multiple objective GEP problem with one objective to minimize investment, maintenance, and variable costs, and the other objective to minimize air pollutant emissions, which would consequently minimize adverse human health effects (Selcuklu, Coit, Felder, & Rodgers, 2013). In addition to the multiple objective aspect of this work, to understand the uncertainty associated with climate change assumptions, this research introduces a Pareto Uncertainty Index (PUI), which allows for the ranking of solutions on the pareto front based on dominance.

One of the simplest ways to address environmental concerns, and consequently human health effects, in the context of generation expansion planning is the inclusion of renewables requirements and emissions limits as constraints in the model. Schenk and Chan leveraged renewables requirements in formulating the GEP problem as a least-cost, mixed integer linear program. The goal of this work was to study the impact of

increasing wind dispatch on the optimal expansion plans. They consider the typical energy balance, and reliability constraints. However, they include a minimum wind dispatching constraint as well (Schenk & Chan, 1981). They probabilistically simulate the load duration curve, and solve the GEP over a 19 year time horizon. Farghal and Aziz extended this research by limiting the available generation technologies for expansion to renewable sources, ultimately leading to significant emissions reductions, but increases in market costs (Farghal & Abdel Aziz, 1988).

Nualhong *et al.* consider CO₂ emissions by considering a carbon tax charge for excessive CO₂ emissions, but also include a requirement for biomass to be included in the expansion plan (Nualhong, Chusanapiputt, Jantarang, & Pungprasert, 2005). They solve the GEP by using a tabu search. Chen *et al.* extended this work to not only consider expansion decisions, but also the trading of CO₂ permits (Chen & Hobbs, 2005). Essentially the decision variables in this problem are the typical investment and dispatching decisions. However, they also include the ability to buy and sell CO₂ permits in a regulated environment. Thus, this work studies how expansion planning is viewed in a low-carbon economy.

Nordlund *et al.* solved the GEP in systems where there are large shares of hydro power (Nordlund, Sjelvgren, Pereira, & Bubenko, 1987). This GEP model is a least-cost capacity expansion model that determines the optimal expansion plan subject to water balance, minimum and maximum water outflow, minimum and maximum water storage, and capacity limits. Additionally, the load duration curve is simulated from a normal distribution. The problem is formulated as a linear program. Similar work was done by Kenfack *et al.* for hydro dominated systems. However, the solution approach

decomposed the problem via the use of Benders Decomposition (Kenfack, Guinet, & Ngundam, 2001).

In one of the first attempts to quantify the human health externalities from the electricity sector using advanced statistical methods, Rowe *et al.* assessed the critical factors in computing externalities for electricity generation sources. They utilize the New York State Environmental Externalities Cost Study and computerized externality model (EXMOD) to simulate and develop response surfaces to approximate the negative health externalities of electricity generation by varying 15 different factors, including the selection of generation type, location, and operating characteristics (Rowe, Lang, & Chestnut, 1996). Additionally, they considered various air emissions and economic valuation scenarios in their analysis. The output of this work was a multivariate probabilistic distribution of the externalities as a function of the aforementioned key factors. The most significant factors are the selection and application of air dispersion models, selection of air pollution thresholds for health impacts, reduced life span risks associated with ozone exposure and with long-term exposure to particulate matter, values for CO₂ damages, and the value to be applied to increased risks of reduced life span for individuals age 65 or older.

Thanh and Lefevre apply the impact pathway approach (IPA), which links the origin of the environmental burden with the ultimate impact on human health, to estimate health impacts and damage costs of sulfur dioxide and particulate matter emissions from four power units using different fuels (lignite, oil, natural gas, and coal) at four locations in Thailand. Using response surfaces, the results show that damage costs related to health effects of electricity generation in Thailand are relatively small, but not negligible,

ranging from 0.006 U.S. cent to 0.05 U.S. cent per kilowatt-hour (in 1995 dollars) (Thanh & Lefevre, 2000). Additionally, these damage costs to the public health depend strongly on power plant location, which implies that the assessment of adverse health impacts is very important for technology choice and siting of new power plants.

Burtaw *et al.* take a different but related route and quantify the economic benefits of reduced air pollution in the US as a result of greenhouse gas emissions mitigation policies in the electricity sector by using historical emissions data and an EPA screening tool to generate response surfaces of human health externalities. Using a detailed hourly day-ahead electricity market model (Haiku) linked to an integrated assessment framework to value changes in human health, they found that a tax of \$25 per metric ton of carbon emissions would yield NO_x-related health benefits of about \$8 per metric ton of carbon reduced in the year 2010 (1997 dollars) (Burtaw, et al., 2003).

In addition to analyzing the health externalities of electricity sector emissions, other researchers have worked to understand the impact of air quality on human health in a more general sense. This is particularly important due to the fact that air quality is not only dependent upon emissions from the electricity sector. Emissions from the transportation and agricultural sectors, amongst others also contribute to air quality conditions. To investigate this further, Voorhees *et al.*, conducted a sensitivity analysis in an industry standard air quality model, and analyzed the associated human health externalities. The tool used, BenMAP (Environmental Benefits Mapping and Analysis Program), is an EPA economic model used for estimating health effects and the associated externalities associated with changes in air quality. Furthermore, BenMAP is a geographic information system-based program that estimates population-level exposure

rates, changes in incidence of health outcomes associated with air quality, and the human health externalities (Abt Associates Inc., 2012). ICF International also employed BenMAP as a resource to create a response surface, called a damage function, to evaluate changes in health endpoints as a result of the emissions reduction goals proposed in the Clean Power Plan (U.S. Environmental Protection Agency, 2015). In the work done by Voorhees *et al.*, various climate change scenarios within the air quality model govern scenarios in this work, and the output is then quantified in BenMAP (Voorhees, et al., 2011).

As it relates to the electricity sector, Farkas et al., study the impact of electricity generation on air quality during high-energy demand days. In this work, the researchers investigate the accuracy with which emissions from the electricity sector are represented in the Pennsylvania-New Jersey-Maryland (PJM) Interconnection, specifically $PM_{2.5}$. By analyzing the input parameters to the SMOKE model and testing 267 units from 91 stationary sources in the PJM network, the researchers found that, due to inaccurate temporalization assumptions, emissions from this sector are over-predicted and under-predicted at various snapshots in time and space.

2.3.2. *Stochastic GEP Models*

In the GEP problem, there are many elements that can be considered random variables due to their inherent uncertainty. For instance, fuel prices and electricity demand are considered to be random variables, since there is no way that we can predict their values at times in the future. Researchers have studied uncertainties in the GEP problem, by various methods. Stochastic optimization is a powerful tool that is

commonly used. Other methods involve scenario generation, fuzzy theory, and even dynamic programming.

In early work, Mo *et al.* consider uncertainty in different variables in the least-cost GEP such as energy demand, water inflow for hydropower, fuel prices, and investment costs (Mo, Hegge, & Wangenstein, 1991). A discrete-time Markov chain is used to model these variables in each year of the model. Backwards stochastic dynamic programming is applied in order to find the optimal expansion plan.

Gorenstin *et al.* solve the least-cost GEP problem by combining stochastic optimization with Benders' decomposition. They consider uncertainty in several factors such as demand growth, fuel cost, project completion, dates, and financial constraints. They split the GEP (via Benders cuts) into 2 sub-problems – an investment sub-problem, which determines the optimal investment plan and the operational sub-problem, which determines the optimal aggregate dispatch plan (Gorenstin, Campodonico, Costa, & Pereira, 1993). Stochastic programming is used to obtain an optimal solution. Additionally, the authors also study the “regret”, which is the deviation from deterministic optimality due to the stochastic nature of the future. Malcolm and Zenios did similar work; however, they approach the uncertainty by generating scenarios. They use stochastic programming to determine the least-cost expansion plan across all scenarios (Malcolm & Zenios, 1994).

Pokharel and Ponnambalam solve the least-cost stochastic GEP problem by generating multiple scenarios, where equipment, costs, and demand all take on different values in the model (Pokharel & Ponnambalam, 1997). They use linear programming to

determine the optimal expansion plan in each individual scenario, and also combine all of the scenarios in one model that aims to find an optimal solutions across all scenarios.

Su *et al.* apply fuzzy theory to parameters in the model in order to solve the least-cost stochastic GEP problem. Specifically, they use fuzzy sets to classify the "degree" to which each unit type produces different kinds of pollutants (Su, Lii, & Chen, 2000). This allows them to determine different Markovian states of pollution for different generation units. Dynamic programming is used to solve the optimization problem. Liu *et al.* also apply fuzzy sets, but in a more general approach for constraints and constraint limits (Liu, Huang, & Li, 2008).

Min and Subramanian study the stochastic GEP problem from a different perspective. When a plant is forced to shut down or close, there is some economic value associated with that stranded cost. The model uses stochastic programming to determine the optimal expansion plan, and also incorporates the mean-variance method in order to model the probability of a plant being stranded (Min & Subramanian, 2002). This allows for estimation or quantification of stranded cost. Lopez *et al.* also utilize mean-variance theory in a two-stage model, where risk is minimized in the first stage, and cost is minimized in the second stage (Lopez, Ponnambalam, & Quintana, 2007).

Mazadi *et al.* solve the stochastic least-cost GEP problem via chanced-constrained programming, which is a variation of stochastic programming. The authors add constraints in the form of density functions to describe the demand (Mazadi, Rosehart, Malik, & Aguado, 2009). More specifically, they assign boundaries, and use a density function to determine the probability the demand falls within those bounds at any given time.

2.3.3. GEP Models Using Metaheuristics

Due to the complexity of the generation expansion planning problem, typically nonlinear, mixed integer-programming methods are often used. These solution approaches work very well when there are small systems, or if there aren't many nonlinear constraints in the problem. However, often times, researchers study large-scale systems with very complex constraints to make the problem more realistic. Common nonlinear mixed integer programming methods take a long time to reach optimality in such conditions, and often encounter cycling and other challenges in finding an optimal solution. For these reasons, researchers in this field have begun using meta-heuristics to find optimal or near optimal solutions to the GEP problem. The most commonly used meta-heuristic to solve the GEP is the genetic algorithm (GA).

Fukuyama and Chiang used a GA to solve the GEP problem. Specifically, they apply a parallel genetic algorithm to optimal long-range generation expansion planning using multiple processors in parallel (Fukuyama & Chiang, 1996). They formulate the problem as a combinatorial optimization problem, where binary and decimal coding for the string representation method is compared. They apply this model to solve the least-cost optimization problem on a system with four technologies and five planning intervals.

Park *et al.* also used a GA to solve the GEP problem; however, they observed that a challenge of GA's is local optimality. In order to avoid local optimality, they developed a hybrid approach, which combines GA with dynamic programming (Park, Park, & Won, 1998). Essentially, the hybrid algorithm uses dynamic programming to hone in on candidate populations obtained from the GA. Additionally, they include

reliability constraints (LOLP and reserve margin), which adds further complexity to the model.

Park *et al.* solved the GEP problem by combining a GA with a decomposition method similar to the Dantzig-Wolfe method (Park J.-B. , Park, Won, & Lee, 2000). This translates the problem into several subproblems, one for each generating company. Each generator's subproblem is solved individually with a GA, and extreme directions are passed up to the master problem, which is again solved with a GA. Firmo and Legey did similar work, by combining GA with Benders' decomposition to split the least-cost GEP problem into 2 subproblems (an investment problem & operational problem), which are both solved by a GA (Firmo & Legey, 2002).

Kannan *et al.* apply particle swarm optimization (PSO) to solve the least-cost GEP problem. As previously mentioned, the GEP problem is a highly constrained, combinatorial optimization problem. PSO is applied in this problem along with a 'virtual mapping procedure' (VMP) to enhance the effectiveness of the PSO meta-heuristic (Kannan, Baskar, McCalley, & Murugan, 2009). In addition, the authors apply a penalty function approach (PFA) is used to reduce the size of the search space for subsequent iterations of the PSO algorithm. Prior to this research, however, Kannan *et al.* applied the VMP and PFA to various meta-heuristic techniques, including GA, Evolutionary Programming, Ant Colony Optimization, Tabu Search, and Simulated Annealing among others (Kannan S. , Slochanal, Baskar, & Murugan, 2007).

Yildirim *et al.* apply a method referred to as an adaptive simulated annealing genetic algorithm to solve the GEP problem for Turkey's power system (Yildirim, Erkan, & Ozturk, 2006). Since GA's show some limitations in large-scale optimization

problems, the authors decided to use simulated annealing in place of the mutation operator to improve the GA's convergence. Along these lines, Chen *et al.* combined a tabu search with a GA in order to avoid local optimality (Chen, Kang, Xia, & Zhong, 2010).

2.4. Metamodeling Techniques

In optimization problems, metamodels can be used as surrogate functions to replace the objective function or constraints. These metamodels are built from sampled data obtained from randomly searching the design space. Once the metamodel is built an optimization algorithm can be used to search the new design that is most likely to be optimal (or near-optimal). Predictions with a metamodel are generally more efficient. Furthermore, the computational cost associated with the search based on the metamodel is usually negligible.

Metamodeling is referred to as a technique that utilizes the sampled data to build metamodels, which are sufficient to predict the output of an expensive computer code at unobserved locations in the design space. Choosing sample points, building metamodels, and evaluating the accuracy of the metamodel are among the key issues in metamodeling.

2.4.1. Experimental Design

Experimental design methods are usually used to determine the locations of sample points in the design space. Experimental design (sometimes called “design of experiments” or “DoE”) is a procedure with the ultimate goal of maximizing the information obtained from a limited number of sample points. DoE methods can be classified into two major categories - “classic” and “modern” (Han & Zhang, 2012). Traditional DoE methods include full-factorial design, central composite design (CCD),

Box-Behnken and D-Optimal Design (DOD), were development of laboratory experiments, with the consideration of reducing random error effects. Oppositely, “modern” DoE methods such as Latin Hypercube Sample (LHS), Orthogonal Array Design (OAD) and Uniform Design (UD) were developed for deterministic computer experiments.

2.4.2. Metamodels

While there are many metamodels available in the literature, this discussion will be limited to three popular techniques – polynomial Response Surface Models (pRSM), Radial Basis Functions (RBF), and Kriging.

Consider an m -dimensional problem, where the goal is to predict the output of a high-fidelity, thus expensive computer code, which corresponds to an unknown function $y: \mathbb{R}^m \rightarrow \mathbb{R}$. By running the computer code, y is observed at n sites, which are determined by the DoE.

$$\mathbf{S} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}]^T \in \mathbb{R}^{n \times m} \quad (2.22)$$

$$\mathbf{x} = [x_1, \dots, x_m] \in \mathbb{R}^m \quad (2.23)$$

$$\mathbf{y}_s = [y^{(1)}, \dots, y^{(n)}]^T = [y(\mathbf{x}^{(1)}), \dots, y(\mathbf{x}^{(n)})]^T \in \mathbb{R}^n \quad (2.24)$$

The vector \mathbf{x} corresponds to the independent variables x_1 through x_m . The matrix \mathbf{S} corresponds to the n observations of the m independent variables. The vector \mathbf{y}_s corresponds to the n observed responses that are functions of the independent variables. The pair $(\mathbf{S}, \mathbf{y}_s)$ denotes the sampled data sets in the vector space. Using the relationships defined in Equations 2.22 to 2.24, the objective is to build a metamodel to predict the output of the computer code for any unobserved location \mathbf{x} based on sampled data sets $(\mathbf{S}, \mathbf{y}_s)$, in order to obtain accurate predictions with a minimal number of sample points (Han & Zhang, 2012).

Quadratic Response Surface Method

Polynomial response surface models (pRSMs) denote a class of least-squares regression models that fit the data to a polynomial curve. Quadratic pRSMs provide the best compromise between modeling accuracy and computational expense in comparison

to other linear or higher order polynomial models. An advantage of pRSMs is that it can smooth out the various scales of numerical noise in the data while capturing the population variation (Han & Zhang, 2012). This makes pRSM a very robust approach for optimization problems in engineering. The true quadratic pRSM has the following form (Montgomery, Peck, & Vining, 2006).

$$y(\mathbf{x}) = \hat{y}(\mathbf{x}) + \varepsilon, \mathbf{x} \in \mathbb{R}^m \quad (2.25)$$

$\hat{y}(\mathbf{x})$ is a quadratic polynomial approximation and ε is the random error term, which is assumed to follow a normal distribution with mean zero and variance σ^2 . The errors at each observed location, ε_i , are assumed to be independent and identically distributed. The quadratic pRSM predictor, $\hat{y}(\mathbf{x})$, is defined as follows (Montgomery, Peck, & Vining, 2006).

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 + \sum_{i=1}^m \hat{\beta}_i x_i + \sum_{i=1}^m \hat{\beta}_{ii} x_{ii}^2 + \sum_{i=1}^m \sum_{j \geq i}^m \hat{\beta}_{ij} x_i x_j \quad (2.26)$$

In Equation 2.26, β_0 , β_i , β_{ii} , and β_{ij} are the unknown model parameters that are determined by least-squares (Montgomery, Peck, & Vining, 2006). In this model, there are a total of $p=(m+1)(m+2)/2$ unknown coefficients. Thus, in this case, a quadratic pRSM requires at least p sample points. Letting $\boldsymbol{\beta} \in \mathbb{R}^p$ be the column vector that corresponds to these p unknown coefficients, gives the following least-squares estimator of $\boldsymbol{\beta}$ which is given as follows.

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_s \quad (2.27)$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_m^{(1)} & x_1^{(1)}x_2^{(1)} & \cdots & x_{m-1}^{(1)}x_m^{(1)} & (x_1^{(1)})^2 & \cdots & (x_m^{(1)})^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & \cdots & x_m^{(n)} & x_1^{(n)}x_2^{(n)} & \cdots & x_{m-1}^{(n)}x_m^{(n)} & (x_1^{(n)})^2 & \cdots & (x_m^{(n)})^2 \end{bmatrix} \in \mathbb{R}^{n \times p} \quad (2.28)$$

\mathbf{X} is called the design matrix. The resulting model will be a hyperplane in \mathbb{R}^{p+1} space (Montgomery, Peck, & Vining, 2006). After the coefficients are found by Equation 2.27, the fitted model response at any unobserved location can be predicted at any unobserved location by Equation 2.26.

Radial Basis Functions

Radial Basis Functions (RBFs) are interpolation models for metamodeling. In RBFs, the approximation of an unknown function $y(\mathbf{x})$ at an unobserved location \mathbf{x} is defined as a linear combination of the radial basis functions and a global trend function (Han & Zhang, 2012).

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^n \omega_i \varphi(\mathbf{x}) + P(\mathbf{x}) \quad (2.29)$$

In Equation 2.29, ω_i is the i -th weight coefficient, $\varphi(\mathbf{x}) = \varphi(\|\mathbf{x}^{(i)} - \mathbf{x}\|)$ are the basis functions that depend on the Euclidean distance between the observed location $\mathbf{x}^{(i)}$ and the unobserved location \mathbf{x} , and $P(\mathbf{x})$ is the global trend function, which is usually chosen to be linear or constant. Also, to ensure that the prediction function reproduces the responses at observed locations, the following constraints need to be satisfied (Han & Zhang, 2012).

$$\hat{y}(\mathbf{x}^{(i)}) = y^{(i)}, i = 1, \dots, n \quad (2.30)$$

$$\sum_{i=1}^n \omega_i = 0 \quad (2.31)$$

In the case where $P(\mathbf{x})$ is a constant ($P(\mathbf{x})=\beta_0$), solving the system of equations formed by Equations 2.30 and 2.31 for ω_i and β_0 , and substituting into Equation 2.29 gives the RBF predictor as follows.

$$\hat{y}(\mathbf{x}) = \beta_0 + \boldsymbol{\varphi}^T(\mathbf{x})\boldsymbol{\Psi}^{-1}(\mathbf{y}_S - \beta_0\mathbf{1}) \quad (2.32)$$

In Equation 2.32, $\beta_0=(\mathbf{1}^T\boldsymbol{\Psi}^{-1}\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{\Psi}^{-1}\mathbf{y}_S$, where $\boldsymbol{\Psi}$ and $\boldsymbol{\varphi}(\mathbf{x})$ are defined as follows.

$$\boldsymbol{\Psi} = \left[\varphi(\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|) \right]_{ij} \in \mathbb{R}^{n \times n} \quad (2.33)$$

$$\boldsymbol{\varphi}(\mathbf{x}) = \left[\varphi(\|\mathbf{x}^{(i)} - \mathbf{x}\|) \right]_i \in \mathbb{R}^n \quad (2.34)$$

In order to build RBF models, one needs to specify the appropriate basis functions, which depends on the Euclidean distance $r=\|\mathbf{x}-\mathbf{x}'\|$ between any two locations \mathbf{x} and \mathbf{x}' . Common basis functions for RBF models are given in Table 2.1.

Table 2.2 – Basis Functions for RBF Metamodels (*Han & Zhang, 2012*)

Basis functions	Formulations
Gaussian (GAUSS)	$\phi(r) = e^{-r^2/2\sigma^2}$ (e.g. $\sigma^2 = 1$)
Power function (POW)	$\phi(r) = r^\beta, 1 \leq \beta \leq 3$ (e.g. $\beta = 1.8$)
Thin Plate Spline (TPS)	$\phi(r) = r^2 \ln(r)$
Hardy's Multiquadric (HMQ)	$\phi(r) = \sqrt{1 + r^2}$
Hardy's Inverse Multiquadric (HIMQ)	$\phi(r) = 1 / \sqrt{1 + r^2}$

2.4.2.1 Kriging Models

Kriging is an interpolation method that uses the observed data at all sample points to provide a statistical prediction of an unknown function by minimizing its' Mean Squared Error (MSE) (Bohling G. , 2005). To develop a kriging metamodel, the output of a deterministic computer experiment is treated as a realization from a stochastic

process, which is then defined as the sum of a global trend function $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ and a Gaussian stochastic process $Z(\mathbf{x})$ as follows.

$$y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^m \quad (2.35)$$

In Equation 2.35, $\mathbf{f}(\mathbf{x})=[f_0(\mathbf{x}), \dots, f_{p-1}(\mathbf{x})]^T \in \mathbb{R}^p$ is defined with a set of regression basis functions and $\boldsymbol{\beta}=[\beta_0, \dots, \beta_{p-1}]^T \in \mathbb{R}^p$ is the vector of the corresponding coefficients. Typically, $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ is taken as a constant value or a low-order polynomial. In practice, the constant trend function is sufficient for most problems. Also, $Z(\mathbf{x})$ represents a stationary, zero-mean, stochastic process with variance σ^2 and nonzero covariance given as follows.

$$\text{Cov}[Z(\mathbf{x}), Z(\mathbf{x}')] = \sigma^2 R(\mathbf{x}, \mathbf{x}') \quad (2.36)$$

In Equation 2.36, $R(\mathbf{x}, \mathbf{x}')$ is the correlation function, which depends on the Euclidean distance between any two locations \mathbf{x} and \mathbf{x}' in the design space (Bohling G. , 2005). The most commonly used correlation function is a Gaussian exponential function of the following form.

$$R(\mathbf{x}, \mathbf{x}') = \exp\left[-\sum_{k=1}^m \theta_k |x_k - x'_k|^{p_k}\right], 1 < p_k \leq 2 \quad (2.37)$$

In Equation 2.37, $\boldsymbol{\theta}=[\theta_1, \theta_2, \dots, \theta_m]^T$ and $\mathbf{p}=[p_1, p_2, \dots, p_m]^T$ define the vectors of the unknown model parameters to be tuned (also commonly referred to as hyper-parameters) (Han & Zhang, 2012). In the case where a constant global trend function is used, using the previous equations, the Kriging predictor, $\hat{y}(\mathbf{x})$, for any unobserved location can be written as defined in Equations 2.38 and 2.39.

$$\hat{y}(\mathbf{x}) = \boldsymbol{\beta}_0 + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y}_s - \boldsymbol{\beta}_0\mathbf{1}) \quad (2.38)$$

$$\beta_0 = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}_s \quad (2.39)$$

\mathbf{R} and \mathbf{r} are the correlation matrix and correlation vector respectively, and are given in Equations 2.40 and 2.41.

$$\mathbf{R} = \begin{bmatrix} R(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & R(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & R(\mathbf{x}^{(1)}, \mathbf{x}^{(n)}) \\ R(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & R(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & R(\mathbf{x}^{(2)}, \mathbf{x}^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ R(\mathbf{x}^{(n)}, \mathbf{x}^{(1)}) & R(\mathbf{x}^{(n)}, \mathbf{x}^{(2)}) & \dots & R(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}) \end{bmatrix} \in \mathbb{R}^{n \times n} \quad (2.40)$$

$$\mathbf{r} = \begin{bmatrix} R(\mathbf{x}^{(1)}, \mathbf{x}) \\ R(\mathbf{x}^{(2)}, \mathbf{x}) \\ \vdots \\ R(\mathbf{x}^{(n)}, \mathbf{x}) \end{bmatrix} \in \mathbb{R}^n \quad (2.41)$$

$R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is the correlation between any two observed locations $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ and $R(\mathbf{x}^{(i)}, \mathbf{x})$ is the correlation between the i -th observed location $\mathbf{x}^{(i)}$ and the unobserved location \mathbf{x} (Bohling G. , 2005).

The MSE for prediction using a Kriging metamodel is defined in Equation 2.42. The MSE, as mentioned in the previous section, provides an estimate of the prediction uncertainty.

$$\hat{s}^2(\mathbf{x}) = \sigma^2 [1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + (\mathbf{r}^T \mathbf{R}^{-1} \mathbf{1} - 1)^2 / \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}] \quad (2.42)$$

Furthermore, in Kriging models, a major assumption is that the sampled data follows a normal distribution, and the responses at sampled locations are correlated random functions with a likelihood function given in Equation 2.43 (Bohling G. , 2005).

$$L(\beta_0, \sigma^2, \theta, \mathbf{p}) = \frac{1}{\sqrt{2\pi(\sigma^2)^n |\mathbf{R}|}} \exp \left(-\frac{(\mathbf{y}_s - \beta_0 \mathbf{1})^T \mathbf{R}^{-1} (\mathbf{y}_s - \beta_0 \mathbf{1})}{2\sigma^2} \right) \quad (2.43)$$

The optimal estimates of β_0 and the process variance are obtained by maximizing the likelihood function in Equation 2.43. This can be achieved by various numerical

optimization algorithms such as GA or Newton-Raphson (Bohling G. , 2005). These estimates are given in Equation 2.44 and 2.45.

$$\boldsymbol{\beta}_0(\boldsymbol{\theta}, \mathbf{p}) = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}_s \quad (2.44)$$

$$\boldsymbol{\sigma}^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}, \mathbf{p}) = \frac{1}{n} (\mathbf{y}_s - \boldsymbol{\beta}_0 \mathbf{1})^T \mathbf{R}^{-1} (\mathbf{y}_s - \boldsymbol{\beta}_0 \mathbf{1}) \quad (2.45)$$

2.4.3. Evaluation of Metamodels

A critical aspect of metamodeling is estimating the prediction error of the model. Metamodels with sufficient accuracy and prediction capabilities allow for more reliable searches for optimality. Two commonly used criteria for evaluating the prediction error are the average relative error (\bar{e}) and root mean squared error (σ_e) (Montgomery, Peck, & Vining, 2006). The average relative error is defined in Equation 2.46.

$$\bar{e} = \frac{1}{n_t} \sum_{i=1}^{n_t} \left\| \frac{\hat{y}_t^{(i)} - y_t^{(i)}}{y_t^{(i)}} \right\| \quad (2.46)$$

The number of test points is denoted by n_t . $y_t^{(i)}$ and $\hat{y}_t^{(i)}$, are the true and predicted values corresponding to the i -th test point, respectively. The root mean squared error is defined in Equation 2.47.

$$\sigma_e = \sqrt{\frac{1}{n_t} \sum_{i=1}^{n_t} \left\| \frac{\hat{y}_t^{(i)} - y_t^{(i)}}{y_t^{(i)}} \right\|^2} \quad (2.47)$$

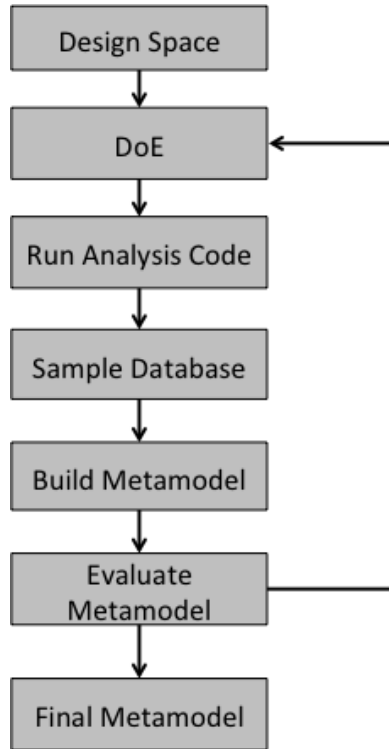


Figure 2.5 – Framework for Building Metamodels (*Han & Zhang, 2012*)

The framework of building a metamodel is presented in Figure 2.5. The initial metamodel can be evaluated by Equations 2.46 and 2.47. Additionally, upon evaluating the metamodel, the design space is then resampled and a new metamodel is fit, as evidenced by the arrow that circulates backward in the chain from “Evaluate Metamodel” to “DoE”.

2.4.4. Metamodels for Optimization

A typical optimization problem has the following form given in Equations 2.48 to 2.50.

Minimize

$$y(\mathbf{x}) \quad (2.48)$$

Subject to

$$g_i(\mathbf{x}) \leq 0, i = 1, \dots, n_c \quad (2.49)$$

$$\mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \quad (2.50)$$

In this formulation, n_c is the number of constraints (denoted by $g_i(\mathbf{x})$). \mathbf{x}_l and \mathbf{x}_u are the lower and upper bounds of the decision variables, respectively. The objective function is $y(\mathbf{x})$. In this formulation, the objective function and/or the constraints are evaluated by expensive analysis code. Typically, such problems are solved by a gradient-based algorithm or a metaheuristic. However, problems with these characteristics are computationally expensive to reach an optimal or near-optimal solution. To avoid this pitfall, metamodeling methods are utilized to improve the efficiency of optimization algorithms.

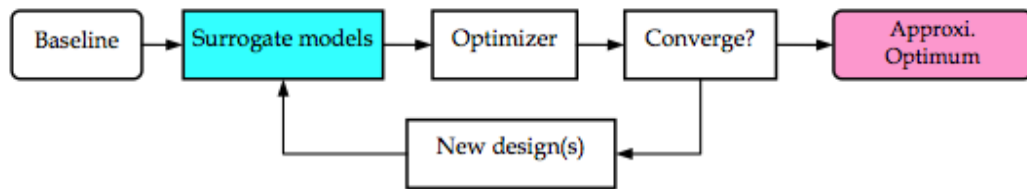


Figure 2.6 – Simple Framework for Simulation-Based Optimization (Han & Zhang, 2012)

Figure 2.6 illustrates the simple framework for simulation-based optimization. The general concept of using metamodels in optimization is quite simple. First, the metamodels for the objective function and constraints are built with sufficient accuracy. These metamodels serve as surrogates in an optimization problem. Then, an approximate optimal solution is found by an optimization algorithm or a metaheuristic.

In the remaining portion of this sub-section, the infill criterion for determining new sample locations by solving sub-optimization problems is presented. Specifically, Searching Metamodels (SM) and Expected Improvement (EI) are discussed.

Searching Metamodels

Given that the initial metamodels have been built with sufficient accuracy, an optimization algorithm or metaheuristic can be used to find an approximate optimal solution. This, in turn, can be used to refine the metamodels. The mathematical formulation of the sub-optimization for determining the new sample location is given in Equations 2.51 to 2.53 (Gosavi, 2003).

Minimize

$$\hat{y}(\mathbf{x}) \quad (2.51)$$

Subject to

$$g_i(\mathbf{x}) \leq 0, i = 1, \dots, n_c \quad (2.52)$$

$$\mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \quad (2.53)$$

In this formulation, metamodels for the objective function and the constraints serve as surrogates for the true functions. With optimal solution of this problem, \mathbf{x}^* , in hand, one needs to run the expensive analysis code to compute the true function value and compare it with that is predicted by the metamodels. If this error is below a specified threshold, then the optimization process can be terminated. Otherwise, a new sample location is augmented to the sampled data sets and the metamodels are then rebuilt. This procedure is repeated until an approximate optimal solution is found.

Expected Improvement

The expected improvement function is calculated to determine the improvement of an objective function that is expected to be achieved at any unobserved location \mathbf{x} (Han & Zhang, 2012). The mathematical representation is given in Equation 2.54.

$$E[I(\mathbf{x})] = \begin{cases} (y_{\min} - \hat{y}(\mathbf{x}))\Phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}(\mathbf{x})\phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) & \text{if } \hat{s} \geq 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \quad (2.54)$$

$\Phi(\cdot)$ and $\phi(\cdot)$ represent the cumulative distribution function and probability density function of a standard normal distribution respectively in Equation 2.54. Also, y_{\min} is the minimum response of the observed data thus far. s , defined in Equation 2.42, is the square root of the MSE. The point with the maximum EI is located by a global optimization algorithm or metaheuristic, and then observed by running the analysis code. In this case, the constraints are accounted for by introducing the probability that the constraints are satisfied into a sub-optimization problem (Gosavi, 2003).

Maximize

$$E[I(\mathbf{x})] = \prod_{i=1}^{n_c} P[G_i(\mathbf{x}) \leq 0] \quad (2.55)$$

Subject to

$$\mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \quad (2.56)$$

$P[G_i(\mathbf{x}) \leq 0]$ in Equation 2.54 denotes the probability that the i -th constraint is satisfied. $G_i(\mathbf{x})$ is a random function corresponding to the i -th constraint, $g_i(\mathbf{x})$. If the constraint is satisfied, $P[G_i(\mathbf{x}) \leq 0]$ will approach 1, but if the constraint is violated it approaches 0 (Han & Zhang, 2012). $P[G_i(\mathbf{x}) \leq 0]$ is given in Equation 2.56.

$$P[G_i(\mathbf{x}) \leq 0] = \Phi\left(\frac{-\hat{g}_i(\mathbf{x})}{\hat{s}_i(\mathbf{x})}\right) \quad (2.57)$$

The solution obtained from solving this sub-optimization problem, \mathbf{x}^* , is observed by running the analysis code, and the new sampling location is added to the sampled data sets. Metamodels are then rebuilt and the whole process is repeated until an approximate global optimal solution is obtained.

2.4.5. Applications

In early works, Roux *et al.* researched techniques to find parts of the feasible region of an optimization problem with a surrogate objective function, which contain the optimal solution. Additionally, they studied methods for finding more accurate regression models to approximate the value of the objective function. They found that for problems with large feasible regions, the problem could be decomposed into multiple sub-problems by partitioning the feasible region, similar to Dantzig-Wolfe decomposition (Roux, Stander, & Haftka, 1998). They apply regression theory, experimental design, and feasible region partitioning approaches in various structural optimization examples. The results indicate that finding the true global optimum solution is extremely difficult. Accurate solutions could only be achieved by considering a smaller section of the feasible solution space.

Response surface methodology is a very common approach to solving chemical and biochemical process problems. Kalil *et al.* apply response surface techniques along with factorial design in an industrial bioprocess optimization problem. They apply regression analysis to the output of a factorial designed experiment. From the model, they determined that five of the ten total process parameters were statistically significant in predicting the responses (yield and productivity) (Kalil, Maugeri, & Rodrigues, 2000). These served as surrogate objective functions in an optimization model, which aimed to determine the optimal settings of these parameters that maximize yield and productivity subject to process constraints. In similar work, Quanhong and Calil applied RSM methods to study the effect of liquid:solid ratio, NaCl concentration and reaction time on the production of protein from germinant pumpkin seeds. They apply regression models

to study the shape of the surface and the statistical significance of key covariates (Quanhong & Caili, 2005). They then use this regression model as a surrogate objective function in an optimization problem.

Often times, RSM is used in optimization problems where there are no constraints. In this case, the response surface is analyzed to find a minimum (or maximum depending on the problem). Chen *et al.* apply RSM to evaluate the effects of enzyme concentration, temperature, molar ratio of methanol to oil, and stirring rate on biodiesel production yields. Using a 5-factor experiment with 5-levels, they use a second order regression model to predict biodiesel production yield as a function of these parameters. From these results, they found that all model covariates were statistically significant. Furthermore, by investigating the fitted surface, they determined the optimal solution. By conducting further experiments, these results were validated.

RSM methods are commonly used in design problems. Specifically, researchers have applied RSM methods to determine the best design specifications based on metrics of quality or some other relative scale, while considering budget constraints among others. Youn and Choi developed a method to design a system based on an estimate of its' reliability, which is called reliability-based design optimization (RBDO). In previous works, RBDO involves evaluation of probabilistic constraints that may be exorbitantly expensive or may even diverge in some cases. Because of this, they modify the commonly used RBDO approach by integrating this with response surface methodology (Youn & Choi, 2004).

Kriging metamodels are more accurate than regression models as predictors. When the cost of evaluating the response of a system by experiments is high, it is critical

to use the most accurate metamodel in predicting the response surface. In comparison to other metamodeling methods used in optimization, kriging provides the best predictions, and serves as a more accurate and reliable surrogate objective function in comparison to pRSM and RBFs (Han & Zhang, 2012).

Huang *et al.* developed a sequential kriging-optimization method, which can be applied to surrogate systems to reduce the total cost of response evaluation. They utilize data from computer experiments to develop a kriging metamodel that provides a global prediction of the objective function and a measure of prediction uncertainty (Huang D. , Allen, Notz, & Miller, 2006). The location and accuracy of subsequent evaluations are obtained by maximizing an expected improvement function, which is based on the cost of evaluation. They apply this method to the metal-forming process design, which demonstrated the robustness and repeatability of this approach. These researchers later extended this work to address black-box systems, where, in such systems, metamodel predictions have a measure of prediction uncertainty at every point (Huang, Allen, Notz, & Zeng, 2006). Similar results were obtained when the procedure was applied to a production-inventory problem (Huang, Allen, Notz, & Miller, 2006). Jakumeit *et al.* also applied kriging to the metal-forming process, by combining kriging with gradient and direct search optimization algorithms (Jakumeit, Herdy, & Nitsche, 2005).

In aerospace and aircraft design applications, determining the optimal design configuration requires metamodels with a high degree of accuracy. Jeong and Murayama researched this subject, and developed a kriging-based genetic algorithm for aerodynamic design problems. They found that kriging models drastically reduce computational time required for finding optimal (or near-optimal) solutions in comparison to regression-

based methods (Jeong & Murayama, Efficient Optimization Design Method Using Kriging Model, 2005). Additionally, by sequentially maximizing an expected improvement function, additional sample points are used to fit the kriging metamodel. Through an iterative procedure of metamodeling, optimization, and sampling, the researchers found that not only does the prediction accuracy of the metamodel improve, but the genetic algorithm's ability to explore optimal solutions improves accordingly. Moreover, they were able to reduce the number of design variables by studying the results of a functional ANOVA study.

Sakata *et al.* applied kriging approximation and optimization in a structural optimization problem. The response measured in the kriging model is a metric that measures the reinforcement properties of beams (Sakata, Ashida, & Zako, Structural optimization using Kriging approximation, 2003). The optimization model determines the optimal layout. Kriging estimation is compared with neural network approximation, and thus demonstrated that kriging is a better approach. The obtained results clearly show the applicability of the method. They later extended this work to develop an algorithm to improve a computational cost for estimation using the kriging method with a large number of sampling data (Sakata, Ashida, & Zako, An efficient algorithm for Kriging approximation and optimization with large-scale sampling data, 2004). This improved algorithm computes the weighting coefficients for kriging by applying the Sherman–Morrison–Woodbury formula. Ratle applied a similar method in a fitness landscape optimization problem by combining kriging with various evolutionary programming methods (Ratle, 2001).

Li *et al.* combined simulation, kriging, and multi-objective optimization in their work. They present a new multi-objective design optimization method where the kriging-based metamodel is embedded within a multi-objective genetic algorithm (MOGA) (Li, Li, & Azarm, 2008). The proposed approach is called Kriging assisted MOGA (K-MOGA). The key difference between K-MOGA and a conventional MOGA is that in K-MOGA some of the decision variables are evaluated on-line using kriging metamodeling. To determine whether simulations are needed or a kriging metamodel should be used is based on evaluating a set of decision variables is determined by a metric that determines whether its “domination status” in the current generation can be changed.

3. Analytical Framework

In this section, the analytical framework of simulation-based optimization in the context of solving the generation expansion-planning problem is presented. Specifically, a conceptual and mathematical roadmap is presented which outlines the models used in the analyses and defines their respective relationships within the framework. These models were identified and selected based on literature surveys and industry practices. In assessing the inputs and outputs of these models, linear linkages between the chosen models are established. This allows for the execution of controlled, simulated, computer experiments to observe the response in question, human health externalities, as a function of key variables that are parameters in our suite of models. Applying a metamodel to the output of these experiments allows human health externalities to be quantified explicitly as a function of electricity generation. This allows for the formulation of a generation expansion planning problem that not only minimizes investment, variable and fixed operating and maintenance costs, and fuel costs, but minimizes human health externalities and the social cost of carbon and methane leakage as well.

3.1 Establishing Model Relationships

Our primary objectives of this research are to: (i) establish a mathematical relationship to predict human health externalities associated with electricity generation and emissions; and (ii) develop a generation expansion planning (GEP) model that minimizes investment costs, fixed operating and maintenance costs, variable operating and maintenance costs (including fuel), the social cost of carbon and methane leakage, and human health externalities associated with pollutant emissions from electricity generation. Conceptually, however, these goals present a major challenge. From a health

perspective, current GEP research focuses on considering health and environmental impacts by considering human health externalities as deterministic multipliers in the objective function of an optimization model, minimizing the cost of purchasing emissions permits, or defining emissions limits and renewables requirements as constraints. While these approaches indirectly account for human health externalities, there is a need for a more direct approach to solving this problem.

As identified in the literature, there are various approaches to quantifying human health externalities. Mainly, the research is saturated with less rigorous analyses of electricity market activity using air quality and economic models. Research in this area can be advanced by explicitly including expansion planning decisions into such analyses.

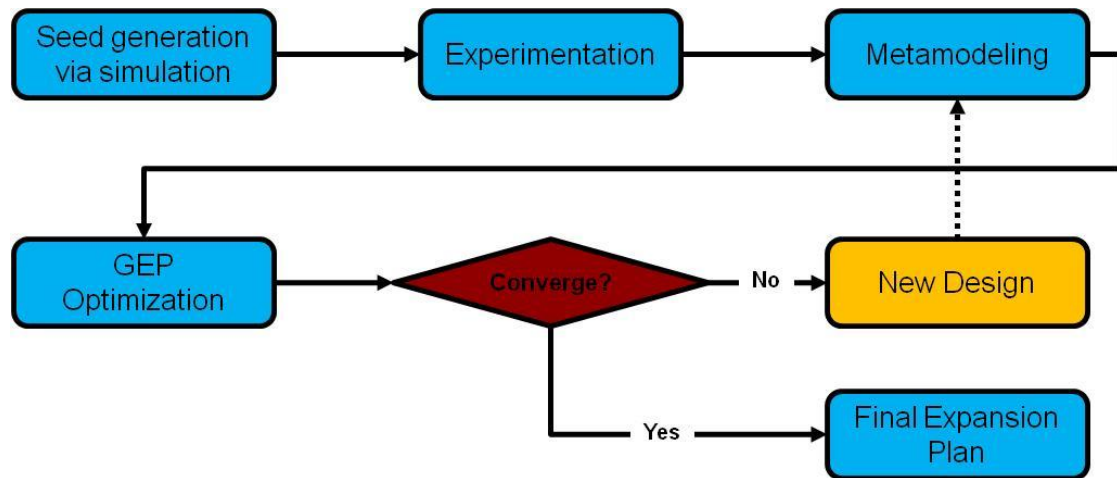


Figure 3.1 – GEP Simulation-Based Optimization Framework Considering Human Health Externalities

Upon reviewing literature related to generation expansion planning models, the analytical framework presented in Figure 3.1 was developed. The first step in this process is to generate expansion planning scenarios (also known as seed generation) to be evaluated in an EPA screening tool, COBRA, to quantify human health externalities. A kriging metamodel is then applied to this data set, which predicts human health

externalities as a function of electricity dispatch decisions. This metamodel then serves as a surrogate function for human health externalities in the objective function in a generation expansion planning model. Criteria are then evaluated to determine when to terminate the algorithm or where to sample data points that better enhance the metamodel.

3.2. Seed Generation via Generation Expansion Planning Simulations and Experimentation

In order to create a metamodel of human health externalities as part of GEP, we must have a diverse sample space of data points from our expansion-planning model, which are then evaluated via an EPA screening tool – we call this process seed generation. To execute the seed generation process, first we solve a baseline generation expansion planning model. We then solve an experimental GEP model where key parameters are simulated from a normal distribution. Next, for each experimental trial, we compare the expansion decisions and obtain a change in MWh for each unit available within the system. Concurrently, for each trial we obtain the percent change in NO_x and SO_2 emissions from the baseline. The emissions changes we obtain from our simulated expansion planning models are evaluated in the COBRA model. Thus, at the end of this procedure, we obtain a dataset with the following structure outlined in Figure 3.2.

	MWh Change from Baseline						% Change from Baseline		
Simulation Trial	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	NO _x	SO ₂	Human Health Externalities (\$)
Trial 1	-155,875	-370,736	-94,619	279,400	-383,391	-254,012	-5%	38%	(\$4,679,291.00)
Trial 2	-376,755	248,947	-340,339	-163,159	124,657	208,962	-76%	-13%	(\$1,673,385.00)
Trial 3	-285,236	323,306	437,813	282,549	444,718	302,453	-74%	42%	(\$3,286,246.00)
Trial 4	-285,920	-174,044	-316,507	-92,520	-314,648	-175,169	-62%	-43%	\$3,556,307.00
Trial 5	189,508	455,590	270,230	-411,113	345,879	470,925	47%	45%	\$2,258,830.00
Trial 6	-199,258	-200,412	-327,419	318,627	486,077	212,394	22%	-91%	(\$6,107,685.00)
Trial 7	481,059	214,521	407,007	80,895	-125,196	199,188	-93%	74%	\$8,664,379.00
Trial 8	-354,102	-482,033	110,901	346,963	-87,192	182,781	-36%	-30%	\$6,555,750.00
Trial 9	-385,125	-391,491	279,862	-201,527	-478,628	482,235	-17%	-17%	(\$8,549,241.00)
Trial 10	-185,611	484,510	313,342	-141,125	-281,002	372,838	-81%	27%	\$3,432,419.00
Trial 11	-166,328	245,148	-110,929	-260,372	-250,323	287,905	67%	37%	(\$3,427,381.00)
Trial 12	411,629	-463,315	84,026	-306,282	-352,023	-183,367	88%	-89%	\$6,817,444.00
Trial 13	-223,595	-161,325	167,153	-477,120	-215,226	-433,026	-83%	-29%	(\$5,368,394.00)
Trial 14	108,827	-416,463	210,351	74,420	417,860	78,222	-46%	-21%	(\$4,106,579.00)
Trial 15	-363,222	167,212	-436,544	-112,195	-209,450	464,673	49%	-32%	(\$2,497,533.00)
Trial 16	-182,308	-236,162	-445,906	129,127	474,508	-320,458	-24%	19%	\$3,006,855.00

Figure 3.2 – GEP and COBRA Simulation Output Data Structure with Illustrative Data

3.2.1. Generation Expansion Planning Mathematical Model for Simulation

The electricity GEP problem determines the optimal investment schedule of generation technology options to be added to an existing grid. Additionally, dispatching decisions determine how much energy is produced from each generation source. In this model, there are three classes of decision variables, which are:

1. Investment decisions, which determine the type of generation technology to be added to the system, and when and where the investment is made.
2. Dispatching decisions, which determine how much energy each generation technology produces in aggregate to satisfy the demand in a particular region over a given time interval.
3. Transmission decisions, which determines how much energy from each generation technology is transmitted between regions within the network.

The problem studied is a least-cost generation expansion plan over a multi-period planning horizon starting with an existing centralized power system. The goal is to

integrate generation expansion and dispatching decisions, while considering human health externalities and the social cost of carbon and methane leakage along with investment, fixed operating and maintenance, and variable operating and maintenance costs (including fuel) in the objective function.

3.2.1.1. GEP Objective Function

The goal of this model is to determine the optimal expansion plan that minimizes total system-wide costs including investment costs, fixed operating and maintenance costs, variable operating and maintenance costs (including fuel), the social cost of carbon and methane leakage, and human health externalities associated with pollutant emissions from electricity generation. Also included in the objective function is revenue from steam generation. All future costs are discounted to the present time using a discount rate r to yield a net present value (NPV).

Investment Cost:

$$O_I = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I} \alpha_{yi} y_{yr_1 i} \quad (3.1)$$

Investment costs are the capital costs associated with expansion decisions. $y_{yr_1 i}$ is the continuous investment decision determining the capacity expansion values (in MW) by generation unit type i in year y in region r_1 . α_{yi} is investment cost (\$/MW) of unit generation unit type i in year y . In this equation, r is the interest rate, Y is the set of all years in the planning horizon, and I is the set of all generating unit options.

Fixed Operating and Maintenance Costs:

$$O_M = \sum_{y \in Y} \frac{1}{(1+r)^y} \left[\sum_{r_1 \in R} \sum_{i \in I} \left(\sum_{u (\forall u \leq y) \in Y} \beta_{ui} y_{ur_1 i} \right) + \sum_{r_1 \in R} \sum_{i \in I} \beta_{yi} q_{r_1 i} \right] \quad (3.2)$$

Fixed operating and maintenance costs are the annual maintenance costs associated with each unit. β_{yi} is the maintenance and operational cost (\$/MW) for unit i

in year y . I^* is the set of all existing generating unit options, and as mentioned previously, I is the set of all available generating unit options. $q_{r_1 i}$ is the initial capacity of unit i in region r_1 .

Variable Operating and Maintenance Cost:

$$O_G = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I} v_{yti} x_{ytr_1 i} \quad (3.3)$$

Variable operating and maintenance costs (including fuel costs) are costs per MWh of generation for each unit. $x_{ytr_1 i}$ is the amount of aggregate energy generation dispatched (MWh) from unit i in region r_1 period t of year y . v_{yti} is the variable generation cost (\$/MWh) associated with generation from unit i during period t of year y . The set of all periods in the model is given by T , and includes: spring/fall-offpeak, spring/fall-peak, summer-offpeak, summer-peak, winter-offpeak, winter-peak.

Social Cost of Carbon:

$$O_C = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I} \sum_{e \in E_c} c_y x_{ytr_1 i} \rho_{i,e} \quad (3.4)$$

The social costs of carbon are the costs per ton of CO₂ emissions from electricity generation. These costs are given by c_y , and ρ_{ie} denotes the CO₂ emissions rate of unit i . $x_{ytr_1 i}$ is the amount of aggregate energy generation dispatched (MWh) from unit i in region r_1 period t of year y .

Social Cost of Methane Leakage:

$$O_L = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I_L} \sum_{e \in E_c} c_y x_{ytr_1 i} \lambda_{i,e} \quad (3.5)$$

The social costs of methane leakage are the costs associated with methane leakage from natural gas and combined-cycle units. In this case, λ_{ie} is the methane leakage rate, which is calculated using the following methodology:

- Natural gas leaks at a rate between 1% and 8% (Brandt, et al., 2014). We consider this leakage rate as a random variable, uniformly distributed on this interval. We call this value A .
- Approximately 3,448.3 cubic feet of natural gas is used to generate 1 MWh of electricity from natural gas turbine, and approximately 2,664.6 cubic feet of natural gas is used to generate 1 MWh of electricity from combined-cycle gas turbines (United States Environmental Protection Agency, 2016). We denote these values using B_i , where i is the index of the units using natural gas as a fuel source.
- Per the EPA, natural gas is approximately 90% methane by composition, which is denoted by C (United States Environmental Protection Agency, 2016).
- The 20 year global warming potential (or GWP) of methane is approximately 72 times that of CO_2 (United States Environmental Protection Agency, 2016). To elaborate, the GWP is a measure of how much heat a greenhouse gas traps in the atmosphere, relative to CO_2 . We denote the GWP of methane by D .
- Finally, per the ideal gas law, the density of methane is function of temperature. However, per EPA projections, due to climate change, temperatures are expected to increase by 0.5F to 8.6F by the year 2100 (United States Environmental Protection Agency, 2016). Assuming 1 atm of pressure, and prorating the temperature growth to the time-horizon considered in our GEP model, we simulate the density of methane by applying a uniform distribution to the temperature within the region of the prorated projected boundaries. We denote the density of methane by Γ .

- The methane leakage rate, λ_{ie} , is the product of A, B_i, C, D , and Γ . Since A and Γ are random variables, we take the average of 100 simulated trials to use as the methane leakage rate in our model.

Multiplying this value by MWh of natural gas or combined-cycle unit generation computes the methane leakage from these units in CO₂e tons. Using the social cost of carbon multiplier (c_y), the costs of methane leakage are determined.

Human Health Externalities:

$$O_H = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I} \gamma_{yti} x_{ytr_1i} \quad (3.6)$$

Human health externalities costs are the costs per MWh of generation for each unit produced. γ_{yti} is the human health cost (\$/MWh) associated with generation from unit i in period t of year y . In a deterministic case, these values can be retrieved from less rigorous approximations from literature, as discussed in Chapter 2. However, in this research we approximate these multipliers via simulation and metamodeling.

Steam Generation Revenue:

$$O_R = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I_{REV}} \phi_{yi} x_{ytr_1i} \quad (3.7)$$

Steam generation revenue is generated from combined heat and power units (also known as co-generation). ϕ_{yi} is the revenue from these units (\$/MWh) and x_{ytr_1i} is the amount of aggregate steam generation dispatched (MWh) from in region r_1 period t of year y .

The total cost objective function (z) is the sum of the NPV of these above costs and steam revenue, and is given by Equation 3.8.

$$z = O_G + O_H + O_C + O_L + O_I + O_M - O_R \quad (3.8)$$

3.2.1.2. Problem Formulation

The mathematical formulation for the GEP model is presented in this section. As discussed previously, the objective is to minimize the total costs on the system subject to system constraints. Model sets and parameters are given in Tables 3.1 and 3.2 respectively.

Table 3.1 – GEP Model Set Definitions

Y	Years in the planning horizon (2015 through 2040) – indexed by y (or u)
T	Periods in the planning horizon (summer – peak/offpeak, winter – peak/offpeak, spring/fall – peak/offpeak – indexed by t)
I	All generating units (nuclear, combined-cycle gas turbine, natural gas turbine, wind (land and offshore), biomass, coal, combined heat and power, solar, petroleum, and hydro)
I*	All existing units within the network – indexed by i
I _R	Renewable units (biomass, wind, and solar) – indexed by i
I _L	Units with methane leak rates considered (only natural gas turbines) – indexed by i
I _{MAX}	Units with construction limits – indexed by i
I _{REV}	Units with steam revenue (combined heat and power only) – indexed by i
I _{ND}	Non-dispatchable units (wind and solar) – indexed by i
R	Regions within the Northeastern US network (New England, NY, NYC, NJ, MD/DE, and Rest of PJM) – indexed by r_1 (or r_2)
R*	Regional Greenhouse Gas Initiative (RGGI) subregions within the Northeastern US network – indexed by r_1 (or r_2)
E	Emissions considered in the GEP model (NO _x , SO ₂ , CO ₂) – indexed by e
E _C	Emissions considered in the GEP model to compute the social cost of carbon (CO ₂) – indexed by e
E _T	Emissions considered in the GEP model for specific time periods (SO ₂) – indexed by e
E _R	Emissions considered in the GEP model for specific regions – indexed by e
E _{MAX}	Emissions considered in the GEP model with maximum limits for all regions (NO _x only) – indexed by e
E*	Emissions considered in the GEP model specific to RGGI regions (CO ₂) – indexed by e
$\Lambda_{(t,e)}$	Set of emissions considered by time period – indexed by (t, e)
$\Omega_{(y,r_1,r_2)}$	Set of transmission lines by year – indexed by (y, r_1, r_2)
$\Phi_{(r_1,r_2)}$	Network of regions for renewable trading – indexed by (r_1, r_2)
$\Theta_{(r_1,i)}$	Set of all renewable units within each region – indexed by (r_1, i)

Table 3.2 – GEP Model Parameter Definitions

$\eta_{tr_1 i}$	Capacity factor for non-dispatchable units	$\pi_{yr_1 i}$	Emissions limits by region (regional limits only)
θ_i	Capacity value by unit	$\pi_{yr_1 i}^*$	Emissions limits by region (time dependent)
$\kappa_{y,i}$	Construction limits for each unit by year	$\psi_{yr_1 i}$	Minimum renewable requirement by region
d_{ytr_1}	Demand by period and region	$\psi_{yr_1}^*$	Total minimum renewable requirement
δ_{ti}	Derating value by season and unit	d_{yi}^*	Peak demand by region and year
$\rho_{i,e}$	Emissions rate by unit	ϕ_{yi}	Steam Revenue (\$/MWh)
$\lambda_{i,e}$	Methane leak rate by unit	κ_i^*	Total construction limits by unit
β_{yi}	Fixed cost (\$/MW)	ξ_{ye}^{RGGI}	Emissions limits by year in RGGI
h_t	Hours in each period	ξ_{ye}	Emissions limits by year in Northeastern US
$q_{r_1 i}$	Initial capacity by unit and region	$\chi_{yr_1 r_2}$	Transmission capacity by year
α_{yi}	Investment costs (\$/MW)	v_{yti}	Variable costs (\$/MWh)
l_t	Transmission losses by season	γ_{yti}	Health costs (\$/MWh)
m_{r_1}	Reserve margin by region	c_y	Carbon costs (\$/ton)
r	Interest rate (3%)		

The generation expansion planning formulation is given as follows:

$$\min z = O_G + O_H + O_C + O_L + O_I + O_M - O_R$$

s.t.

$$\sum_{r_1 \in \Omega(y, r_1, r_2)} l_t w_{ytr_1 r_2} - \sum_{r_1 \in \Omega(y, r_1, r_2)} w_{ytr_2 r_1} + \sum_{i \in I} x_{ytr_1 i} = d_{ytr_1} \quad \forall y \in Y, t \in T, r_2 \in R \quad (3.9)$$

$$x_{ytr_1 i} \leq \left(q_{r_1 i} + \sum_{u(\forall u \leq y) \in Y} y_{ur_1 i} \right) \delta_{t,i} h_t \quad \forall y \in Y, t \in T, r_1 \in R, i \notin I_{ND} \quad (3.10)$$

$$x_{ytr_1 i} \leq \left(q_{r_1 i} + \sum_{u(\forall u \leq y) \in Y} y_{ur_1 i} \right) \delta_{t,i} h_t \eta_{tr_1 i} \quad \forall y \in Y, t \in T, r_1 \in R, i \in I_{ND} \quad (3.11)$$

$$w_{ytr_1 r_2} - w_{ytr_2 r_1} \leq \chi_{yr_1 r_2} h_t \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, t \in T, i \in I \quad (3.12)$$

$$w_{ytr_2 r_1} - w_{ytr_1 r_2} \leq \chi_{yr_2 r_1} h_t \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, t \in T, i \in I \quad (3.13)$$

$$\sum_{i \notin I_{ND}} q_{r_1 i} + \sum_{u(\forall u \leq y) \in Y} \sum_{i \notin I_{ND}} y_{ur_1 i} + \sum_{i \in I_{ND}} q_{r_1 i} \theta_i + \sum_{u(\forall u \leq y) \in Y} \sum_{i \in I_{ND}} y_{ur_1 i} \theta_i \geq d_{yi}^* m_{r_1} \quad \forall y \in Y, r_1 \in R \quad (3.14)$$

$$\sum_{t \in T} \sum_{r_2 \in \Phi(r_1, r_2)} x_{ytr_2 i} \geq \psi_{yr_1 i} \sum_{t \in T} d_{ytr_1} \quad \forall y \in Y, (r_1, i) \in \Theta_{(r_1, i)} \quad (3.15)$$

$$\sum_{t \in T} \sum_{(r_2) \in \Theta_{(r_1, i)}} x_{ytr_2 i} \geq \psi_{yr_1}^* \sum_{t \in T} d_{ytr_1} \quad \forall y \in Y, r_1 \in R \quad (3.16)$$

$$\sum_{t \in T} \sum_{r_1 \in R^*} \sum_{i \in I} x_{ytr_1 i} \rho_{i,e} \leq \xi_{ye}^{RGGI} \quad \forall y \in Y, e \in E^* \quad (3.17)$$

$$\sum_{(t,e) \in \Lambda_{(t,e)}} \sum_{i \in I} x_{ytr_1 i} \rho_{i,e} \leq \pi_{yr_1 i}^* \quad \forall y \in Y, r_1 \in R, e \in E_T \quad (3.18)$$

$$\sum_{t \in T} \sum_{i \in I} x_{ytr_1 i} \rho_{i,e} \leq \pi_{yr_1 i} \quad \forall y \in Y, r_1 \in R, e \in E_R \quad (3.19)$$

$$\sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I} x_{ytr_1 i} \rho_{i,e} \leq \xi_{ye} \quad \forall y \in Y, e \in E \quad (3.20)$$

$$\sum_{r_1 \in R} y_{yr_1 i} \leq \kappa_{y,i} \quad \forall y \in Y, i \in I_{MAX} \quad (3.21)$$

$$\sum_{y \in Y} \sum_{r_1 \in R} y_{yr_1 i} \leq \kappa_i^* \quad i \in I_{MAX} \quad (3.22)$$

$$x_{ytr_1i} \geq 0, w_{ytr_1r_2} \geq 0, y_{yr_1i} \geq 0 \quad \forall (y, r_1, r_2) \in \Omega_{(y, r_1, r_2)}, Y, t \in T, i \in I \quad (3.23)$$

Equation 3.9 is the demand (or energy balance) constraint. For each of the periods in a given year, the total generation from both existing and new generating units should be at least as much as the corresponding demand in that region. Equations 3.10 and 3.11 are related to generating unit capacities for dispatchable and nondispatchable units. Equations 3.12 and 3.13 are the transmission limit equations. Equation 3.14 is the reserve margin constraint. Equations 3.15 and 3.16 specify minimum renewable levels for electricity dispatching. Equations 3.17 through 3.20 are the emissions constraints. The set of all emissions considered includes NO_x, SO₂, and CO₂. Equation 3.21 and 3.22 are investment limit constraints on an annual basis and throughout the time horizon of the model respectively. Equation 3.23 specifies non-negativity for generation, transmission, and investment decision variables.

3.2.2. COBRA Overview

The Co-Benefits Risk Assessment (COBRA) tool is a EPA screening model that helps state and local governments approximate air quality, human health implications, and the economic benefits of clean energy policies and programs (United States Environmental Protection Agency, 2015). The outcomes from this model are approximated based on changes in particulate matter (PM_{2.5}), SO₂, NO_x, NH₃, and volatile organic compounds (VOCs) at the county, state, regional, or national level.

In practice, COBRA users generate scenarios by specifying changes in the emissions estimates for the analysis year at the appropriate regional level. The model then quantifies the associated changes in changes in PM_{2.5} concentrations between a baseline scenario and an experimental (or control) scenario. A source-receptor matrix

translates the air pollution emissions changes into changes in ambient $PM_{2.5}$. Using linear regression models, health impact functions, COBRA utilizes the ambient $PM_{2.5}$ changes to quantify changes in the incidence of human health effects. Finally, COBRA assigns a dollar value to these health effects. COBRA estimates the change in air pollution-related health impacts, and estimates the economic value of these impacts, using an approach that is consistent with EPA Regulatory Impact Analyses, and reflects the current state of the science regarding the relationship between particulate matter and adverse human health effects (United States Environmental Protection Agency, 2015).

The basic tools and data sources that the COBRA model employs to determine emissions reductions are given in Table 3.3.

Table 3.3 – COBRA Tools and Data Sources (United States Environmental Protection Agency, 2015)

Online Tool	Description
EPA's Emissions & Generation Resource Integrated Database (eGrid): http://www.epa.gov/cleanenergy/energyresources/egrid/index.html	Provides data on the environmental characteristics of electric generation by power plants in the United States.
EPA's AVOIDed Emissions and geneRATION Tool (AVERT): http://epa.gov/avert/	Estimates displaced emissions (at the county, state, and regional levels) at electric power plants due to renewable energy or energy efficiency policies and programs.
eCalc: http://ecalc.tamu.edu/	Uses both energy and emissions modeling to determine emission reductions from energy efficiency and renewable energy programs in the Electric Reliability Council of Texas region.
EPA's Motor Vehicle Emission Simulator (MOVES): http://www.epa.gov/otaq/models/moves/index.htm	Estimates emissions from mobile sources, including emissions from cars, trucks, and motorcycles.
National Emissions Inventory: http://www.epa.gov/ttnchie1/net/2008inventory.html	Allows users to view emissions by sector (for 60 emissions inventory sectors) for specific pollutants at varying levels of geographic aggregation.
OTC Workbook: http://www.otcair.org	Predicts emission reductions from energy portfolio policies and energy efficiency programs and other measures affecting renewable resources or multiple pollutants
Power Profiler: http://www.epa.gov/powerprofiler/	Allows users to view the emissions that can be attributed to electricity use in homes or businesses

The COBRA model can be adjusted to study the impact of pollutant emissions changes at the national level or for smaller geographic areas. Based on the changes in ambient PM_{2.5} concentrations, COBRA predicts the associated changes in the number of cases of a variety of health endpoints including the following (United States Environmental Protection Agency, 2015):

- Adult and infant mortality;
- Non-fatal heart attacks;
- Respiratory-related and cardiovascular-related hospitalizations;
- Acute bronchitis;
- Upper and lower respiratory symptoms;
- Asthma-related emergency room visits;
- Asthma exacerbations;
- Minor restricted activity days; and
- Work days lost due to illness.

The economic impact of each of these health impacts is approximated by using industry standard response functions. The response functions and assumptions for each health endpoint are given in Table 3.4. These functions are the as follows:

- Value of Statistical Life (VSL);
- Cost of Illness (COI);
- Opportunity Cost (OC); and
- Willingness to Pay (WTP).

Table 3.4 – Canned Response Functions (United States Environmental Protection Agency, 2015)

Health Endpoint	Canned Response Function
Adult Mortality	Value of Statistical Life (VSL)
Infant Mortality	Value of Statistical Life (VSL)
Non-Fatal Heart Attacks	<ul style="list-style-type: none"> • Cost of Illness (COI) = Direct Medical Costs • Opportunity Costs (OC)
Hospital Admissions	<ul style="list-style-type: none"> • COI = Hospital Charges • OC

Asthma ER Visits	COI = Hospital Costs
Acute Bronchitis	Willingness to Pay (WTP) = Coughing and chest tightness/restricted activity day
Respiratory Symptoms	WTP = Coughing, head/sinus congestion, eye irritation, chest tightness, coughing up phlegm, and/or wheezing
Asthma Exacerbations	WTP = Bad asthma day
Minor Restricted Activity Days	WTP = Combination of coughing, throat congestion, and sinusitis
Work Loss Days	WTP = Median annual earnings \div # of working days (52 * 5)

COBRA is a very user-friendly tool that allows researchers the flexibility to study various emissions reduction (or increase) scenarios quickly and visualize the data. However, since COBRA is a screening tool based on external inputs and linear regression models, there is the potential for oversimplification (United States Environmental Protection Agency, 2015). Furthermore, since the generation expansion planning problem assesses investment and dispatching decisions in aggregate over a long-term planning horizon, COBRA is the ideal screening tool for this application.

3.2.3. Simulation and Experimentation Approach for Seed Generation

In order to create a metamodel approximating human health externalities in terms of electricity generation, we must have a diverse sample space of data points from our expansion-planning model, which are then evaluated in the COBRA tool. This is accomplished via the development of structured experiments that are driven by generation expansion planning simulations. The approach we follow in this research is outlined in Figure 3.3.

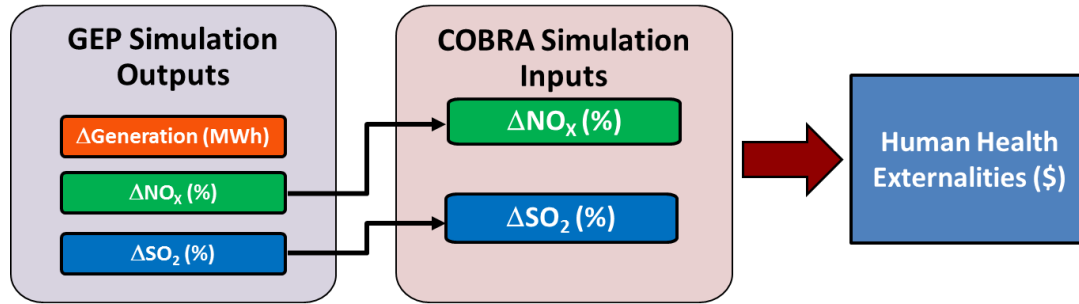


Figure 3.3 – Schematic Flow Diagram of GEP & COBRA Experiments

The first step in this process is to solve a baseline generation expansion planning model, as outlined in Section 3.2.1., that minimizes total costs, excluding human health externalities, the social cost of carbon, and the social cost of methane leakage. The rationale behind excluding these costs in the baseline model is driven by obtaining a diverse set of data points (or dispatching decision) in the expansion plan.

Secondly, we solve experimental generation expansion planning models, similar to the baseline model, where we minimize total costs, excluding human health externalities, the social cost of carbon, and the cost of methane leakage. However, the major difference here is that we vary several key parameters in the model by simulating from a normal distribution with a coefficient of variation following a uniform random variable on the interval [0,1]. The parameters in the GEP model we vary are as follows:

- Demand;
- Peak demand;
- Reserve margin;
- Minimum and maximum limits;
- Minimum renewable generation levels;
- Emissions rates;

- Emissions limits;
- Costs (Investment, Fixed, and Variable);
- Steam revenue;
- Transmission capacity;
- Capacity factors for non-dispatchable units; and
- Derating values.

For each experimental generation expansion planning trial, we compute three values for every year in the time horizon.

$$\Delta x_{y,i} = \sum_{t \in T} (x_{ytr_1i,experimental}^* - x_{ytr_1i,baseline}^*) \quad (3.24)$$

$$\Delta NO_{Xy,r_1}(\%) = \frac{NO_{Xy,experimentl} - NO_{Xy,baseline}}{NO_{Xy,baseline}} \quad (3.25)$$

$$\Delta SO_{2y,r_1}(\%) = \frac{SO_{2y,experimentl} - SO_{2y,baseline}}{SO_{2y,baseline}} \quad (3.26)$$





















Equation 3.24 quantifies the deviation in electricity dispatch in comparing the optimal generation from an experimental trial to the baseline trial. This value is determined on an annual basis for each generating unit and region in the network, and has units of MWh. Equations 3.25 and 3.26 quantify the percent deviation in NO_X and SO_2 respectively in comparing an experimental trial to the baseline trial. These values are determined on an annual basis for each region, and are inputs to the COBRA model to determine the associated human health externalities. Ultimately, the end deliverable of these simulated experiments is a data set that is used to develop a metamodel, which predicts human health externalities as a function of electricity dispatch decisions at unobserved locations.

The data set resembles Figure 3.2, and the resulting metamodel serves as a surrogate objective function in a generation expansion-planning model.

3.3. Constructing a Metamodel of Human Health Externalities

As evidenced by Sections 3.1 and 3.2, quantifying human health externalities as a function of electricity dispatch decisions is a computationally expensive process. In order to solve the generation expansion planning problem efficiently and effectively, we need to develop a model that explicitly quantifies human health externalities as a function of dispatching decisions, while capturing the highly nonlinear relationship between electricity generation, pollutant emissions, air quality, and human health externalities. This process is commonly referred to as metamodeling.

As summarized in Chapter 2 of this dissertation, there are various approaches that can be used in metamodeling. The most common tools are response surfaces, radial basis functions, neural networks, and kriging. The schematic diagram in Figure 3.4 qualitatively compares each of these metamodeling approaches within the context of this research.

	Response Surfaces	Radial Basis Functions	Neural Networks	Kriging
Accommodates Nonlinearity				
Returns the exact response for an observed point				
Treatment of clustered observations				
Model Training Requirements				
Computational Burden				

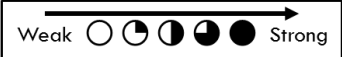


Figure 3.4 – Qualitative Comparison of Metamodeling Methods

We considered five key criteria to compare metamodeling methods: (i) the ability to accommodate nonlinear problems, (ii) the ability to return an exact response for an observed point, (iii) the treatment of clustered observations, (iv) training requirements, and (v) the computational burden.

In qualitatively assessing each of these methods, we determined that kriging is the most suitable approach for this application. Since kriging is an optimal interpolation-based method that is based on regression against observed responses and surrounding data points (which are weighted according to spatial covariance value), it accommodates nonlinearity and treats clustered observations effectively. Furthermore, for an observed location, kriging models return the observed response. Additionally, these models require no training, and are able to generate results quickly and efficiently with a high degree of accuracy.

3.3.1. Kriging Metamodel

The kriging metamodel we utilize to quantify human health externalities as a function of electricity generation is given in equation 3.27.

$$\hat{\gamma}_{yi}(\Delta \mathbf{x}_y) = \mu(\Delta \mathbf{x}_y) + \sum_{k=1}^{n(\Delta \mathbf{x}_y)} \omega_k [\gamma_{yi}(\Delta \mathbf{x}_{yk}) - \mu(\Delta \mathbf{x}_{yk})] \quad (3.27)$$

In equation 3.27, $\hat{\gamma}_{yi}(\Delta \mathbf{x}_y)$ is the predicted value of human health externalities in year y for unit i , which is a function of a vector of unobserved dispatch values, $\Delta \mathbf{x}_y$. $\Delta \mathbf{x}_{yk}$ is a vector of observed dispatch deviations obtained from the procedure outlined in Section 3.2.3, indexed by k . The observed value of $\Delta \mathbf{x}_{yk}$ is given by $\gamma_{yi}(\Delta \mathbf{x}_{yk})$. $n(\Delta \mathbf{x}_y)$ is the number of nearest neighbors to consider in the model, and ω_k is the kriging weight,

which is derived from the residuals (see Chapter 2, Section 2.4.2). The means of the predicted response and observed response are given by $\mu(\Delta\mathbf{x}_y)$ and $\mu(\Delta\mathbf{x}_{yk})$ respectively.

Residuals, $R(\Delta\mathbf{x}_{yk})$, in a kriging model have a stationary mean and stationary covariance. That is, $R(\Delta\mathbf{x}_{yk}) = \gamma_{yi}(\Delta\mathbf{x}_{yk}) - \mu(\Delta\mathbf{x}_{yk})$, with $E[R(\Delta\mathbf{x}_{yk})] = 0$ and $Cov\{R(\Delta\mathbf{x}_{yk}), R(\Delta\mathbf{x}_{yk} + h)\} = C(h)$ for some lag h . $C(h) = C(0) + SV(h)$, where $C(0)$ is defined as the sill and $SV(h)$ is defined as the semivariogram. For this particular application, we employ the Gaussian semivariogram in Equation 3.28.

$$SV(h) = (sill - nugget) \times \left(1 - \exp\left(\frac{-3h^2}{range^2}\right)\right) + nugget \quad (3.28)$$

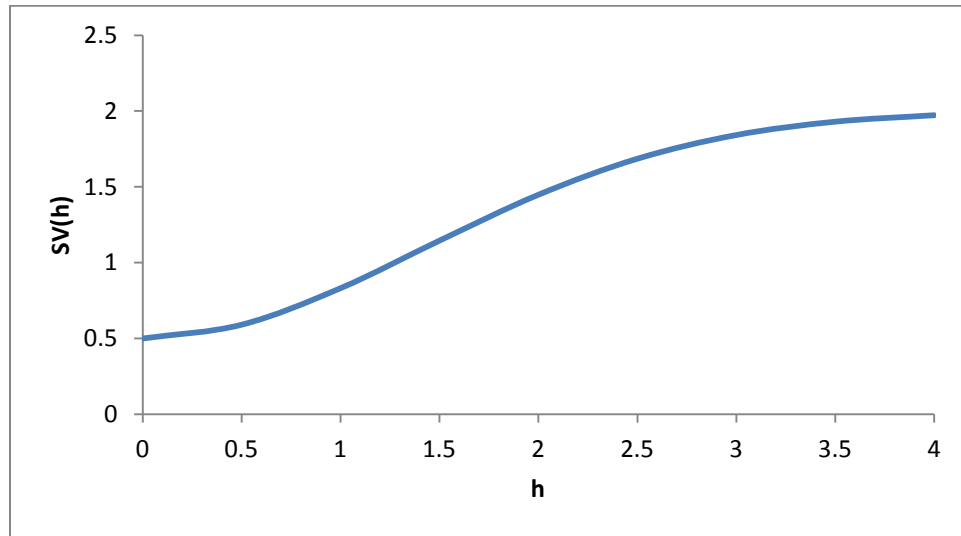


Figure 3.5 – Gaussian Semivariogram Example

The Gaussian semivariogram function has three parameters that the user can specify – sill, nugget, and range. From the graphical representation in Figure 3.5, the sill is stationary value of the semivariogram, when the slopes of the tangent lines become

zero. Similarly, the range is the lag value when the slope of the tangent lines becomes zero. The nugget is the value of the semivariogram function with zero lag.

3.4. Implementing a Metamodel in the Objective Function

Upon successful creation of a metamodel based on the process outlined in Section 3.3, we can now proceed with solving the generation expansion planning problem with the inclusion of human health externalities. In the optimization formulation we present in Section 3.2.1., the deterministic objective function is given by $z = O_G + O_H + O_C + O_L + O_I + O_M - O_R$, where all of these costs are considered to be deterministic. However, with a fully constructed metamodel, we replace the human health externalities, O_H , with a surrogate, \hat{O}_H , where \hat{O}_H is given in Equation 3.29.

$$\hat{O}_H = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{t \in T} \sum_{r_1 \in R} \sum_{i \in I} \hat{\gamma}_{yi}(\Delta \mathbf{x}_y) = x_{ytr_1i} \quad (3.29)$$

Since a portion of the objective function in this generation expansion planning problem is given by a metamodel, we cannot apply solve this problem using traditional linear programming methods from the simplex algorithm family. However, since our metamodel will yield a smooth surface, we can consider alternative heuristics like search algorithms or gradient-based approaches. For this particular problem, we apply a multistart heuristic algorithm. The steps in the algorithm are given below (GAMS Development Corporation, 2014).

- Step 1: Apply a scatter search algorithm an initial set of candidate points from the population in the feasible region. This search algorithm selects these candidate points from the population using a crossover operator – a feature that is used in evolutionary programming methods such as genetic algorithms.

- Step 2: The algorithm then calls a gradient-based nonlinear programming (NLP) solver, which takes the candidate points from Step 1, and executes optimization routines from each point.
- Step 3: The NLP solver tracks all feasible solutions obtained in Step 2, and reports the best of these as its final solution.

3.4.1. Design Criteria

Upon solving the generation expansion planning problem with the multistart heuristic algorithm, we now have a candidate expansion plan that minimizes human health externalities along with total system costs. However, given that the surrogate for human health externalities is a metamodel, it is likely that our solution does not adequately account for metamodeling prediction error. To address this deficiency, we solve an expected improvement problem, which determines candidate dispatch plans to compare against our baseline GEP model from Section 3.2.3., and evaluate in the COBRA tool. Upon obtaining human health externalities from COBRA, this data set is then appended to the experimental data set used to create the metamodel. Equation 3.30 defines the expected improvement function, $E[I(\mathbf{x})]$.

$$E[I(\mathbf{x})] = \begin{cases} (O_{H,min} - \hat{O}_H(\mathbf{x})) \Phi\left(\frac{O_{H,min} - \hat{O}_H(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}(\mathbf{x}) \phi\left(\frac{O_{H,min} - \hat{O}_H(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) & \text{if } \hat{s} \geq 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \quad (3.30)$$

From Equation 3.30, \mathbf{x} is a vector of dispatching decisions, x_{ytr_1i} , as previously defined in Section 3.2.1.2. $O_{H,min}$ is the minimum value of the total human health externalities from all GEP model trials. $\hat{O}_H(\mathbf{x})$ is the current metamodel that predicts human health externalities as a function of the dispatch vector. \hat{s} is the model error (see

Section 2.4.3). The formulation for the expected improvement optimization problem is given as follows:

$$\begin{aligned} \max z &= E[I(\mathbf{x})] \\ \text{s.t.} \end{aligned}$$

$$\sum_{r_1 \in \Omega(y, r_1, r_2)} l_t w_{ytr_1 r_2} - \sum_{r_1 \in \Omega(y, r_1, r_2)} w_{ytr_2 r_1} + \sum_{i \in I} x_{ytr_1 i} = d_{ytr_1} \quad \forall y \in Y, t \in T, r_2 \in R \quad (3.31)$$

$$w_{ytr_1 r_2} - w_{ytr_2 r_1} \leq \chi_{yr_1 r_2} h_t \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, t \in T, i \in I \quad (3.32)$$

$$w_{ytr_2 r_1} - w_{ytr_1 r_2} \leq \chi_{yr_2 r_1} h_t \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, t \in T, i \in I \quad (3.32)$$

$$x_{ytr_1 i} \geq 0, w_{ytr_1 r_2} \geq 0 \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, t \in T, i \in I \quad (3.33)$$

The goal is to determine the dispatch plan that maximizes the expected improvement of the metamodel. From the presented formulation, Equation 3.31 is the energy balance constraint from our baseline GEP model. Similarly, Equations 3.32 and 3.33 are the transmission constraints from the baseline GEP model. The decision variables in this model are $x_{ytr_1 i}$ and $w_{ytr_1 r_2}$, which are the respective dispatching and transmission decision variables outlined in Section 3.3.1.2. To solve this problem, we apply the previously defined multistart heuristic algorithm.

3.4.2. Termination Criteria

The procedures outlined in this Chapter of the dissertation detail the analytical framework for solving the generation expansion planning problem with a metamodel serving as a surrogate function for human health externalities. As discussed at the beginning of this chapter, we follow the algorithm outlined in this Chapter iteratively until we reach termination criteria. We terminate our algorithm under two circumstances.

- Case 1: When the expected improvement model reaches zero, we terminate our model. This occurs when the metamodel is error free, which is highly unlikely in this problem. This typically occurs in smaller optimization problems with a limited feasible region.
- Case 2: From iteration to iteration, if the objective function changes within a pre-specified tolerance range, we terminate our model. This occurs when surface of the metamodel approaches the true surface, but has no or minimal impact on the outcome of the GEP optimization model.

4. Reduced-Form Model Evaluation

In this section, we apply our modeling approach to a series of scenarios to evaluate (i) the impact of constraints on the solution, (ii) the impact of minimizing human health externalities and the social cost of carbon and methane leakage, and (iii) a full evaluation of the approach outlined in Chapter 3. For the numerical examples presented in this Chapter, we consider the Northeastern U.S., as shown in the schematic representation in Figure 4.1.

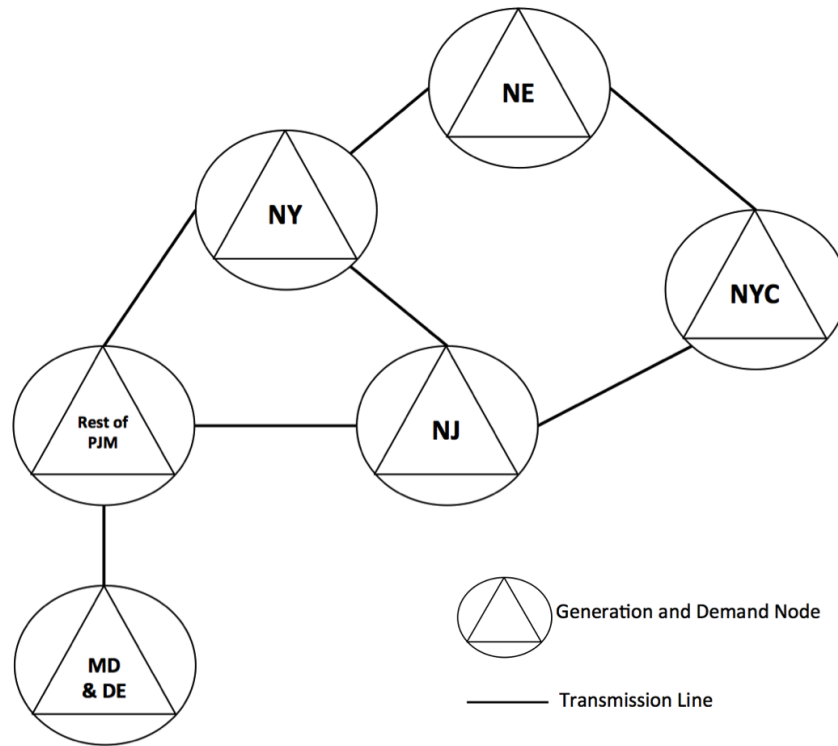


Figure 4.1 – Transmission Network of the Northeastern United States

The network shown in Figure 4.1 has 6 regions, with the notation given as follows:

- NE: New England (Maine, Vermont, New Hampshire, Massachusetts, Connecticut and Rhode Island);
- NY: New York State (excluding New York City);

- NYC: New York City;
- NJ: New Jersey;
- MD & DE: Maryland, Delaware, and the District of Columbia;
- Rest of PJM: Illinois, Indiana, Kentucky, Michigan, North Carolina, Ohio, Tennessee, Virginia, and West Virginia.

Each region (or node) in the network is a source of demand, and also is capable of generating electricity. Additionally, the configuration defines how electricity is generated and transmitted within the network. Within each region, the generating unit technologies considered are combined cycle gas turbines, coal, natural gas turbines, hydro, nuclear, petroleum, solar, biomass, on-shore wind, and off-shore wind.

In order to evaluate the performance of our model, we solve multiple GEP model scenarios in the Northeastern U.S. region; however, for the numerical examples in this Chapter, we solve a reduced-form of the GEP model outlined in Chapter 3 by removing the seasonality element and reducing the time-horizon to 10-years (2015 through 2025). Full details of the reduced-form model are presented in Section 4.1 of this Chapter.

4.1. Reduced-Form Model Formulation

The mathematical formulation for the reduced-form GEP model is presented in this section. As discussed previously, the objective is to minimize the total costs on the system subject to system constraints. For the reduced-form model, we modify the objective function by removing the seasonality index, t , from the formulation presented in Chapter 3. All future costs are discounted to the present time using a discount rate $r=3\%$ to yield a net present value (NPV), and to also align with the United States Energy Information Association's assumptions presented in the 2015 Annual Energy Outlook (U.S. Energy Information Administration, 2015). Applying these assumptions, the costs in the objective function for the test model are now given as follows:

$$O_I = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I} \alpha_{yi} y_{yr_1 i} \quad (4.1)$$

$$O_M = \sum_{y \in Y} \frac{1}{(1+r)^y} \left[\sum_{r_1 \in R} \sum_{i \in I} \left(\sum_{u(\forall u \leq y) \in Y} \beta_{ui} y_{ur_1 i} \right) + \sum_{r_1 \in R} \sum_{i \in I^*} \beta_{yi} q_{r_1 i} \right] \quad (4.2)$$

$$O_G = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I} v_{yi} x_{yr_1 i} \quad (4.3)$$

$$O_C = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I} \sum_{e \in E_c} c_y x_{yr_1 i} \rho_{i,e} \quad (4.4)$$

$$O_L = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I_L} \sum_{e \in E_c} c_y x_{yr_1 i} \lambda_{i,e} \quad (4.5)$$

$$O_H = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I} \gamma_{yi} x_{yr_1 i} \quad (4.6)$$

$$O_R = \sum_{y \in Y} \frac{1}{(1+r)^y} \sum_{r_1 \in R} \sum_{i \in I_{REV}} \phi_{yi} x_{yr_1 i} \quad (4.7)$$

Equations 4.1 through 4.7 follow the same definitions as given in Chapter 3, with the exception that the seasonality index, t , is not considered in the test model. The total cost objective function (z) is the sum of the net present value (NPV) of these above costs and

steam revenue, and is given by Equation 4.8, which analogous to Equation 3.8 in Chapter 3.

$$z = O_G + O_H + O_C + O_L + O_I + O_{OM} - O_R \quad (4.8)$$

The full formulation of the generation expansion planning test problem is given as follows:

$$\begin{aligned} \min z &= O_G + O_H + O_C + O_L + O_I + O_F - O_R \\ \text{s.t.} \end{aligned}$$

$$\begin{aligned} \sum_{r_1 \in \Omega(y, r_1, r_2)} l_t w_{yr_1 r_2} - \sum_{r_1 \in \Omega(y, r_1, r_2)} w_{yr_2 r_1} \\ + \sum_{i \in I} x_{ytr_1 i} = d_{yr_1} \end{aligned} \quad \forall y \in Y, r_2 \in R \quad (4.9)$$

$$x_{yr_1 i} \leq \left(q_{r_1 i} + \sum_{u(\forall u \leq y) \in Y} y_{ur_1 i} \right) \delta_i h \quad \forall y \in Y, r_1 \in R, i \notin I_{ND} \quad (4.10)$$

$$x_{yr_1 i} \leq \left(q_{r_1 i} + \sum_{u(\forall u \leq y) \in Y} y_{ur_1 i} \right) \delta_i h \eta_{r_1 i} \quad \forall y \in Y, r_1 \in R, i \in I_{ND} \quad (4.11)$$

$$w_{yr_1 r_2} - w_{yr_2 r_1} \leq \chi_{yr_1 r_2} h \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, i \in I \quad (4.12)$$

$$w_{ytr_2 r_1} - w_{ytr_1 r_2} \leq \chi_{yr_2 r_1} h \quad \forall (y, r_1, r_2) \in \Omega(y, r_1, r_2), Y, i \in I \quad (4.13)$$

$$\begin{aligned} \sum_{i \notin I_{ND}} q_{r_1 i} + \sum_{u(\forall u \leq y) \in Y} \sum_{i \notin I_{ND}} y_{ur_1 i} \\ + \sum_{i \in I_{ND}} q_{r_1 i} \theta_i \\ + \sum_{u(\forall u \leq y) \in Y} \sum_{i \in I_{ND}} y_{ur_1 i} \theta_i \\ \geq d_{yi}^* m_{r_1} \end{aligned} \quad \forall y \in Y, r_1 \in R \quad (4.14)$$

$$\begin{aligned} \sum_{r_2 \in \Phi(r_1, r_2)} x_{yr_2 i} \geq \psi_{yr_1 i} \sum_{t \in T} d_{yr_1} \\ \sum_{(r_2) \in \Theta(r_1, i)} x_{yr_2 i} \geq \psi_{yr_1 i}^* \sum_{t \in T} d_{yr_1} \end{aligned} \quad \forall y \in Y, (r_1, i) \in \Theta(r_1, i) \quad (4.15)$$

$$\sum_{r_2 \in \Phi(r_1, r_2)} x_{yr_2 i} \geq \psi_{yr_1 i}^* \sum_{t \in T} d_{yr_1} \quad \forall y \in Y, r_1 \in R \quad (4.16)$$

$$\sum_{r_1 \in R^*} \sum_{i \in I} x_{yr_1 i} \rho_{i, e} \leq \xi_{ye}^{RGGI} \quad \forall y \in Y, e \in E^* \quad (4.17)$$

$$\sum_{i \in I} x_{yr_1 i} \rho_{i,e} \leq \pi_{yr_1 i} \quad \forall y \in Y, r_1 \in R, e \in E_R \quad (4.18)$$

$$\sum_{r_1 \in R} \sum_{i \in I} x_{yr_1 i} \rho_{i,e} \leq \xi_{ye} \quad \forall y \in Y, e \in E \quad (4.19)$$

$$\sum_{r_1 \in R} y_{yr_1 i} \leq \kappa_{y,i} \quad \forall y \in Y, i \in I_{MAX} \quad (4.20)$$

$$\sum_{y \in Y} \sum_{r_1 \in R} y_{yr_1 i} \leq \kappa_i^* \quad i \in I_{MAX} \quad (4.21)$$

$$x_{yr_1 i} \geq 0, w_{yr_1 r_2} \geq 0, y_{yr_1 i} \geq 0 \quad \forall (y, r_1, r_2) \in \Omega_{(y, r_1, r_2)}, Y, t \in T, i \in I \quad (4.22)$$

Equation 4.9 is the demand (or energy balance) constraint. For each year in the test model, the total generation from both existing and new generating units should be at least as much as the corresponding demand in that region. Equations 4.10 and 4.11 are related to generating unit capacities for dispatchable and nondispatchable units. Equations 4.12 and 4.13 are the transmission limit equations. Equation 4.14 is the reserve margin constraint. Equations 4.15 and 4.16 specify minimum renewable levels for electricity dispatching. Equations 4.17 through 4.19 are the emissions constraints. The set of all emissions considered in the test model includes NO_x, SO₂, and CO₂. Equation 4.20 and 4.21 are investment limit constraints on an annual basis and throughout the time horizon of the model respectively. Equation 4.22 specifies non-negativity for generation, transmission, and investment decision variables, given by $x_{yr_1 i}$, $w_{yr_1 r_2}$, and $y_{yr_1 i}$ respectively.

4.2. Reduced-Form Test Case Definitions and Assumptions

In this chapter, we evaluate six (6) test cases as presented in Table 4.3. For Cases 1.1 through 1.5, we apply a CPLEX linear programming (LP) solver to obtain an optimal solution. However, for Case 1.6, we apply the simulation-based optimization approach discussed in Chapter 3. Capacity, costs, and emissions rates for the reduced-form

generation expansion-planning model are given in Table 4.4. We apply these assumptions to the network shown in Figure 4.1 in order to solve the test generation expansion-planning problem.

Table 4.1 – Reduced-Form Model Test Cases

		Objective Function Costs						Constraints					
		Market Costs			Health and Other Damages								
		Investment	Fixed O&M	Variable O&M (including fuel)	Social Cost of CO ₂ and CH ₄ Leakage	Deterministic NO _x and SO ₂ Multipliers	Simulated Human Health Externalities (COBRA)	Energy Balance	Capacity	Construction Limits	Reserve Margin	Emissions Limits	Renewable Portfolio Standards (RPS)
Case 1.1	Baseline	✓	✓	✓				✓	✓	✓			
Case 1.2	Reserve Margin	✓	✓	✓				✓	✓	✓	✓		
Case 1.3	Emissions & RPS Constraints	✓	✓	✓				✓	✓	✓	✓	✓	✓
Case 1.4	Social Cost of CO ₂ and CH ₄ Leakage	✓	✓	✓	✓			✓	✓	✓	✓		
Case 1.5	Deterministic Health Damages	✓	✓	✓	✓	✓		✓	✓	✓	✓		
Case 1.6	Simulated Health Damages	✓	✓	✓	✓		✓	✓	✓	✓	✓		

Table 4.2 – Generating Units and Their Costs, Capacities, and Emissions Rates

	¹ Aggregate Capacity (MW)	² Investment Costs (\$/MW)	² Maintenance Costs (\$/MW)	² Variable (including Fuel) Costs (\$/MWh)	² Fuel Costs (\$/MWh)	² Steam Revenue (\$/MWh)	³ Human Health Externalities (\$/MWh)	⁴ Social Cost of CO ₂ (\$/MT of CO ₂ emitted)	⁵ CO ₂ (lbs/MWh)	⁵ NO _x (lbs/MWh)	⁵ SO ₂ (lbs/MWh)
Biomass	0	\$4,359,382	\$74,627	\$30.29	\$0.00	\$0.00	\$0.00	\$36	0	4	10
Combined Cycle Gas Turbine	21,828	\$1,096,358	\$13,546	\$69.37	\$15.69	\$0.00	\$4.28		1,135	2	0
Combined Heat & Power	0	\$1,744,772	\$18,563	\$75.86	\$14.30	\$56.63	\$0.00		1,135	2	0
Coal	59,968	\$2,713,713	\$38,327	\$25.75	\$12.91	\$0.00	\$44.58		2,249	6	13
Natural Gas Turbine	16,172	\$733,926	\$12,198	\$50.07	\$15.69	\$0.00	\$4.28		1,135	2	0
Hydro	24,082	\$2,909,653	\$12,198	\$0.00	\$0.00	\$0.00	\$0.00		0	0	0
Nuclear	27,799	\$4,326,537	\$104,245	\$6.35	\$5.21	\$0.00	\$2.54		0	0	0
Petroleum	4,703	\$1,114,480	\$14,452	\$99.31	\$64.87	\$0.00	\$9.27		1,672	4	12
Solar	446	\$6,989,283	\$13,523	\$0.00	\$0.00	\$0.00	\$1.50		0	0	0
Off-Shore Wind	0	\$4,459,051	\$98,446	\$0.00	\$0.00	\$0.00	\$0.00		0	0	0
Wind On Land	1,985	\$2,226,694	\$35,088	\$0.00	\$0.00	\$0.00	\$0.00		0	0	0

Notes: ¹ (Nuclear Energy Institute, 2012)
² (U.S. Energy Information Administration, 2015)
³ (National Research Council of the National Academies, 2010)
⁴ (U.S. Environmental Protection Agency, 2016)
⁵ (U.S. Environmental Protection Agency, 2013)

In Case 1.6, where we apply our proposed simulation-based optimization framework to the network, we vary the key model parameters, as identified in Section 3.2.3 in Chapter 3, by simulating from a normal distribution with a coefficient of variation uniformly distribution on the interval $[0,1]$. Next, we execute multiple trials of our experimental GEP model and compare the results to the baseline GEP model, Case 1.1, to obtain changes in investment and dispatching decisions by unit and changes in NO_x and SO_2 emissions. Upon obtain results from the simulations, we then input the changes in emissions into the COBRA model to quantify the associated human health externalities for each trial. Once we generate the full data set, we now fit a kriging metamodel, as outlined in Section 3.3.1 of Chapter 3, that predicts human health externalities as a function of dispatching decisions for each unit in the network. This metamodel serves as a surrogate curve for human health externalities in the objective function of our model.

We test the validity and robustness of our metamodel by applying a modified K-fold cross-validation procedure as follows:

- Step 1: Randomly partition the data set obtained by simulation into eleven (11) equal subsets. Designate one of these subsets to be a testing set, and the remaining ten sets (or validation sets) to fit our metamodels.
- Step 2: Using the validation and test sets identified in Step 1, fit kriging metamodels with a Gaussian variogram.
- Step 3: For each validation set, determine the model error against the test set. The metric used for model error in this procedure is the mean absolute prediction error.

- Step 4: Aggregate the all of the validation sets into a single data set. Use this set to build a full metamodel of the test set data, and calculate the model error as described in Step 3.

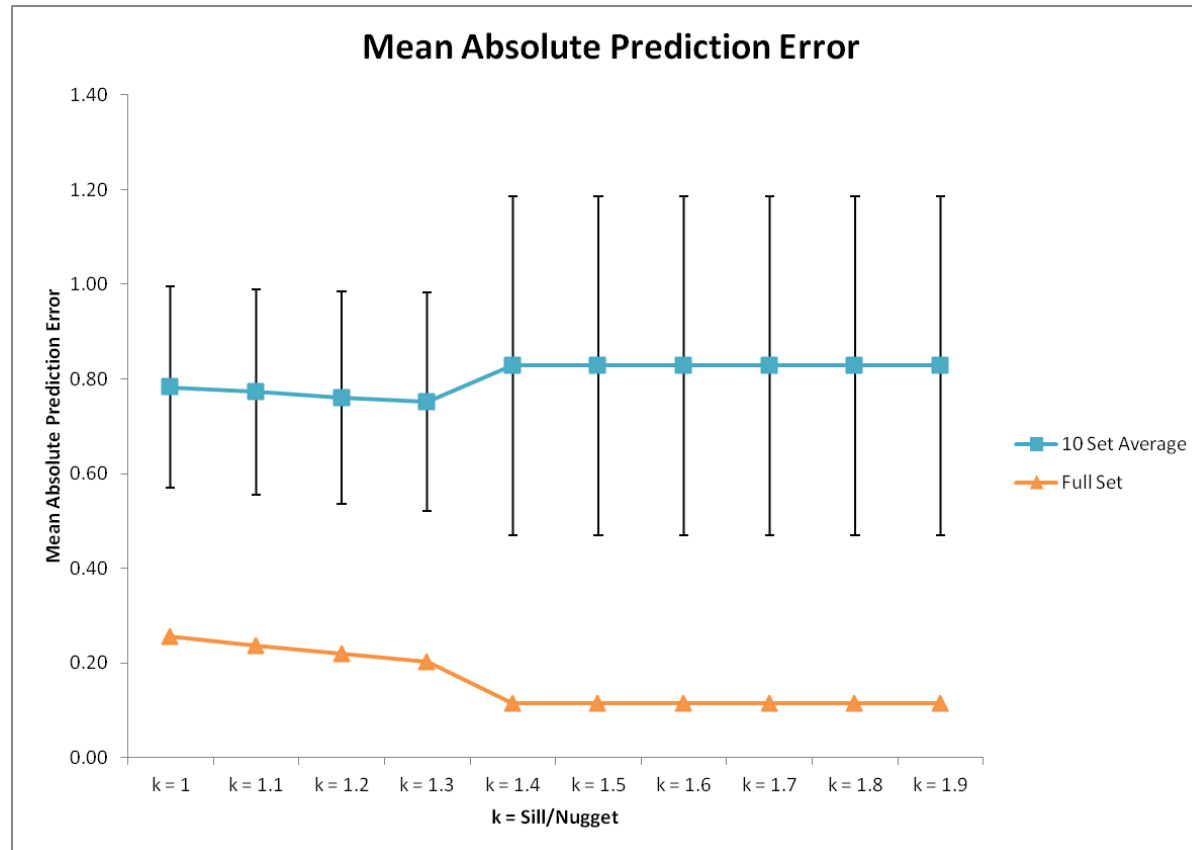


Figure 4.2 – Mean Absolute Prediction Error Values from Cross Validation Procedure

As evidenced by Table 4.3 and Figure 4.2, we execute our modified K-fold cross-validation procedure for various sill and nugget parameters in the Gaussian variogram. In this case, we denote the variable k as the ratio of the sill to the nugget parameters. Based on the individual validation sets, on average, as we increase the k -ratio, the prediction error increases minimally. Additionally, the error variability (or standard deviation) begins to increase nominally after k -ratios greater than 1.4. Upon applying the full data set to fit our metamodel, however, we notice that the prediction error is significantly reduced. In general, the more exhaustive and expansive the sample space, the model will exhibit correspondingly greater prediction properties. From a qualitative perspective, our metamodel is robust for all values of k . Thus, we select $k=1.5$ as our parameter to apply to the variogram. Using this metamodel as a surrogate in the objective function of our GEP test model, we apply the full framework in Chapter 3 to determine an optimal (or near-optimal) expansion plan.

4.3. Dispatching Results

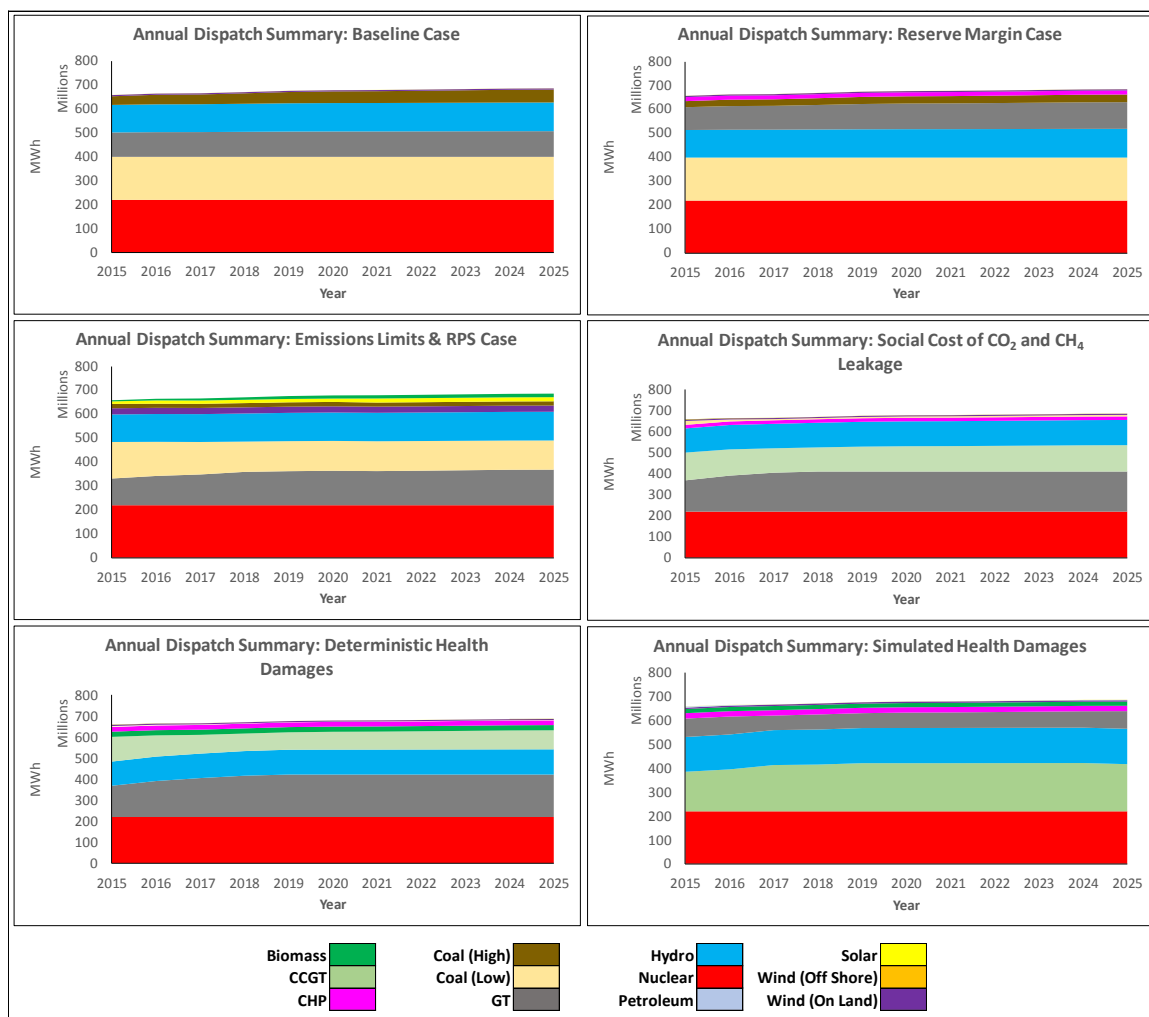


Figure 4.3 – Annual Dispatch Summary Dashboard

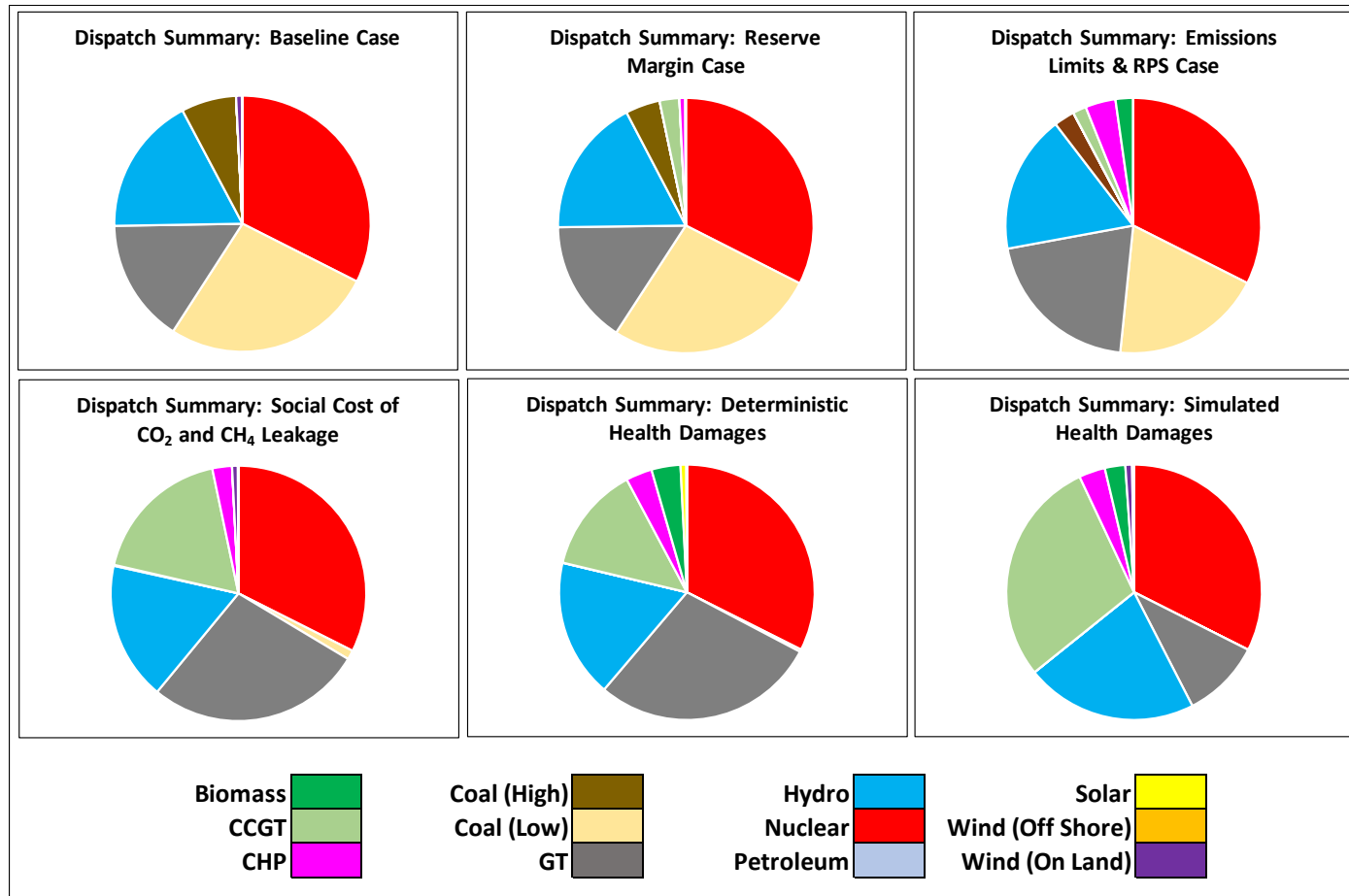


Figure 4.4 – Total Dispatch Percentage Dashboard

Table 4.4 – Regional Dispatch Summary

Case	Region	Dispatch Summary by Region (10 ⁶ MWh)											Totals
		Nuclear	CCGT	GT	Hydro	Coal (Low)	Coal (High)	CHP	Biomass	Wind (On Land)	Solar	Petroleum	
Baseline	MDDE	515.0	-	-	229.4	279.4	46.7	-	-	-	1.6	-	1,072.1
	NE	248.9	-	173.5	788.3	197.3	170.3	-	-	0.0	3.6	-	1,581.9
	NJ	356.9	-	389.7	25.4	101.5	94.1	-	-	0.1	0.2	-	967.8
	NY	457.7	-	24.0	144.9	364.8	4.5	-	-	36.1	-	-	1,032.1
	NYC	-	-	285.5	-	634.7	-	-	-	-	-	-	920.2
	RoPJM	830.4	-	287.1	112.2	401.4	200.3	-	-	17.7	0.0	-	1,849.1
	Totals	2,408.9	-	1,159.8	1,300.2	1,979.1	515.8	-	-	54.0	5.4	-	7,423.2
Reserve Margin	MDDE	515.0	-	33.5	180.9	279.4	-	61.7	-	-	1.6	-	1,072.2
	NE	248.9	-	173.5	788.3	197.3	165.0	-	-	0.0	3.6	-	1,576.7
	NJ	356.9	-	422.3	25.6	101.5	38.7	61.7	-	0.1	0.2	-	1,007.0
	NY	457.7	-	8.9	156.1	364.8	-	-	-	36.1	-	-	1,023.6
	NYC	-	-	238.2	-	634.7	-	61.7	-	-	-	-	934.6
	RoPJM	830.4	-	282.4	149.2	401.4	120.9	-	-	17.7	0.0	-	1,802.0
	Totals	2,408.9	-	1,158.8	1,300.2	1,979.1	324.6	185.1	-	54.0	5.4	-	7,416.1
Emissions Limits & RPS	MDDE	515.0	-	194.9	223.9	135.3	-	-	-	-	1.6	-	1,070.7
	NE	248.9	-	133.4	788.3	148.0	2.8	-	119.4	110.7	4.7	-	1,556.3
	NJ	356.9	-	440.5	24.4	59.7	-	-	-	0.1	0.2	-	881.8
	NY	457.7	-	13.4	263.6	264.7	-	-	-	36.1	-	-	1,035.5
	NYC	-	-	523.9	-	415.9	-	-	-	-	-	-	939.8
	RoPJM	830.4	-	213.7	-	401.4	192.0	-	7.2	137.3	159.5	-	1,941.5
	Totals	2,408.9	-	1,519.8	1,300.2	1,425.1	194.8	-	126.6	284.3	166.0	-	7,425.6
Social Cost of CO ₂ and CH ₄ Leakage	MDDE	515.0	58.1	126.5	292.2	17.7	-	61.7	-	-	1.6	-	1,072.9
	NE	248.9	16.4	386.6	788.3	41.6	9.0	-	-	0.0	3.6	-	1,494.6
	NJ	356.9	143.0	530.5	1.0	-	-	61.7	-	0.1	0.2	-	1,093.4
	NY	457.7	410.9	25.1	161.8	-	-	-	-	36.1	-	-	1,091.6
	NYC	-	197.7	614.6	-	25.0	-	61.7	-	-	-	-	899.0
	RoPJM	830.4	519.1	350.9	56.8	2.1	-	-	-	17.7	0.0	-	1,777.0
	Totals	2,408.9	1,345.2	2,034.1	1,300.2	86.5	9.0	185.1	-	54.0	5.4	-	7,428.4
Deterministic Health Damages	MDDE	515.0	44.2	126.5	227.5	-	-	61.7	97.6	-	1.6	-	1,074.1
	NE	248.9	7.5	425.9	788.3	19.6	2.1	61.7	-	0.0	3.6	-	1,557.6
	NJ	356.9	79.7	530.5	5.9	-	-	61.7	60.5	0.1	0.2	-	1,095.5
	NY	457.7	299.8	25.1	203.2	-	-	-	-	36.1	-	-	1,021.8
	NYC	-	105.7	659.5	-	2.8	-	61.7	114.7	-	-	-	944.4
	RoPJM	830.4	464.3	350.9	75.2	-	-	-	-	17.7	0.0	-	1,738.5
	Totals	2,408.9	1,001.2	2,118.3	1,300.2	22.4	2.1	246.8	272.8	54.0	5.4	-	7,432.0
Simulated Health Damages	MDDE	515.0	72.1	20.6	327.6	-	-	61.7	64.1	-	1.6	1.8	1,064.5
	NE	248.9	147.5	131.5	951.6	-	-	61.7	-	8.9	3.6	8.7	1,562.4
	NJ	356.9	513.2	68.8	11.6	-	-	61.7	64.1	0.1	0.2	-	1,076.7
	NY	457.7	406.8	3.0	173.5	-	-	-	-	36.1	-	-	1,077.1
	NYC	-	348.1	399.4	-	0.3	-	61.7	64.1	-	-	2.7	876.3
	RoPJM	830.4	648.9	120.5	152.7	-	-	-	-	17.7	0.0	-	1,770.2
	Totals	2,408.9	2,136.6	743.9	1,616.9	0.3	-	246.8	192.2	62.9	5.4	13.2	7,427.1

Figures 4.3 and 4.4 summarize the dispatching results on a case-by-case basis for each of the available unit type within the network. In the Baseline case, over 80% of the generation is from nuclear, coal, and natural gas units, with additional generation from hydro units and incremental amounts of wind on land generation. The Reserve Margin case shows similar results to the Baseline case; however, a small portion of coal generation is displaced by combined heat and power units. As evidenced by Table 4.4, combined heat and power generation is concentrated in the Maryland/Delaware, New Jersey, and New York City regions. This finding is also observed in the cases where health and other damages are minimized in the objective function along with market costs.

As we expand the costs in the objective function to include health and other damages, we observe significant decreases in coal generation. The inclusion of health and other damages in the objective function of the generation expansion planning model essentially behaves as a penalty for dispatching units that emit CO_2 , NO_x , and SO_2 . Because of this behavior, in cases where these damages are minimized in the objective function, coal generation is largely displaced by natural gas turbines, combined-cycle gas turbines, combined heat and power, hydro, and renewable generation under our assumptions.

4.4. Investment Results

Table 4.5 – Regional Investment Summary

		Investment Summary by Region (MW)								
Case	Region	CCGT	GT	Coal (High)	CHP	Petroleum	Wind (On Land)	Biomass	Solar	Totals
Baseline	MDDE	-	775	277	971	1,774	-	-	-	3,796
	NJ	-	513	443	1,032	1,281	-	-	-	3,268
	NYC	-	812	364	709	2,015	-	-	-	3,900
	Totals	-	2,099	1,084	2,712	5,070	-	-	-	10,964
Reserve Margin	MDDE	748	1,631	1,147	741	745	-	-	-	5,012
	NJ	745	1,749	1,820	753	752	-	-	-	5,818
	NYC	742	1,591	1,332	748	744	-	-	-	5,156
	Totals	2,235	4,971	4,298	2,242	2,241	-	-	-	15,987
Emissions Limits & RPS	MDDE	-	5,084	-	-	-	-	-	-	5,084
	NE	-	-	-	-	-	3,782	1,799	105	5,687
	NJ	-	5,888	-	-	-	-	-	-	5,888
	NYC	-	5,178	-	-	-	-	-	-	5,178
	RoPJM	-	-	-	-	-	4,841	107	14,796	19,745
	Totals	-	16,150	-	-	-	8,624	1,906	14,901	41,581
Social Cost of CO ₂ and CH ₄ Leakage	MDDE	759	1,637	1,137	755	756	-	-	-	5,044
	NE	-	2,988	-	-	-	-	-	-	2,988
	NJ	756	1,740	1,793	746	758	-	-	-	5,792
	NYC	762	2,256	1,362	745	759	-	-	-	5,884
	RoPJM	-	759	-	-	-	-	-	-	759
	Totals	2,277	9,379	4,293	2,246	2,272	-	-	-	20,467
Deterministic Health Damages	MDDE	750	1,635	-	752	751	-	1,140	-	5,029
	NE	-	3,749	-	757	-	-	-	-	4,506
	NJ	750	1,754	1,090	751	758	-	702	-	5,806
	NYC	756	3,013	-	751	754	-	1,364	-	6,639
	RoPJM	-	761	-	-	-	-	-	-	761
	Totals	2,256	10,913	1,090	3,011	2,263	-	3,207	-	22,741
Simulated Health Damages	MDDE	922	1,451	-	735	1,148	-	743	-	4,998
	NE	2,232	2,456	-	747	-	317	-	-	5,753
	NJ	756	1,745	-	747	1,816	-	744	-	5,808
	NYC	2,313	1,858	-	743	1,336	-	748	-	6,998
	Totals	6,223	7,511	-	2,973	4,300	317	2,234	-	23,558

Table 4.5 summarizes the technology investments selected by the generation expansion-planning model within our proposed network over for each of the cases studied. The Baseline case results propose no new technology investments. The Baseline case proposes investments in coal, natural gas turbines, combined heat and power, and petroleum units. The inclusion of the reserve margin constraint introduces combined cycle gas turbines into the investment portfolio, as well as increases in natural gas, coal, and combined heat and power generation. Furthermore, the inclusion of emissions limits and renewable portfolio standards proposes natural gas investments in the Maryland/Delaware, New Jersey, and New York City regions, which is consistent with the Baseline and Reserve Margin cases. Additionally, emissions limits and renewable portfolio standards drive investments in wind (on land), biomass, and solar units in the New England and Rest of PJM regions.

The inclusion of the social cost of CO₂ and CH₄ leakage into the objective function yields investments in combined cycle gas turbines, natural gas units, as well as investments in coal, combined heat and power, and petroleum at the similar levels observed in the reserve margin case. Introducing deterministic health damages into the objective function decreases coal investments, but introduces biomass into the investment plan. Replacing deterministic health damages with simulated health damages, via the framework presented in Chapter 3, eliminates coal from the investment plan, but significantly offsets these investments with combined cycle gas turbines. Additional investments are made in natural gas, biomass, wind (on land), and petroleum units.

4.5. Emissions and Human Health Results

Table 4.6 – CO₂ Emissions Summary

	Regional CO ₂ Emissions (Short Tons, Millions)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Total
Baseline	3,668	5,118	4,410	4,289	8,757	8,396	34,637
Reserve Margin	3,682	5,059	4,322	4,153	8,839	7,476	33,532
Emissions & RPS Constraints	2,627	2,453	3,171	3,053	7,650	7,886	26,840
Social Cost of CO₂ and CH₄ Leakage	1,597	2,857	4,172	2,474	5,241	4,961	21,303
Deterministic Health Damages	1,319	3,054	3,813	1,843	4,724	4,626	19,379
Simulated Health Damages	892	2,006	3,653	2,326	4,618	4,366	17,861

Table 4.7 – NO_x Emissions Summary

	Regional NO _x Emissions (Short Tons, Thousands)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Total
Baseline	9,784	12,501	9,178	11,284	21,467	20,492	84,706
Reserve Margin	9,192	12,344	8,317	11,021	21,590	18,070	80,534
Emissions & RPS Constraints	5,715	8,047	5,535	8,055	16,930	19,763	64,046
Social Cost of CO₂ and CH₄ Leakage	2,626	4,947	6,249	3,706	8,180	7,458	33,165
Deterministic Health Damages	3,927	4,860	6,921	2,761	9,408	6,929	34,805
Simulated Health Damages	2,630	3,070	6,753	3,483	8,223	6,539	30,700

Table 4.8 – SO₂ Emissions Summary

	Regional SO ₂ Emissions (Short Tons, Thousands)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Total
Baseline	21,200	23,976	12,903	24,018	41,397	39,256	162,751
Reserve Margin	18,211	23,637	9,350	23,718	41,405	34,093	150,413
Emissions & RPS Constraints	8,890	15,842	4,101	17,214	27,294	39,039	112,380
Social Cost of CO₂ and CH₄ Leakage	1,277	3,496	368	218	2,063	573	7,995
Deterministic Health Damages	4,994	1,659	3,361	162	6,333	408	16,917
Simulated Health Damages	3,390	692	3,526	205	3,790	385	11,988

Tables 4.6 through 4.8 display emissions results for CO₂, NO_x, and SO₂ respectively for each of the six cases studied in this chapter. In comparing our emissions findings with the corresponding dispatching plans, as we include additional constraints in the cases where only market costs are minimized, we observed that coal generation is displaced by cleaner units. Consequently, this reduces air emissions, mainly CO₂, NO_x, and SO₂ relative to the Baseline case. decrease due to the displacement of fossil fuels in the dispatch plan.

The inclusion of health and other damages into the objective function of the generation expansion planning model further reduces coal generation, displacing these units with natural gas, combined cycle gas turbines, combined heat and power, and hydro generation. This is the main driver of the significant reductions in air emissions relative to the cases where only market costs are minimized. However, in cases where human health externalities from NO_x and SO₂ are minimized, whether they are treated deterministically or simulated via the procedure outlined in Chapter 3, we notice slight increases in SO₂ emissions, which is driven by biomass generation.

Table 4.9 – Expected Negative Health Outcomes Summary

	Expected Health Outcomes					
	Baseline	Reserve Margin	Emissions Limits & RPS	Social Cost of CO₂ and CH₄ Leakage	Deterministic Health Damages	Simulated Health Damages
Adult Mortality	48,165	43,528	35,124	21,208	19,551	17,006
Infant Mortality	64	59	48	29	27	24
Non-fatal Heart Attacks	19,461	17,654	14,598	8,695	8,049	6,969
Respiratory-Related Hospital Admissions	8,295	7,738	6,288	3893	3,534	3,077
Cardiovascular-Related Hospital Admissions	10,341	9,435	7,802	4716	4,380	3,852
Acute Bronchitis	44,440	40,548	33,395	20,267	18,921	16,718
Upper Respiratory Symptoms	761,589	696,882	582,393	351,833	323,773	282,377
Lower Respiratory Symptoms	564,004	519,285	428,456	258,550	236,480	206,081
Asthma ER Visits	15,343	14,358	11,484	6940	6,411	5,666
Minor Restricted Activity Days	21,001,716	19,423,414	15,989,868	9,487,067	8,810,074	7,669,236
Work Loss Days	3,568,670	3,293,161	2,694,189	1,648,502	1,533,920	1,355,337
Asthma Exacerbations	800,580	737,471	605,492	364,626	330,542	282,428

Table 4.9 displays the expected number of negative health outcomes as a result of NO_x and SO₂ emissions, as modeled in the COBRA tool for each of the cases studied. Generally speaking, negative health outcomes are reduced from the Baseline for all cases. However, in the cases where health damages are included in the objective function, we observe dramatic reductions in negative human health effects. Again, this is driven by significant reductions in fossil fuel generation, which drives reductions in pollutant emissions.

4.6. Cost and Pricing Results

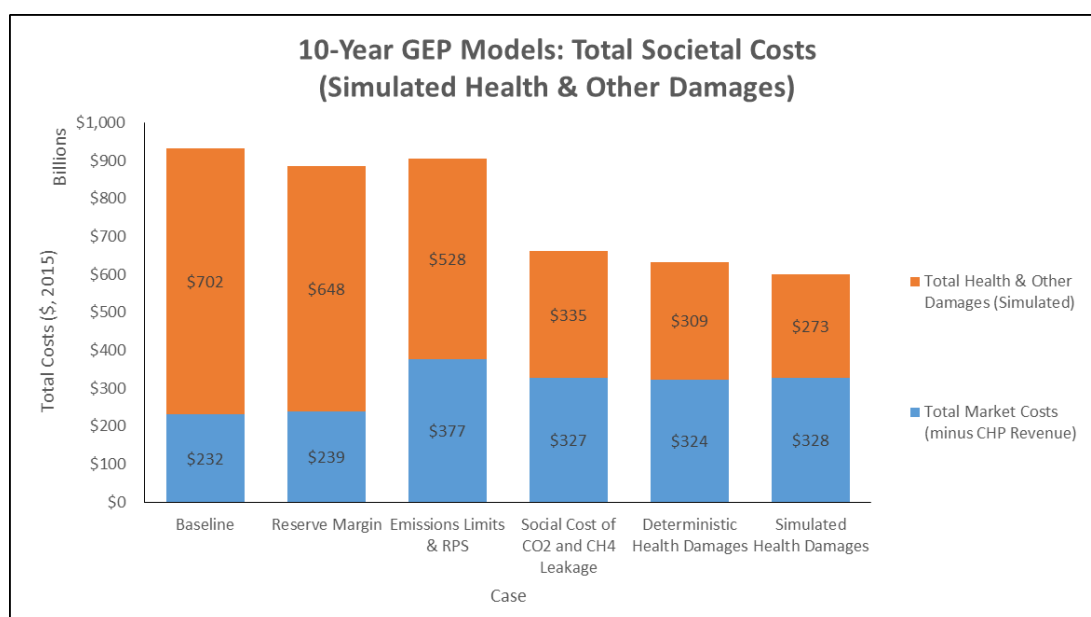


Figure 4.5 – Total Cost Summary (Including Simulated Health Damages)

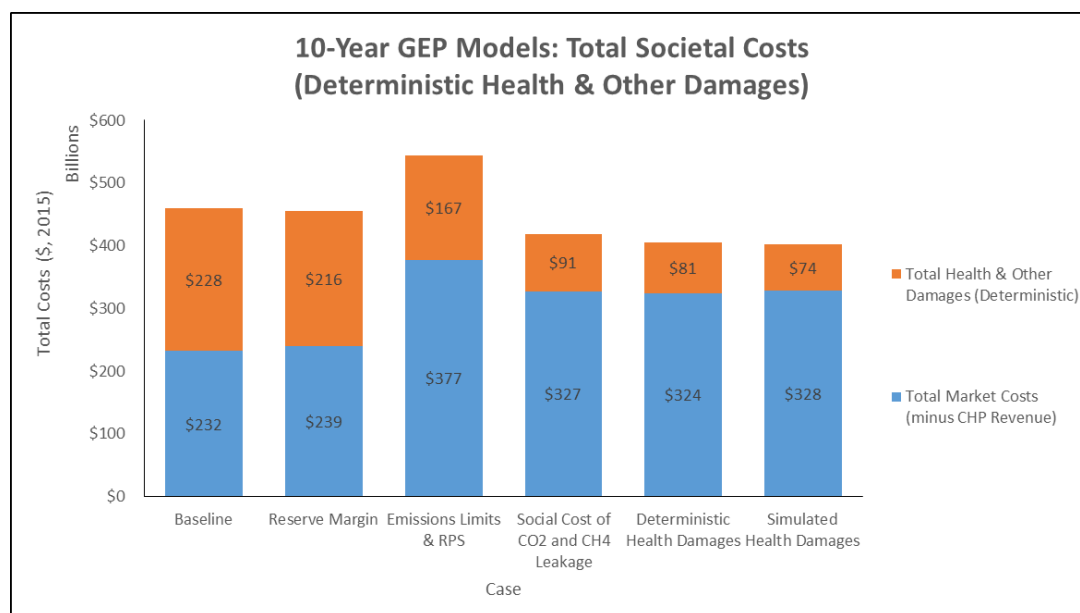


Figure 4.6 – Total Cost Summary (Including Deterministic Health Damages)

Figures 4.5 and 4.6 displays the total objective function costs for each of the six cases presented in this chapter. Market costs, specifically investment costs, fixed O&M costs, and variable O&M costs (including fuel) are minimized in the Baseline, Reserve Margin, and Emissions Limits & RPS cases. Whereas, total societal costs, including health and other damages (specifically the social cost of CO₂ and CH₄ leakage and human health externalities from NO_x and SO₂ emissions) are minimized in the Social Cost of CO₂ and CH₄ Leakage, Deterministic Health Damages, and Simulated Health Damages cases. In the Deterministic Health Damages case, human health externalities from NO_x and SO₂ emissions are obtained from the values shown in Table 4.4; however, human health externalities from NO_x and SO₂ emissions in the Simulated Health Damages case are obtained via the COBRA simulation procedure presented in Chapter 3. In general, in cases where only market costs are optimized, health damages are significantly greater than in cases where total societal costs are optimized.

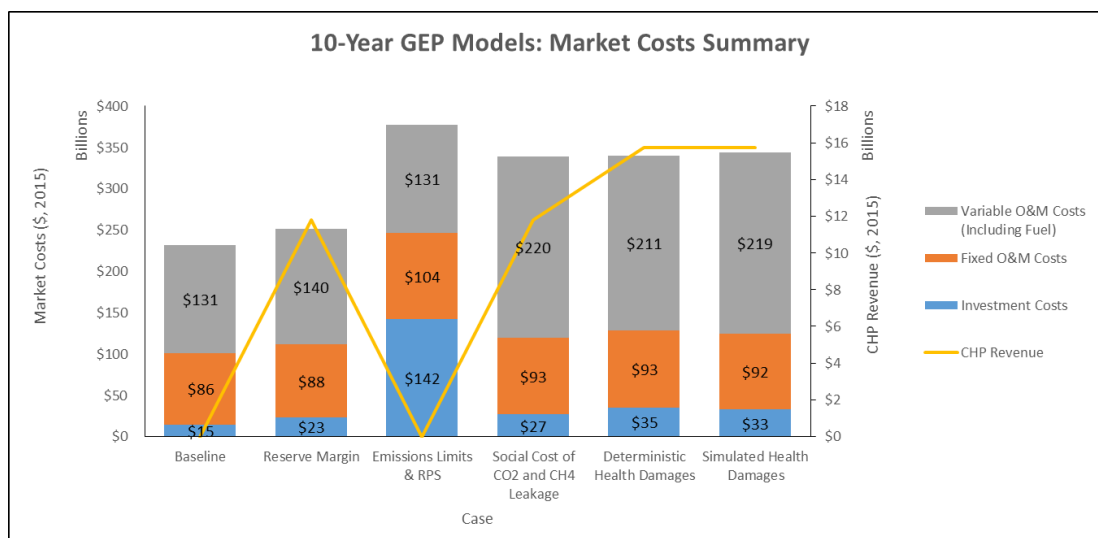


Figure 4.7 – Market Cost Summary

Figure 4.7 displays the market costs for each of the six cases presented in this chapter. The Baseline and Reserve Margin models exhibit similar results as their expansion plans are nearly similar. However, the inclusion of emissions limits and renewable portfolio standards significantly increases investment and fixed O&M costs relative to the Baseline model, which is due to significant investments in renewable units. Including health and other damages in the objective function of the model significantly increases variable O&M (including fuel) costs, which is primarily due to the fact that coal generation is displaced by combined cycle gas turbines and natural gas units.

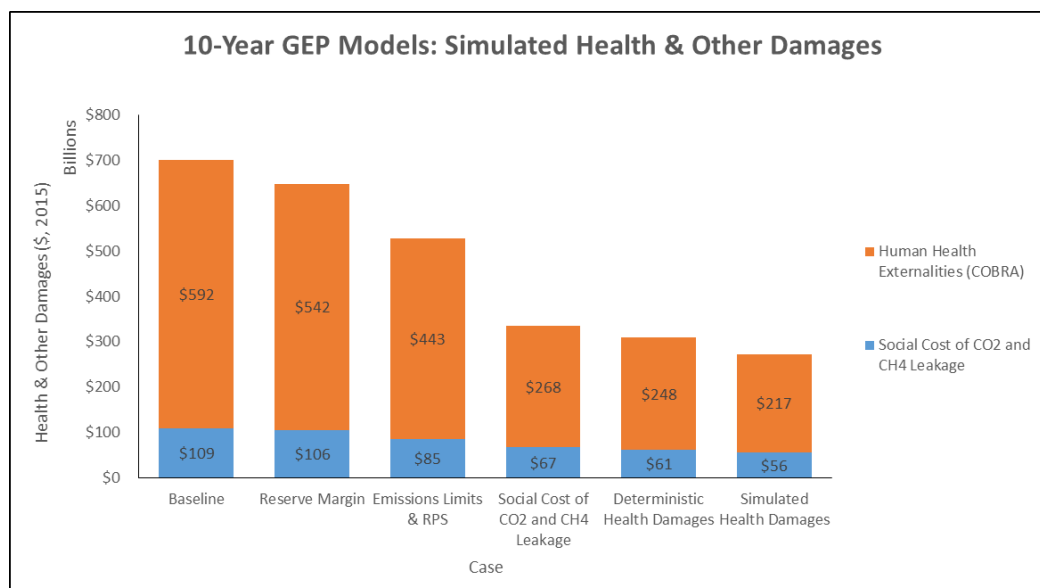


Figure 4.8 – Simulated Health Damages Summary

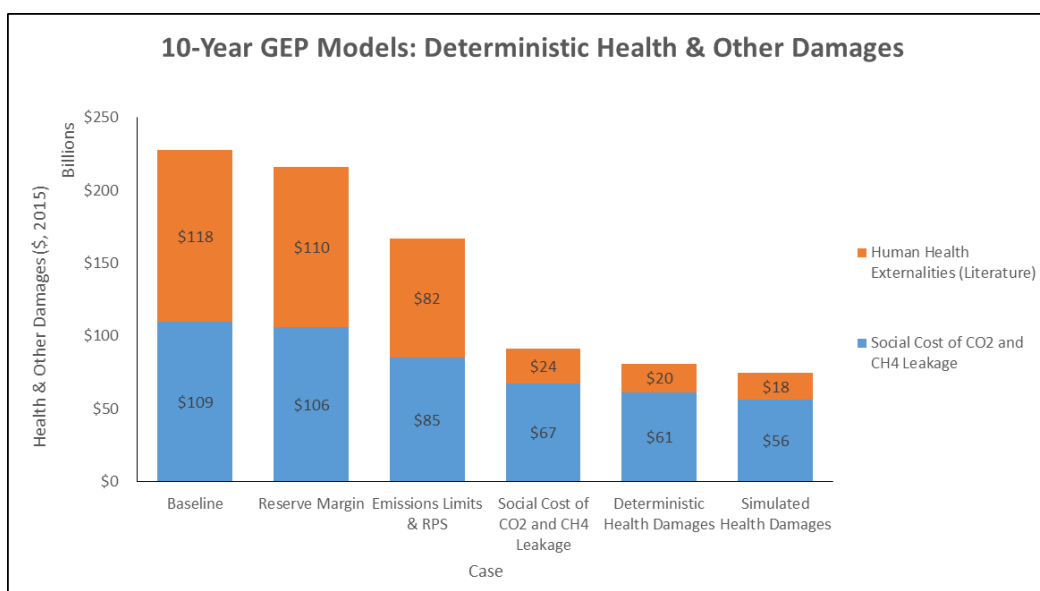


Figure 4.9 – Deterministic Health Damages Summary

Figures 4.8 and 4.9 summarize the simulated and deterministic health damages associated with the expansion plan selected in each of the cases studied. As noticed in these figures, in cases where only market costs are optimized, we observe that human health externalities are significantly greater than cases where total societal costs (including health damages) are minimized. In cases where health damages are

minimized, this drives significant increases in hydro generation, which displaces fossil fuel generation. Ultimately, this results in significant emissions reductions and reductions in health damages within the network.

Furthermore, human health externalities calculated via the COBRA simulation procedure outlined in Chapter 3 are generally much larger in magnitude in comparison to the deterministic approach. As mentioned previously, human health externalities are highly non-linear functions of various factors, such as temperature, location, and emissions from other sources. Since COBRA is able to capture these relationships and approximate them statistically, the simulation procedure is able to monetize the health damages more effectively than simply applying deterministic multipliers to penalize pollutant-emitting units.

Table 4.10 – Expected Human Health Externalities Summary

	Expected Human Health Externalities (\$ 2015, millions)					
	Baseline	Reserve Margin	Emissions & RPS	Social Cost of CO ₂ and CH ₄ Leakage	Deterministic Health Damages	Simulated Health Damages
Adult Mortality	\$581,343.7	\$532,535.1	\$435,422.8	\$262,830.9	\$243,384.0	\$212,850.1
Infant Mortality	\$572.7	\$517.2	\$419.0	\$250.6	\$235.1	\$212.8
Non-fatal Heart Attacks	\$6,444.9	\$5,915.3	\$4,806.0	\$2,788.5	\$2,618.5	\$2,378.7
Respiratory-Related Hospital Admissions	\$371.8	\$337.9	\$274.1	\$165.5	\$154.8	\$141.8
Cardiovascular-Related Hospital Admissions	\$255.9	\$231.2	\$184.9	\$111.1	\$102.0	\$92.3
Acute Bronchitis	\$18.5	\$16.7	\$13.9	\$8.2	\$7.7	\$6.9
Upper Respiratory Symptoms	\$4.8	\$4.4	\$3.6	\$2.2	\$2.0	\$1.8
Lower Respiratory Symptoms	\$22.5	\$20.0	\$16.6	\$10.1	\$9.3	\$8.3
Asthma ER Visits	\$9.4	\$8.6	\$6.9	\$4.2	\$3.9	\$3.5
Minor Restricted Activity Days	\$2,319.0	\$2,057.1	\$1,706.7	\$1,011.7	\$936.8	\$839.3
Work Loss Days	\$781.1	\$718.0	\$580.3	\$341.1	\$323.4	\$288.7
Asthma Exacerbations	\$62.6	\$57.0	\$46.9	\$28.0	\$26.1	\$23.4
Totals	\$592,207	\$542,418	\$443,482	\$267,552	\$247,804	\$216,848

Corresponding to the expected number of negative health outcomes displayed in Table 4.9, Table 4.10 displays the associated monetized value of the human health externalities as a result of NO_x and SO_2 emissions, as modeled in the COBRA tool for each of the cases studied. Again, negative health outcomes are reduced from the Baseline for all cases. Again, in the cases where health damages are included in the objective function, we observe dramatic reductions in human health externalities driven by significant reductions in coal generation. Across all cases, the major contributor to human health externalities is the number of adult lives lost (or mortality), which decreases significantly as health damages are reduced as part of the objective function.

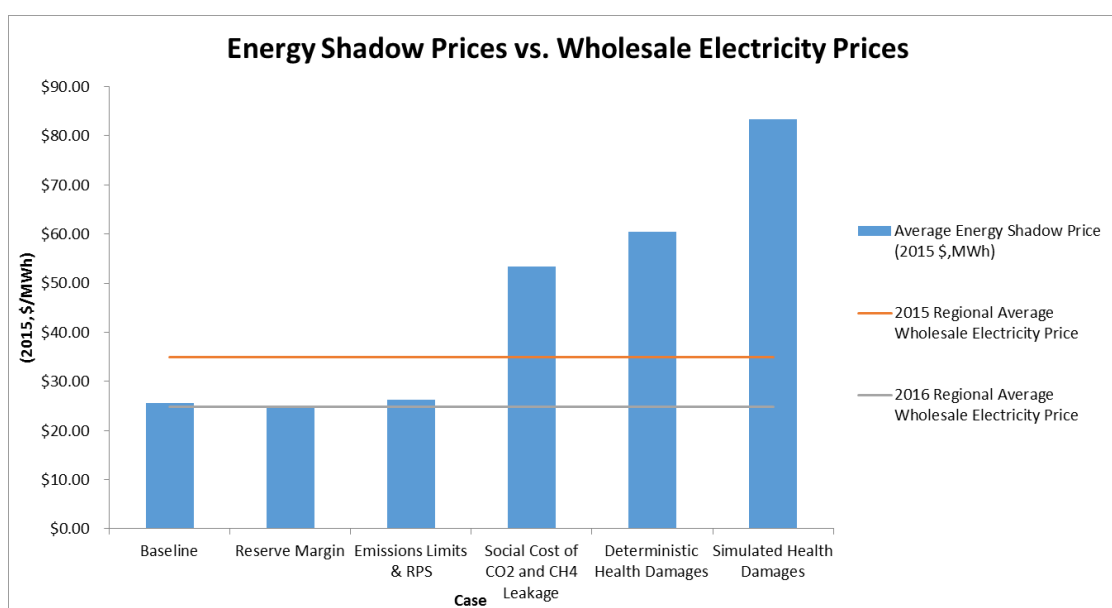


Figure 4.10 – Energy Comparison (U.S. Energy Information Administration, 2015)

In the context of the generation expansion planning model, the dual variables associated with the energy balance constraint are interpreted as the energy shadow price. In the context of electricity markets, this is the analog of wholesale electricity prices, which are derived from day-ahead unit commitment models. The results from the cases in this chapter are presented in Figure 4.10. The ultimate take-away from our analysis is

that minimizing societal costs in the objective function could potentially lead to increased energy prices. Extending this research to day-ahead markets would allow for further insight into this precursory finding.

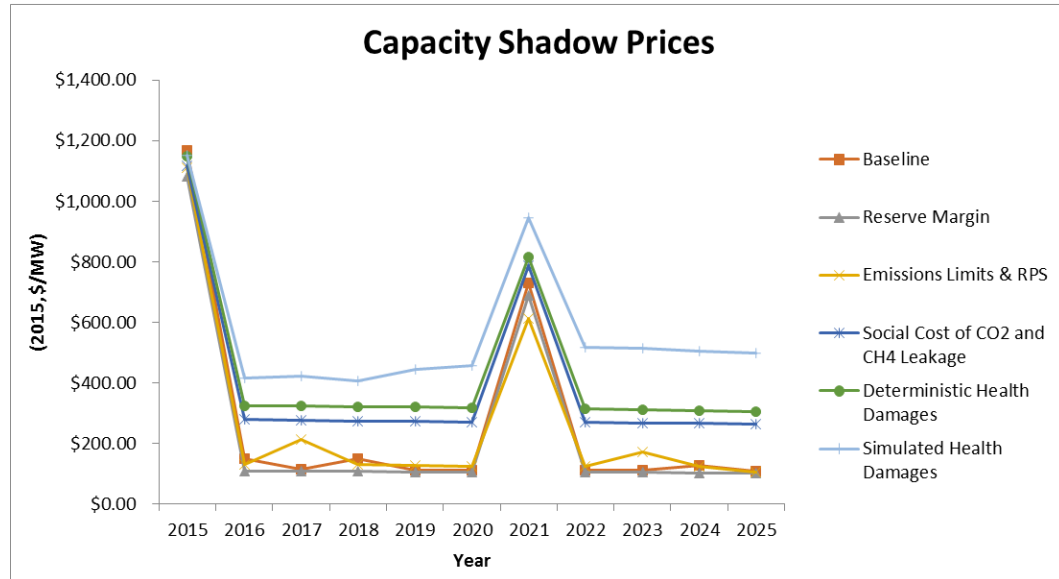


Figure 4.11 – Capacity Shadow Prices

Similar to the interpretation of the dual variables associated with the energy balance constraints, we quantify the dual variables associated with the capacity constraints and refer to them as capacity shadow prices. That is, for each MW of capacity invested, there is an incremental \$/MW increase to the objective function value. To elaborate on this interpretation further, the decision variables on the left-hand side of this constraint are dispatching decisions by region, period, and year. However, on the right-hand side of the capacity constraints, we also have capacity expansion decision variables. Thus, the dual variables, in this case, suggest that as we expand capacity on the right-hand side of the capacity constraints, it ultimately penalizes the objective function over time. The results shown in Figure 4.11 depict this concept for each of the cases studied. In general, based on our findings, the costs of capacity expansion are

much greater in the early stages of the planning process, relative to the remainder of the time-horizon.

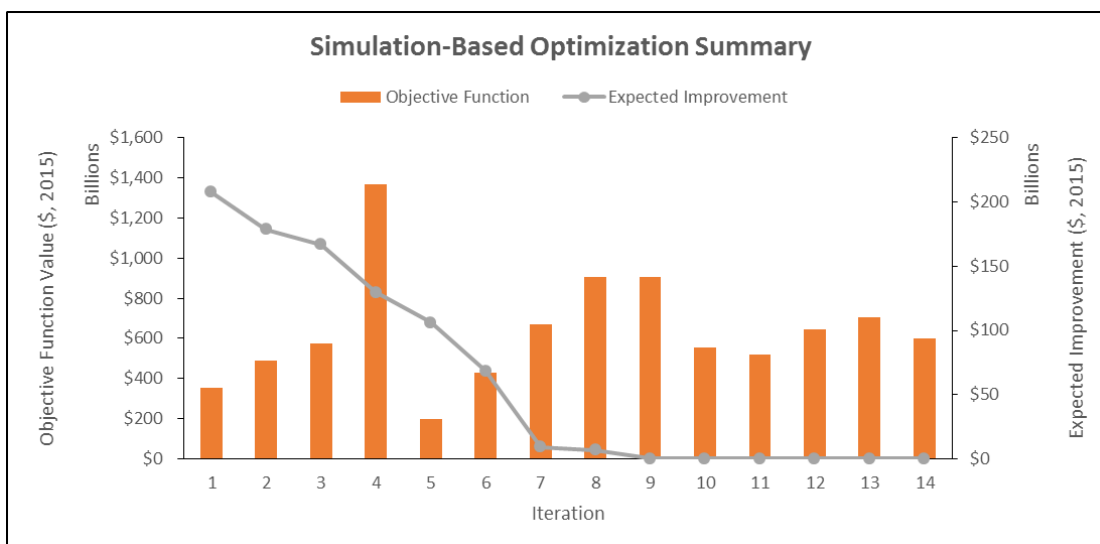


Figure 4.12 – Simulation-Based Optimization Iteration Summary

Case 1.6, alternatively referred to as the Simulated Health Damages case, applies the simulation-based optimization framework presented in Chapter 3 in order to solve the generation expansion-planning problem that minimizes total societal costs, inclusive of market costs and health damages. Following our procedure in Chapter 3, as depicted in Figure 4.12, we are able to achieve a solution within 16 full iterations. As we progress through each of the iterations, the expected improvement value decreases and eventually reaches zero upon termination of the algorithm. The objective function, however, behaves rather erratically initially, but begins to level off beginning at iteration 10.

5. Full-Form Model Evaluation

In this section, we apply our learnings from Chapters 3 and 4 to six (6) expanded cases using the network described in Figure 4.1. The assumptions presented in Section 4.2 of Chapter 4 are applied to all scenarios in this chapter. In comparison to our reduced-form model, the cases in this chapter include seasonality indices, and extend the time-horizon to 25 years (2015 through 2040).

Table 5.1 outlines the definitions of the six cases we consider in this section. Cases 2.1 through 2.6 are extensions of Cases 1.1 through 1.6 from Chapter 4. Cases 2.1 through 2.5 are deterministic models solved by a CPLEX linear programming tool, whereas Case 2.6 is solved via the simulation-based optimization framework presented in Chapter 3.

Table 5.1 – Full Model Test Cases

		Objective Function Costs						Constraints					
		Market Costs			Health Damages								
		Investment	Fixed O&M	Variable O&M (including fuel)	Social Cost of CO ₂ and CH ₄ Leakage	Deterministic NO _x and SO ₂ Multipliers	Simulated Human Health Externalities (COBRA)	Energy Balance	Capacity	Construction Limits	Reserve Margin	Emissions Limits	Renewable Portfolio Standards (RPS)
Case 2.1	Baseline	✓	✓	✓				✓	✓	✓			
Case 2.2	Reserve Margin	✓	✓	✓				✓	✓	✓	✓		
Case 2.3	Emissions & RPS Constraints	✓	✓	✓				✓	✓	✓	✓	✓	✓
Case 2.4	Social Cost of CO ₂ and CH ₄ Leakage	✓	✓	✓	✓			✓	✓	✓	✓		
Case 2.5	Deterministic Health Damages	✓	✓	✓	✓	✓		✓	✓	✓	✓		
Case 2.6	Simulated Health Damages	✓	✓	✓	✓		✓	✓	✓	✓	✓		

Similar to Case 1.6 from Chapter 4, in Case 2.6, where we apply our proposed simulation-based optimization framework to the network, we vary the key model parameters, as identified in Section 3.2.3 in Chapter 3, by simulating from a normal distribution with a coefficient of variation uniformly distribution on the interval [0,1]. Next, we execute multiple trials of our experimental GEP model and compare the results to the baseline GEP model, Case 2.1, to obtain changes in dispatching decisions by unit and changes in NO_x and SO_2 emissions. Upon obtaining results from the simulations, we then input the changes in emissions into the COBRA model to quantify the associated human health externalities for each trial. Once we generate the full data set, we fit a kriging metamodel, as outlined in Section 3.3.1 of Chapter 3, that predicts human health externalities as a function of dispatching decisions for each unit in the network. This metamodel serves as a surrogate curve for human health externalities in the objective function of our model. We test the validity and robustness of our metamodel by applying the same modified K-fold cross-validation procedure to the simulated data set, which generates the results in Table 5.2 and Figure 5.1.

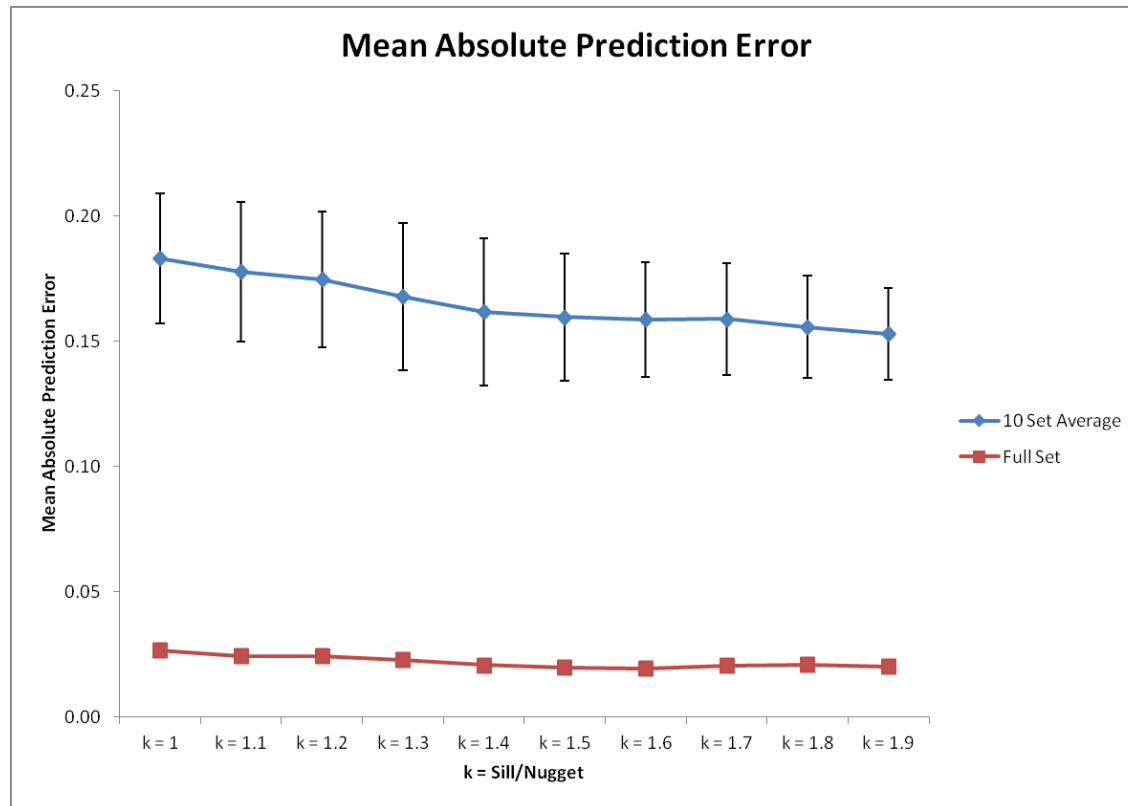


Figure 5.1 – Mean Absolute Prediction Error Values from Cross Validation Procedure

As evidenced by Table 5.2 and Figure 5.1, we execute the modified K-fold cross-validation procedure for various sill and nugget parameters in the Gaussian variogram, as discussed in Chapter 4. Based on the individual validation sets, on average, as we increase the k-ratio, the prediction error decreases. Additionally, the error variability (or standard deviation) begins to decrease after k-ratios greater than 1.4. Upon applying the full data set to fit our metamodel, however, we notice that the prediction error is significantly reduced, and remains relatively stable for all values of k. In general, the more exhaustive and expansive the sample space, the greater the prediction properties exhibited by the model. From a qualitative perspective, our metamodel is robust for all values of k. Thus, we select $k=1.5$ as our parameter to apply to the variogram. Using this metamodel as a surrogate in the objective function of our GEP test model, we apply the full framework in Chapter 3 to determine an optimal (or near-optimal) expansion plan for Case 2.6.

5.1 Dispatching Results

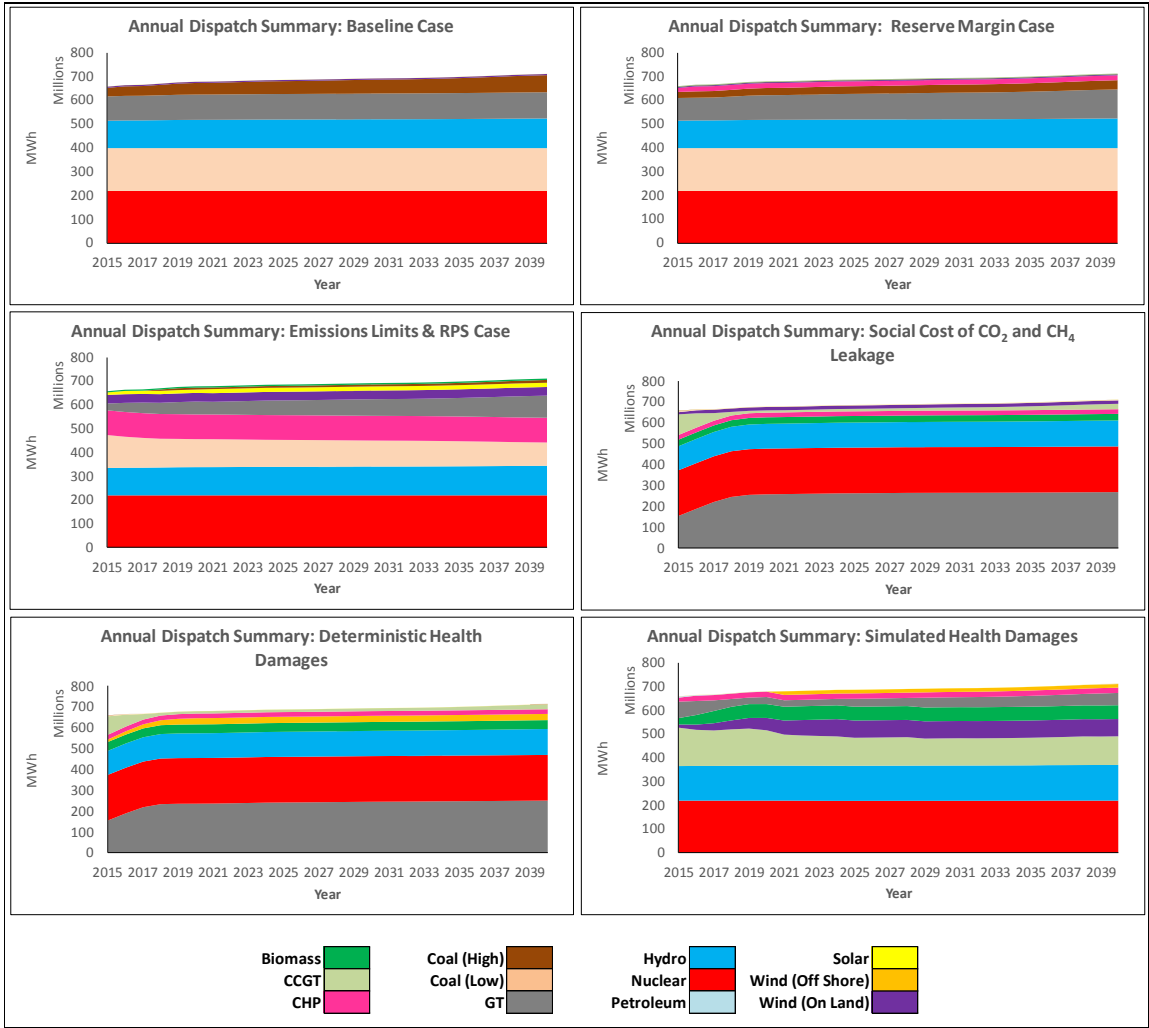


Figure 5.2 – Annual Dispatch Summary Dashboard

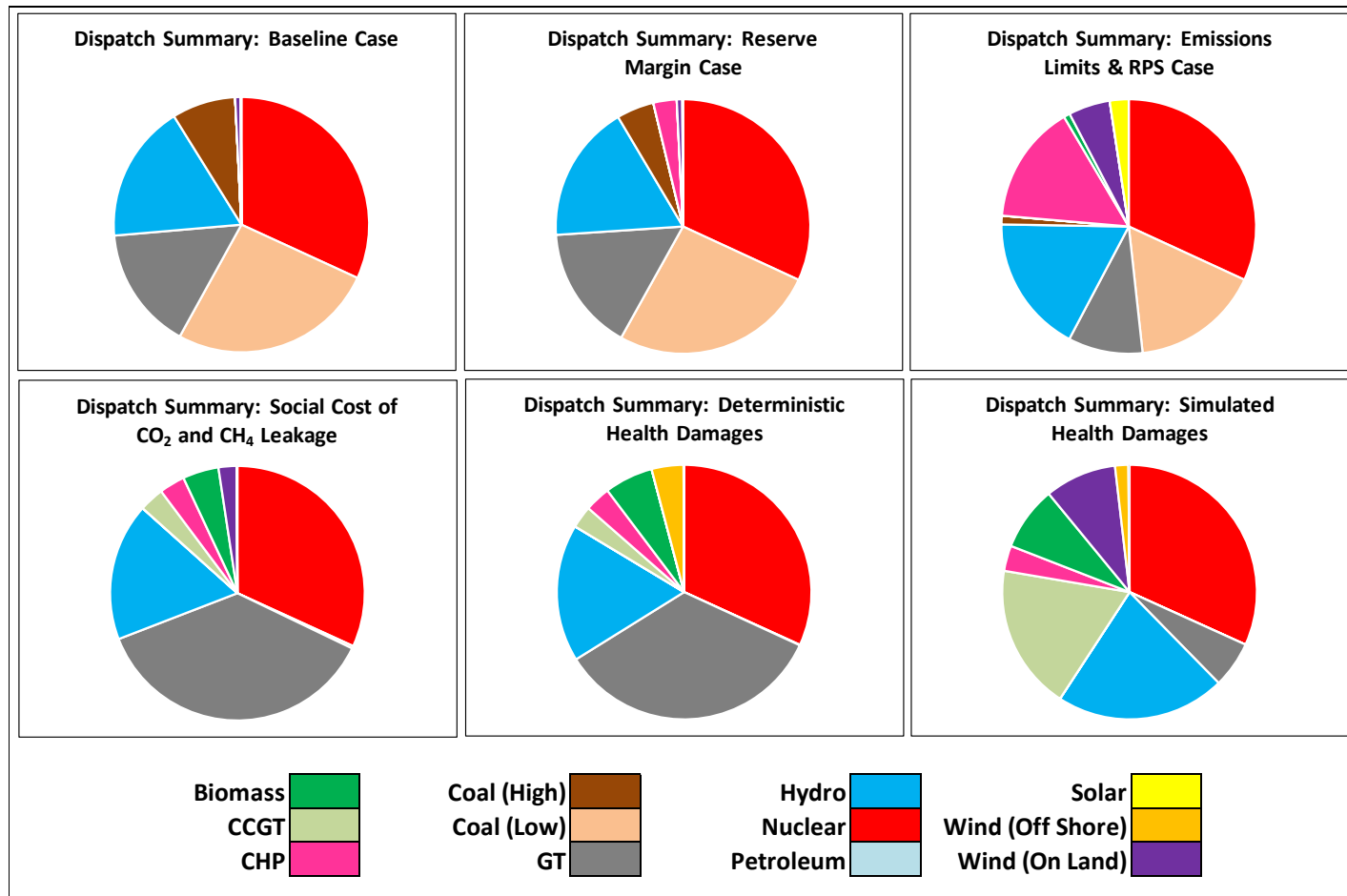


Figure 5.3 – Total Dispatch Percentage Dashboard

Table 5.3 – Regional Dispatch Summary

Case	Region	Dispatch Summary by Region (10 ⁶ MWh)												
		Nuclear	Coal (Low)	GT	Hydro	Coal (High)	CCGT	CHP	Biomass	Wind (On Land)	Wind (Off Shore)	Solar	Petroleum	Totals
Baseline	MDDE	1,217.3	660.5	-	573.7	122.0	-	-	-	-	-	3.9	-	2,577.3
	NE	588.4	466.3	410.1	1,898.2	437.2	-	-	-	0.0	-	8.4	-	3,808.6
	NJ	843.6	239.8	928.6	59.7	246.1	-	-	-	0.3	-	0.4	-	2,318.6
	NY	1,081.8	862.3	58.2	355.9	45.5	-	-	-	85.4	-	-	-	2,489.1
	NYC	-	1,500.2	712.8	-	-	0.3	-	-	-	-	-	-	2,213.3
	RoPJM	1,962.7	948.9	681.4	243.0	591.2	-	-	-	41.9	-	0.0	-	4,469.2
	Total	5,693.8	4,677.9	2,791.1	3,130.5	1,442.0	0.3	-	-	127.7	-	12.8	-	17,876.1
Reserve Margin	MDDE	1,217.3	660.5	149.1	405.0	-	-	145.8	-	-	-	3.9	-	2,581.5
	NE	588.4	466.3	410.1	1,898.2	424.9	-	-	-	0.0	-	8.4	-	3,796.3
	NJ	843.6	239.8	1,009.2	48.7	86.4	-	196.2	-	0.3	-	0.4	-	2,424.6
	NY	1,081.8	862.3	28.2	407.0	-	-	-	-	85.4	-	-	-	2,464.7
	NYC	-	1,500.2	561.9	-	-	-	188.2	-	-	-	-	-	2,250.3
	RoPJM	1,962.7	948.9	674.7	371.7	339.0	-	-	-	41.9	-	0.0	-	4,338.9
	Total	5,693.8	4,677.9	2,833.2	3,130.5	850.2	-	530.2	-	127.7	-	12.8	-	17,856.3
Emissions Limits & RPS	MDDE	1,217.3	134.2	67.2	349.4	-	-	805.3	-	-	-	3.9	-	2,577.1
	NE	588.4	350.5	249.7	1,898.2	6.2	-	-	122.6	522.3	1.0	11.4	-	3,750.2
	NJ	843.6	135.5	305.8	66.1	-	-	1,019.9	-	0.3	-	0.4	-	2,371.6
	NY	1,081.8	520.7	25.6	790.8	-	-	-	-	85.4	-	-	-	2,504.3
	NYC	-	838.1	543.0	-	-	-	880.5	-	-	-	-	-	2,261.6
	RoPJM	1,962.7	948.9	502.0	26.0	190.6	-	-	19.9	336.5	-	413.2	-	4,399.7
	Total	5,693.8	2,927.8	1,693.2	3,130.5	196.7	-	2,705.7	142.5	944.5	1.0	428.8	-	17,864.6
Social Cost of CO ₂ and CH ₄ Leakage	MDDE	1,217.3	0.8	511.7	478.3	-	21.4	145.8	182.1	40.6	-	3.9	-	2,601.9
	NE	588.4	53.9	956.1	1,898.2	2.2	10.3	104.8	-	148.4	-	8.4	-	3,770.7
	NJ	843.6	-	1,282.1	16.4	-	95.6	145.8	364.7	0.3	-	0.4	-	2,749.0
	NY	1,081.8	-	724.6	442.2	-	54.5	-	-	184.0	-	-	-	2,487.1
	NYC	-	3.4	1,607.3	-	-	132.9	145.8	271.1	-	-	-	-	2,160.6
	RoPJM	1,962.7	-	1,536.5	295.5	-	246.5	41.0	-	41.9	-	0.0	-	4,124.2
	Total	5,693.8	58.1	6,618.4	3,130.5	2.2	561.3	583.3	817.9	415.2	-	12.8	-	17,893.4
Deterministic Health Damages	MDDE	1,217.3	-	438.9	453.2	-	7.6	145.8	318.4	-	40.6	3.9	-	2,625.5
	NE	588.4	10.5	994.5	1,898.2	1.2	8.0	145.8	-	-	278.9	8.4	-	3,933.9
	NJ	843.6	-	1,282.7	17.8	-	93.0	145.8	364.7	-	0.3	0.4	-	2,748.4
	NY	1,081.8	-	479.9	438.9	-	57.8	-	-	-	317.4	-	-	2,375.9
	NYC	-	1.0	1,516.3	-	-	102.9	145.8	422.6	-	-	-	-	2,188.6
	RoPJM	1,962.7	-	1,423.7	322.4	-	235.9	-	-	-	86.7	0.0	-	4,031.4
	Total	5,693.8	11.5	6,135.9	3,130.5	1.2	505.3	583.3	1,105.6	-	723.9	12.8	-	17,903.8
Simulated Health Damages	MDDE	1,217.3	-	21.5	692.7	-	58.2	145.8	197.8	251.2	-	3.9	-	2,588.3
	NE	576.7	-	162.2	2,249.2	-	127.8	-	150.4	425.3	109.3	8.4	10.5	3,819.9
	NJ	843.6	-	89.4	70.1	-	873.3	145.8	422.2	99.3	-	0.4	-	2,544.1
	NY	1,081.8	-	0.6	443.3	-	315.0	-	-	501.1	193.8	-	-	2,535.6
	NYC	-	0.3	649.2	-	-	576.6	286.0	591.2	-	-	-	2.5	2,105.8
	RoPJM	1,962.7	-	130.0	395.9	-	1,354.1	-	95.2	351.5	-	0.0	-	4,289.5
	Total	5,682.1	0.3	1,052.9	3,851.2	-	3,304.9	577.7	1,456.7	1,628.4	303.1	12.8	13.0	17,883.2

Figures 5.2 and 5.3 summarize the dispatching results on a case-by-case basis for each of the available unit type within the network. In the Baseline and Reserve Margin cases, we observe similar trends as observed in Chapter 4. In these cases, approximately 80% of total generation is from nuclear, coal, and natural gas sources. In the Reserve Margin case, a small portion of coal generation is displaced by combined heat and power. Implementing emissions limits and renewable portfolio standards as constraints in the model reduces fossil fuel generation, and increases combined heat and power as well as renewable generation from wind on land, solar, and biomass units.

In the cases where health damages are included in the objective function, particularly the Social Cost of CO₂ and CH₄ Leakage, Deterministic Health Damages, and Simulated Health Damages cases, we observe that coal dispatch is significantly displaced by other units. In the Social Cost of CO₂ and CH₄ Leakage case, coal generation is displaced by natural gas and combined cycle gas turbines. The inclusion of deterministic health damages in the objective function also exhibits similar behavior, and also results off-shore wind generation being constructed. However, the inclusion of simulated health damages in the objective function of the model, via the framework presented in Chapter 3, further shows increased levels of combined cycle gas turbine generation, wind (on land and off-shore), biomass, and hydro generation. Coal is no longer in the dispatch plan in this scenario.

5.2 Investment Results

Table 5.4 – Regional Summary of New Investments

Case	Region	Investment Summary by Region (MW)										Totals
		GT	Wind (On Land)	CHP	Solar	Petroleum	Biomass	CCGT	Wind (Off Shore)	Coal (High)	Hydro	
Baseline	MDDE	864	-	927	-	1,981	-	-	-	258	-	4,030
	NJ	469	-	1,029	-	1,231	-	-	-	444	-	3,172
	NYC	704	-	612	-	1,907	-	-	-	311	-	3,534
	Total	2,037	-	2,568	-	5,119	-	-	-	1,013	-	10,736
Reserve Margin	MDDE	2,348	-	648	-	709	-	717	-	404	738	5,564
	NJ	2,379	-	892	-	698	-	649	-	1,446	-	6,065
	NYC	2,169	-	1,033	-	795	-	592	-	1,420	-	6,009
	Total	6,897	-	2,574	-	2,202	-	1,957	-	3,270	738	17,637
Emissions Limits & RPS	MDDE	1,638	-	3,880	-	-	-	-	-	-	-	5,518
	NE	-	5,997	-	99	-	510	-	12	-	-	6,617
	NJ	1,485	-	5,009	-	-	-	-	-	-	-	6,494
	NYC	1,397	-	4,315	-	-	-	-	-	-	-	5,713
	RoPJM	-	4,306	-	14,696	-	107	-	-	-	-	19,109
	Total	4,520	10,302	13,204	14,796	-	616	-	12	-	-	43,450
Social Cost of CO ₂ and CH ₄ Leakage	MDDE	2,907	697	758	-	720	930	744	-	-	-	6,757
	NE	2,900	2,336	513	-	-	-	-	-	-	-	5,749
	NJ	2,734	-	788	-	682	1,879	609	-	-	-	6,691
	NY	3,503	1,580	-	-	-	-	-	-	-	-	5,084
	NYC	3,273	-	817	-	727	1,304	670	-	-	-	6,791
	RoPJM	4,833	-	239	-	-	-	-	-	-	-	5,071
	Total	20,150	4,613	3,114	-	2,129	4,113	2,022	-	-	-	36,143
Deterministic Health Damages	MDDE	2,362	791	742	-	70	-	750	1,558	-	-	6,273
	NE	3,695	4,415	859	-	-	-	-	-	-	-	8,969
	NJ	3,025	-	700	-	778	-	774	1,688	-	-	6,964
	NY	2,288	3,622	-	-	-	-	-	-	-	-	5,909
	NYC	3,010	-	700	-	-	-	638	2,033	-	-	6,381
	RoPJM	4,207	760	-	-	-	-	-	-	-	-	4,967
	Total	18,585	9,588	3,001	-	848	-	2,161	5,279	-	-	39,463
Simulated Health Damages	MDDE	865	5,153	739	-	1,105	992	729	-	-	-	9,584
	NE	1,355	7,655	-	-	-	651	901	911	-	-	11,473
	NJ	803	2,284	748	-	1,434	2,227	777	-	-	-	8,273
	NY	-	7,576	-	-	-	-	-	1,806	-	-	9,382
	NYC	699	-	1,529	-	1,433	3,282	909	-	-	-	7,852
	RoPJM	-	5,826	-	-	-	365	-	-	-	-	6,191
	Total	3,722	28,495	3,016	-	3,972	7,518	3,317	2,717	-	-	52,756

Table 5.4 summarizes the technology investments selected by the generation expansion planning model within our proposed network over for each of the cases studied. In the Baseline and Reserve Margin cases, we observe a similar investment strategy as discussed in Chapter 4. The Baseline case proposes investments in natural gas, coal, combined heat and power, and petroleum units. The Reserve Margin case also includes these investments, along with additional investments in combined cycle gas turbines.

In the cases where health damages are included in the objective function of the generation expansion planning model, we observe that coal is no longer included in the investment portfolio. This is consistent with the dispatching findings, as coal generation is displaced by mostly natural gas and combined cycle gas turbines. In the Social Cost of CO₂ and CH₄ Leakage scenario, investments are dominated by natural gas units. A similar observation is found in the Deterministic Health Damages case; however, this scenario proposes increased investments in wind (on land and off-shore). The Simulated Health Damages case proposes further increases in wind capacity by over 31 GW, with a reduction in natural gas generation relative to the Deterministic Health Damages case.

The increase in wind capacity proposed by the Simulated Health Damages case would bring the total wind capacity of the network to over 33 GW by 2040, which is a significant increase over the initial network capacity of approximately 2 GW. This translates to over 10% of the total electricity generation in this region. However, in 2008, the United States Department of Energy conducted a study investigating a scenario where 20% of the U.S. electricity needs are satisfied by wind energy by 2030. This translates to nearly 300 GW of required growth in the U.S. to achieve this goal (U.S. Department of

Energy, 2008). Relative to the modeling region, this translates to nearly 100 GW of wind capacity growth. This is a much more aggressive wind technology investment strategy, relative to the investment strategy proposed by the Simulated Health Damages case. However, the U.S. Department of Energy's research suggests that significant wind capacity expansions are feasible. Further research is required to fully assess the challenges and feasibility of the wind capacity expansions proposed in the Simulated Health Damages case.

5.3 Emissions and Human Health Results

Table 5.5 – CO₂ Emissions Summary

	Regional CO ₂ Emissions (Short Tons, Millions)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Total
Baseline	8,799	12,488	10,734	10,539	20,916	21,185	84,660
Reserve Margin	9,101	12,349	10,508	9,857	21,126	18,311	81,252
Emissions Limits & RPS Constraints	6,460	5,428	9,047	6,000	17,503	15,662	60,100
Social Cost of CO₂ and CH₄ Leakage	3,862	6,710	8,646	4,422	10,741	10,351	44,733
Deterministic Health Damages	3,361	6,648	8,635	3,052	10,028	9,418	41,142
Simulated Health Damages	1,280	1,734	6,291	1,791	8,604	8,422	28,121

Table 5.6 – NO_x Emissions Summary

	Regional NO _x Emissions (Short Tons, Thousands)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Total
Baseline	23,474	30,592	22,470	27,729	51,066	51,995	207,326
Reserve Margin	22,321	30,221	20,030	26,109	51,381	44,371	194,434
Emissions Limits & RPS Constraints	11,441	15,275	15,332	15,838	37,244	38,847	133,978
Social Cost of CO₂ and CH₄ Leakage	9,439	10,789	20,244	6,623	21,556	15,504	84,154
Deterministic Health Damages	11,402	10,112	20,226	4,571	23,485	14,107	83,903
Simulated Health Damages	5,873	5,683	17,866	2,683	24,733	14,518	71,357

Table 5.7 – SO₂ Emissions Summary

	Regional SO ₂ Emissions (Short Tons, Thousands)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Total
Baseline	50,860	58,934	32,047	59,037	97,867	100,448	399,192
Reserve Margin	43,079	58,132	21,803	56,064	97,886	84,047	361,011
Emissions Limits & RPS Constraints	9,158	29,440	9,468	33,858	55,190	75,307	212,420
Social Cost of CO₂ and CH₄ Leakage	9,501	4,183	18,994	390	14,721	912	48,700
Deterministic Health Damages	16,215	1,335	18,993	269	22,080	830	59,722
Simulated Health Damages	10,004	8,296	21,665	158	30,484	5,500	76,106

Tables 5.5 through 5.7 display emissions results for CO₂, NO_x, and SO₂ respectively for each of the six cases studied in this chapter. In comparing our emissions findings with the corresponding dispatching plans, as coal generation is displaced with other technologies, this leads to corresponding reductions in emissions relative to the Baseline results. Furthermore, as demonstrated by Case 2.3 (or the Emissions Limits and RPS case), the inclusion of emissions limits and renewable portfolio standards yields the largest emissions reductions in cases where only market costs are minimized in the objective function of the model.

Furthermore, in cases where health and other damages are included in the objective function, we observe significant reductions in CO₂, NO_x, and SO₂ emissions relative to the cases where only market costs are minimized. As previously observed in Chapter 4, this is due to the near-elimination of coal from the dispatch plans in these cases.

Table 5.8 – Expected Negative Health Outcomes Summary

	Expected Health Outcomes					
	Baseline	Reserve Margin	Emissions Limits & RPS	Social Cost of CO₂ and CH₄ Leakage	Deterministic Health Damages	Simulated Health Damages
Adult Mortality	114,848	94,430	52,118	40,105	36,572	30,362
Infant Mortality	118	115	65	53	48	37
Non-fatal Heart Attacks	50,575	54,980	36,343	28,436	30,021	23,976
Respiratory-Related Hospital Admissions	19,337	18,037	10,294	6,092	5,813	4,399
Cardiovascular-Related Hospital Admissions	23,682	22,838	16,802	9,890	8,586	6,103
Acute Bronchitis	79,420	58,717	37,036	27,797	23,531	14,206
Upper Respiratory Symptoms	1,778,084	1,498,650	818,666	599,061	627,854	453,905
Lower Respiratory Symptoms	1,386,847	1,269,842	844,169	738,350	704,750	478,534
Asthma ER Visits	35,740	34,728	23,935	16,048	14,974	10,939
Minor Restricted Activity Days	49,269,726	46,317,472	28,288,326	19,485,734	17,744,378	14,285,267
Work Loss Days	8,979,141	8,410,051	5,691,770	4,573,135	3,875,591	2,472,267
Asthma Exacerbations	1,949,020	1,980,709	1,288,379	1,060,425	872,455	617,597

Table 5.8 displays the expected number of negative health outcomes as a result of NO_x and SO₂ emissions, as modeled in the COBRA tool for each of the cases studied. Generally consistent with the findings in Chapter 4, negative health outcomes are reduced from the Baseline for all cases. However, in the cases where health damages are included in the objective function, we observe dramatic reductions in negative human health effects, which is driven by significant reductions in fossil fuel generation yielding reductions in pollutant emissions.

5.4 Cost and Pricing Results

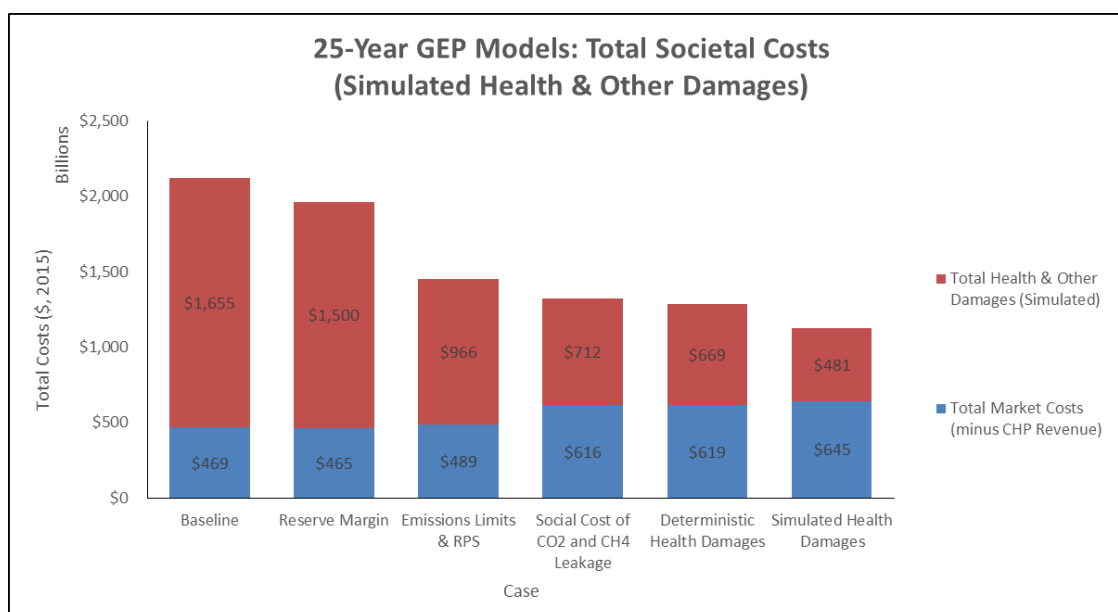


Figure 5.4 – Total Cost Summary (Including Simulated Health Damages)

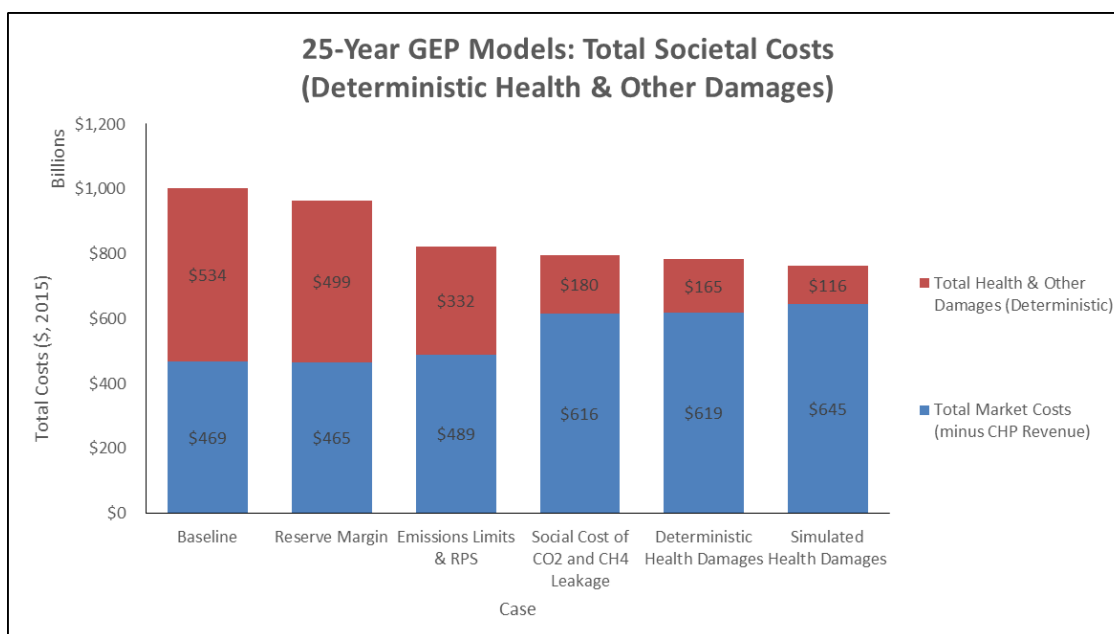


Figure 5.5 – Total Cost Summary (Including Deterministic Health Damages)

Figures 5.4 and 5.5 display the total objective function costs for each of the six cases presented in this chapter. Market costs, specifically investment costs, fixed O&M costs, and variable O&M costs (including fuel) are minimized in the Baseline, Reserve Margin, and Emissions Limits & RPS cases. Whereas, Total societal costs, including health and other damages (specifically the social cost of CO₂ and CH₄ leakage and human health externalities from NO_x and SO₂ emissions) are minimized in the Social Cost of CO₂ and CH₄ Leakage, Deterministic Health Damages, and Simulated Health Damages cases. Like the Chapter 4 cases, in the Deterministic Health Damages case, human health externalities from NO_x and SO₂ emissions are obtained from the values shown in Table 4.4; however, human health externalities from NO_x and SO₂ emissions in the Simulated Health Damages case are obtained via the COBRA simulation procedure presented in Chapter 3. As expected, in cases where only market costs are optimized, health damages are significantly greater than in cases where total societal costs are optimized.

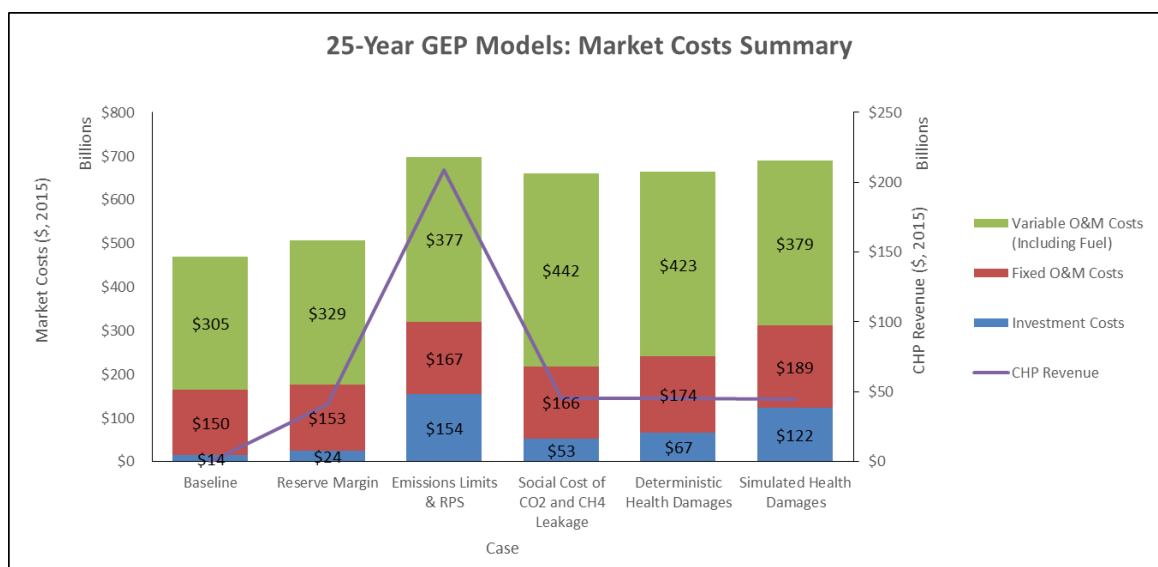


Figure 5.6 – Market Cost Summary

Figure 5.6 breaks down the market costs for each of the six cases presented in this chapter. As observed in the Chapter 4 examples, as we add more complexity to our generation expansion planning models, such as additional constraints or additional costs in the objective function, it drives increased investment and variable O&M costs (including fuel costs) relative to the baseline. This is driven by the displacement of coal generation by natural gas and combined cycle gas turbine dispatch. In general, minimizing total societal costs increases market costs, since the model selects investment and dispatching decisions that have the least impact from an emissions perspective, which consequently benefits society from a human health perspective.

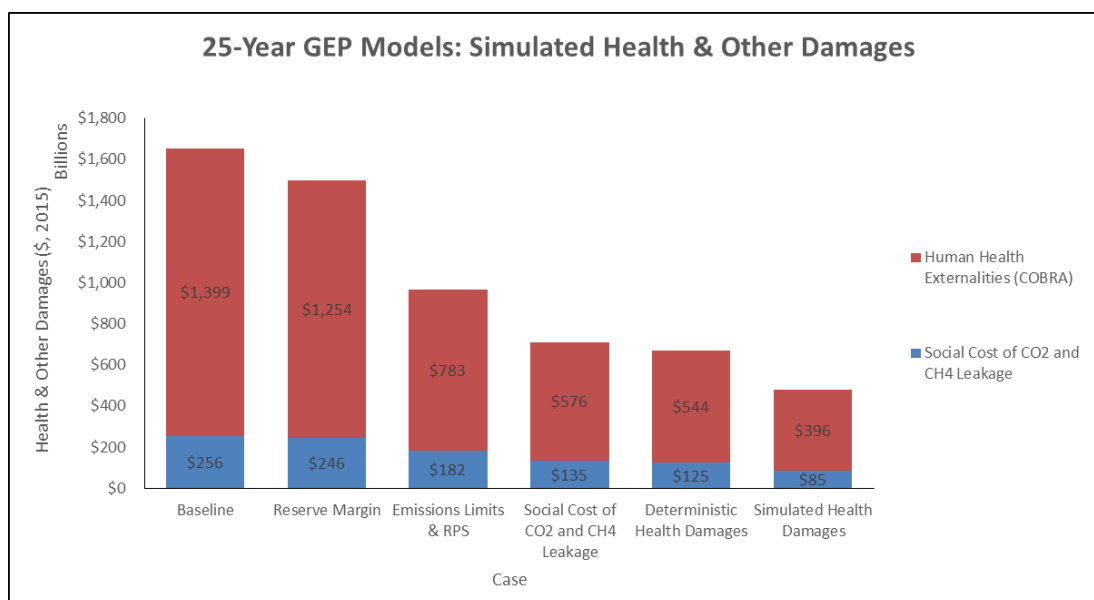


Figure 5.7 – Simulated Health Damages Summary

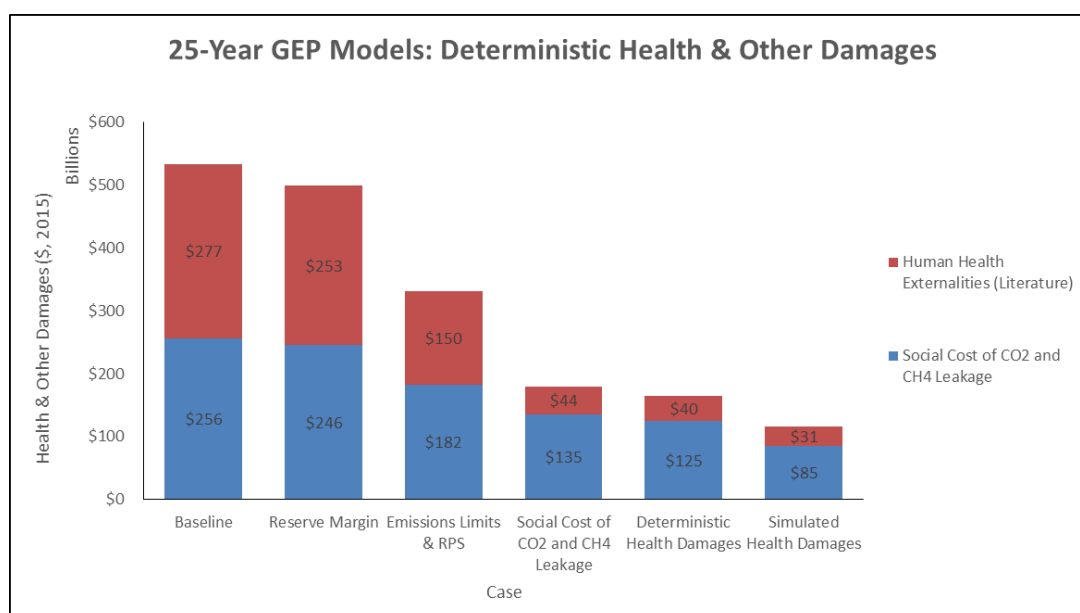


Figure 5.8 – Deterministic Health Damages Summary

Figures 5.7 and 5.8 summarize the simulated and deterministic health damages associated with the expansion plan selected in each of the cases studied. Similar to the observations in Chapter 4, in cases where only market costs are optimized, we observe that total health damages are greater than cases where total societal costs (including

health and other damages) are minimized in the objective function. In cases where health damages are minimized, this drives significant reductions in coal generation, which results in emissions reductions and reductions in health and other damages within the network.

Furthermore, human health externalities calculated via the COBRA simulation procedure outlined in Chapter 3 are generally much larger in magnitude in comparison to the deterministic approach. This is consistent with our findings in Chapter 4, which are due to the fact that human health externalities are highly non-linear functions of various factors, such as temperature, location, and emissions from other sources. Since COBRA is able to capture these relationships and approximate them statistically, the simulation procedure is able to monetize the health damages more effectively than simply applying deterministic multipliers to penalize pollutant emitting units.

Table 5.9 – Expected Human Health Externalities Summary

	Expected Human Health Externalities (\$ 2015, millions)					
	Baseline	Reserve Margin	Emissions & RPS	Social Cost of CO₂ and CH₄ Leakage	Deterministic Health Damages	Simulated Health Damages
Adult Mortality	\$1,373,137	\$1,232,899	\$769,867	\$566,167	\$534,164	\$387,921
Infant Mortality	\$1,336	\$1,103	\$689	\$545	\$538	\$408
Non-fatal Heart Attacks	\$14,882	\$12,469	\$8,207	\$5,852	\$6,019	\$4,595
Respiratory-Related Hospital Admissions	\$854	\$769	\$446	\$355	\$339	\$272
Cardiovascular-Related Hospital Admissions	\$592	\$522	\$300	\$219	\$234	\$193
Acute Bronchitis	\$45	\$39	\$24	\$19	\$16	\$13
Upper Respiratory Symptoms	\$12	\$9	\$6	\$4	\$5	\$4
Lower Respiratory Symptoms	\$57	\$44	\$28	\$20	\$19	\$13
Asthma ER Visits	\$23	\$18	\$11	\$10	\$8	\$7
Minor Restricted Activity Days	\$5,670	\$4,263	\$2,909	\$2,271	\$2,061	\$1,698
Work Loss Days	\$1,834	\$1,577	\$878	\$700	\$724	\$510
Asthma Exacerbations	\$150	\$130	\$78	\$60	\$60	\$47
Totals	\$1,398,592	\$1,253,843	\$783,442	\$576,222	\$544,188	\$395,681

Corresponding to the expected number of negative health outcomes displayed in Table 5.8, Table 5.9 displays the associated monetized value of the human health externalities as a result of NO_x and SO_2 emissions, as modeled in the COBRA tool for each of the cases studied. As expected, negative health outcomes are reduced from the Baseline for all cases, and in the cases where health damages are included in the objective function; we observe dramatic reductions in human health externalities driven by significant reductions in fossil fuel generation. Across all cases, the major contributor to human health externalities is the number of adult lives lost (or mortality), which decreases significantly as health damages are minimized as part of the objective function.

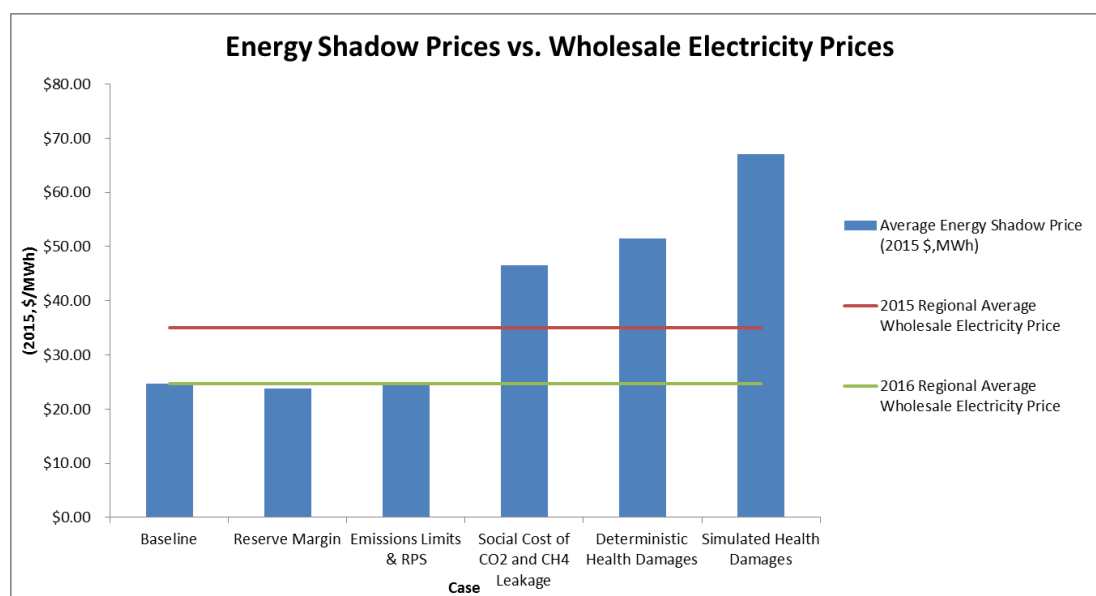


Figure 5.9 – Energy Comparison (U.S. Energy Information Administration, 2015)

As introduced in Chapter 4, the dual variables associated with the energy balance constraint are interpreted as the energy shadow price. In the context of electricity markets, this is the analog of wholesale electricity prices, which are derived from day-ahead unit commitment and real-time dispatch models. The results from the cases in this chapter are presented in Figure 5.9. The ultimate take-away from our analysis is that

minimizing societal costs in the objective function could potentially lead to increased energy prices. Extending this research to day-ahead markets would allow for further insight into this preliminary finding.

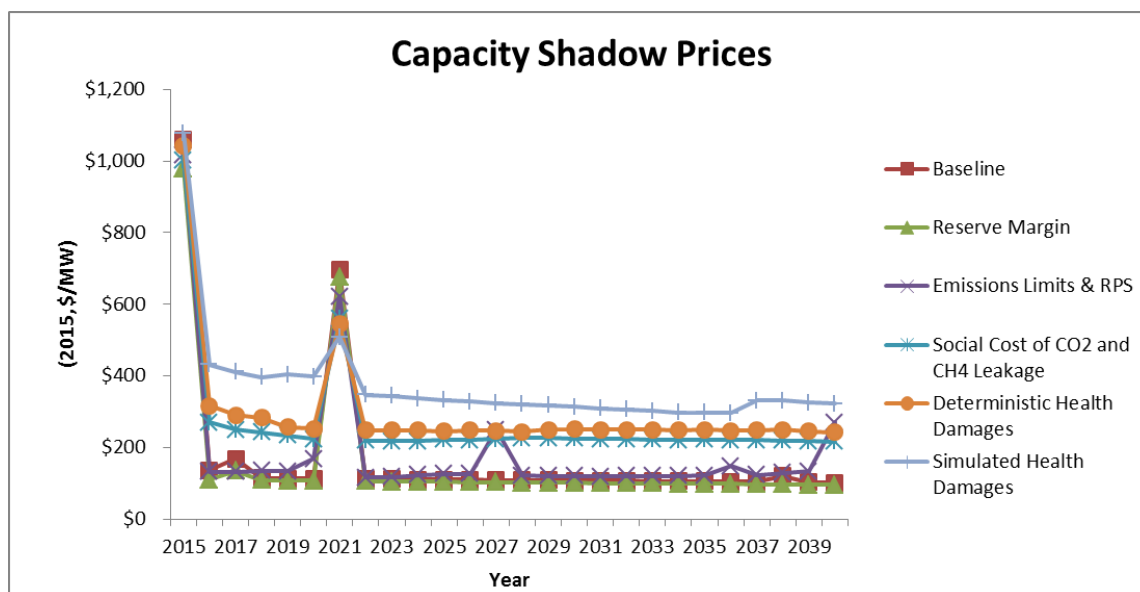


Figure 5.10 – Capacity Shadow Prices

Similar to the interpretation of the dual variables associated with the energy balance constraints, we quantify the dual variables associated with the capacity constraints, and refer to them as capacity shadow prices. That is, for each MW of capacity invested, there is an incremental \$/MW-year cost to the objective function value. The results shown in Figure 5.10 depict this concept for each of the cases studied. In general, based on our findings, the incremental cost of capacity expansion are much greater in the early stages of the planning process, relative to the remainder of the time-horizon.

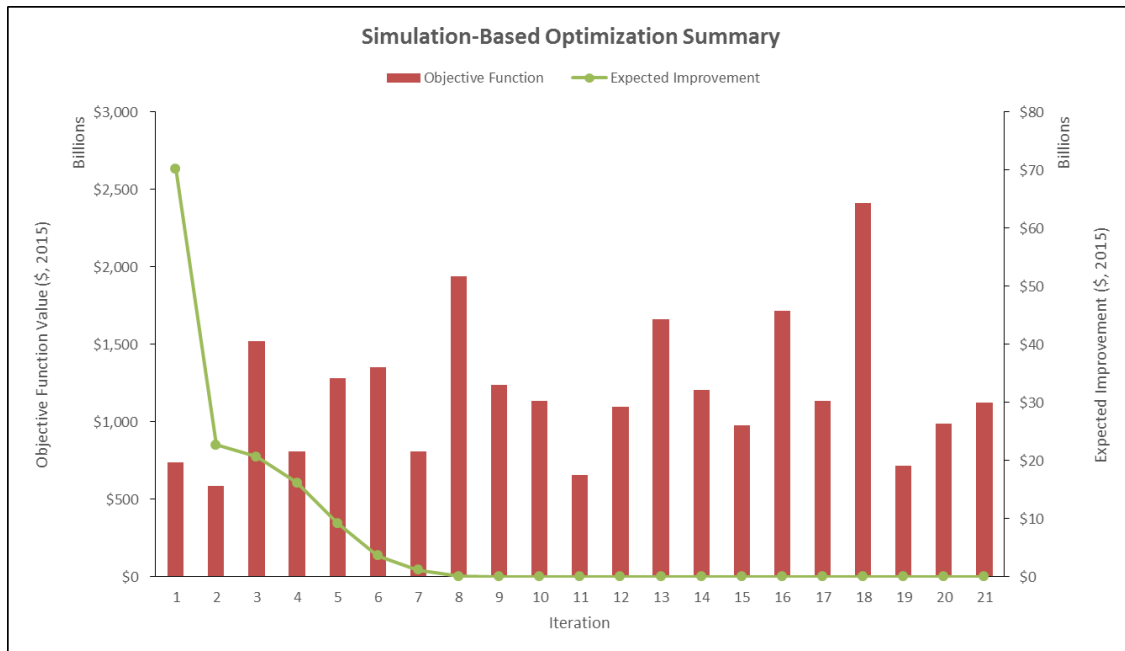


Figure 5.11 – Simulation-Based Optimization Iteration Summary

Case 2.6, the Simulated Health Damages case, applies the simulation-based optimization framework presented in Chapter 3 in order to solve the generation expansion planning problem that minimizes total societal costs, inclusive of market costs and health damages. Following our procedure in Chapter 3, as depicted in Figure 5.11, we are able to achieve a solution within 21 full iterations. In comparison to the reduced-form analogous case in Chapter 4, this example requires more iterations due to the increased complexity of the model stemming from the inclusion of seasonality and the extension of the time horizon to 25 years. As we progress through each of the iterations, the expected improvement value decreases and eventually reaches zero upon termination of the algorithm. The objective function, however, behaves rather erratically initially, but reaches an expected improvement value of zero at iteration 21, satisfying termination criteria.

5.5 Regional Details – Simulated Health Damages Case

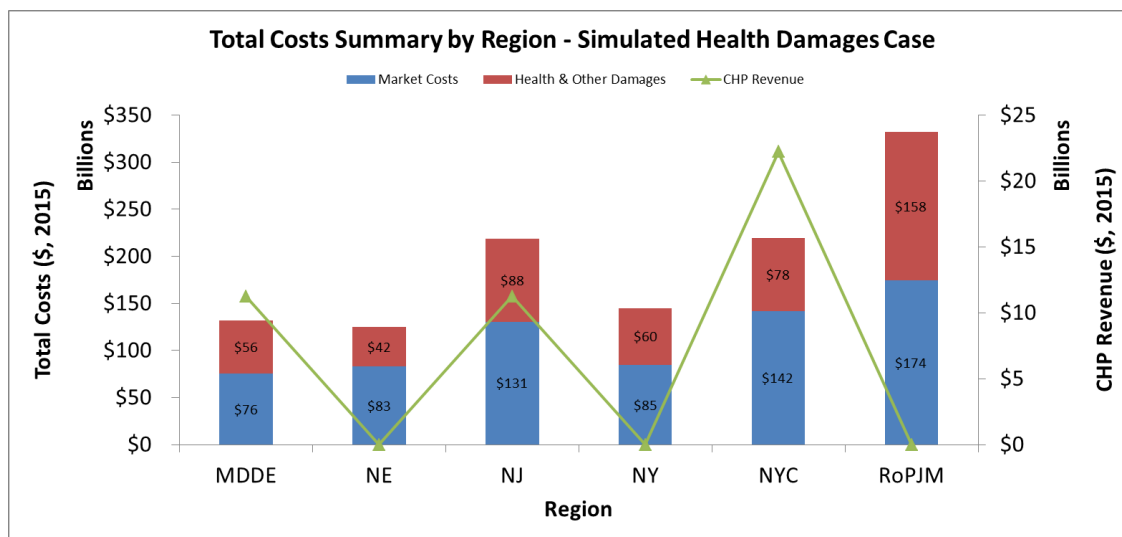


Figure 5.12 – Total Costs by Region

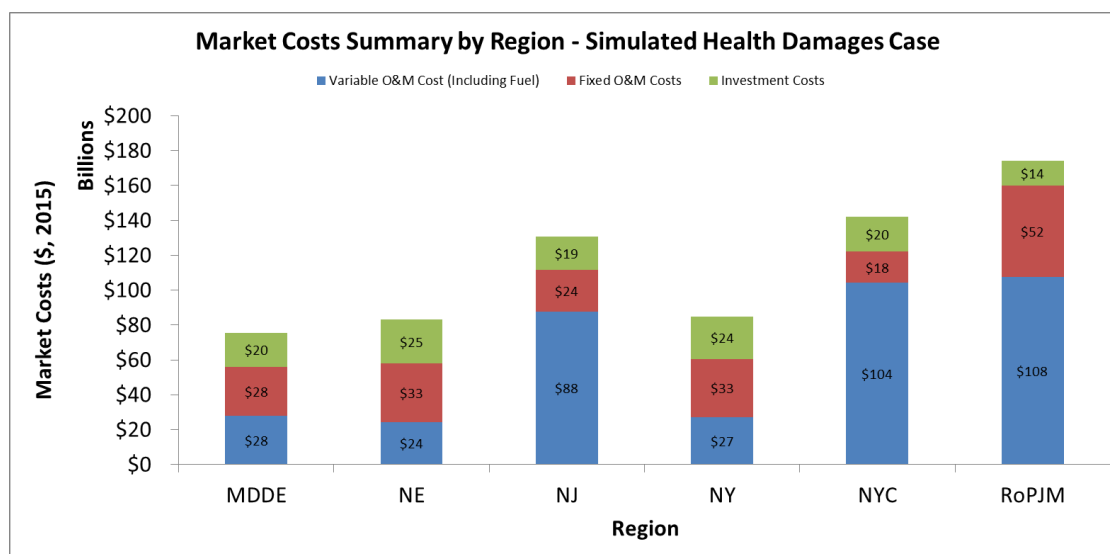


Figure 5.13 – Total Market Costs by Region

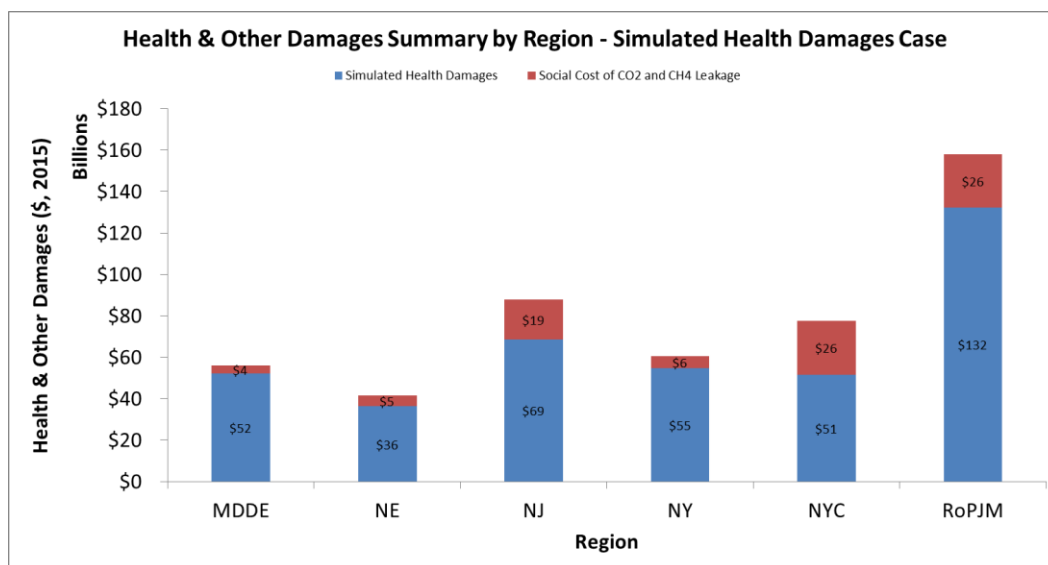


Figure 5.14 – Health & Other Damages by Region

Figures 5.12 through 5.14 display the cost summaries at the regional level for the Simulated Health Damages case. As observed by these figures, the Rest of PJM region, which is the largest geographical region in the model, is where most of the market costs and health damages are concentrated. Market costs in this region are driven by significant investments in wind technology. Additionally, significant combined heat and power technology investments drive revenue in the MDDE, NJ, and NYC regions. The social cost of CO₂ and CH₄ leakage is primarily driven by natural gas generation in the NYC and Rest of PJM regions.

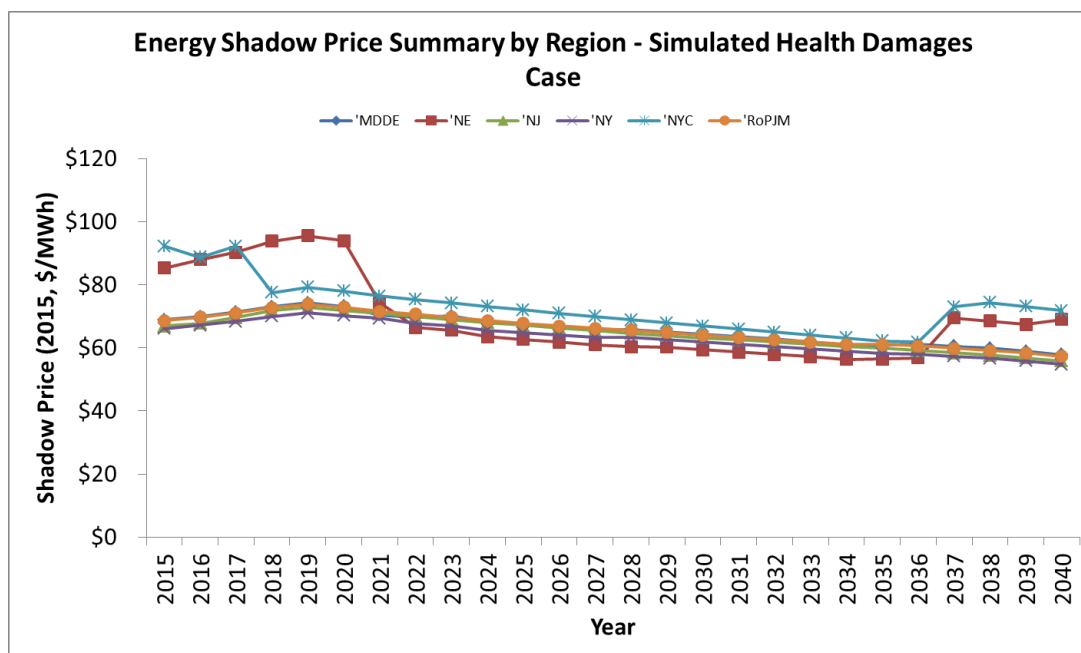


Figure 5.15 – Energy Shadow Prices by Region

Figure 5.15 displays the energy shadow prices for each of the regions considered in the network. In general, shadow prices in the NYC region tend to be slightly more expensive in comparison to the other regions. On average, energy shadow prices follow a slightly decreasing trend over the model duration. This is driven by investments made in the early stages of the model timeframe, where most of the capacity expansion investments are made.

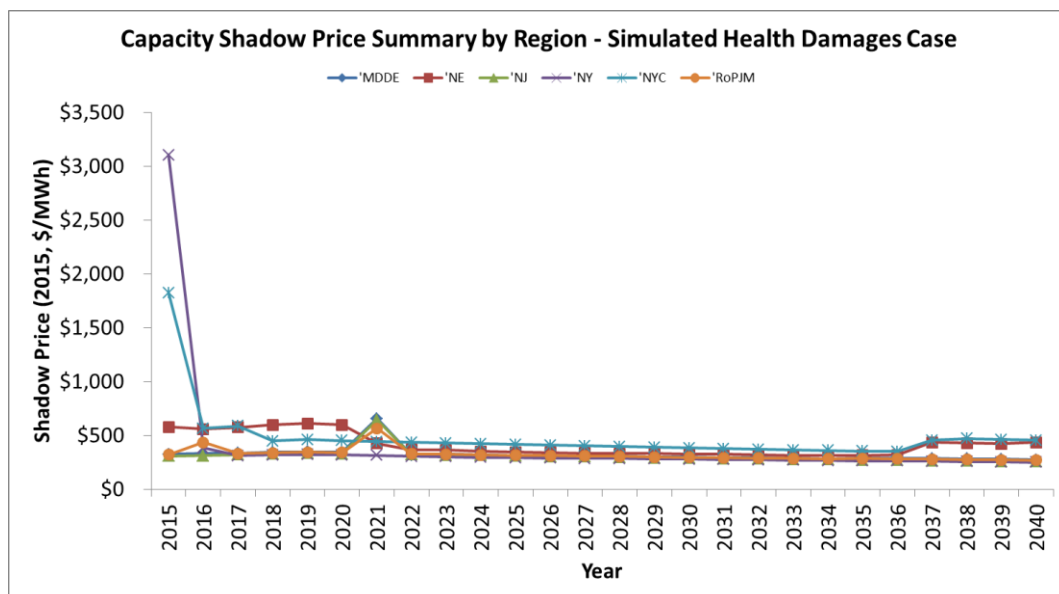


Figure 5.16 – Capacity Shadow Prices by Region

Figure 5.16 displays the capacity shadow prices for each of the regions considered in the network. As noted in Section 5.4, capacity shadow prices tend to be more expensive, on average, in the early years of the model timeframe. This is driven by capacity expansion investments in the NY and NYC regions in 2015.

Table 5.10 – Expected Health Outcomes

	Expected Health Outcomes (25 - Year Model)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Totals
Adult Mortality	3,881	2,491	4,883	3,870	4,188	11,049	30,362
Infant Mortality	4	3	6	5	5	13	37
Non-fatal Heart Attacks	3,476	2,084	4,455	3,157	3,092	7,712	23,976
Respiratory-Related Hospital Admissions	554	440	736	622	586	1,461	4,399
Cardiovascular-Related Hospital Admissions	818	578	1,210	767	821	1,910	6,103
Acute Bronchitis	1,840	1,525	2,583	1,811	2,119	4,327	14,206
Upper Respiratory Symptoms	53,350	32,807	82,936	64,550	62,868	157,394	453,905
Lower Respiratory Symptoms	58,513	46,019	87,649	64,751	61,143	160,459	478,534
Asthma ER Visits	1,347	959	2,005	1,826	1,402	3,400	10,939
Minor Restricted Activity Days	2,074,357	1,374,869	2,185,917	2,139,288	2,011,177	4,499,660	14,285,267
Work Loss Days	292,272	291,009	447,670	345,864	328,832	766,621	2,472,267
Asthma Exacerbations	86,014	55,945	102,590	80,289	75,708	217,051	617,597

Table 5.11 – Expected Human Health Externalities

	Expected Health Outcomes (25 - Year Model)						
	MDDE	NE	NJ	NY	NYC	RoPJM	Totals
Adult Mortality	\$51,080.0	\$35,543.5	\$67,359.4	\$53,719.4	\$50,477.7	\$129,741.0	\$387,921.0
Infant Mortality	\$53.7	\$37.4	\$70.8	\$56.5	\$53.0	\$136.3	\$407.6
Non-fatal Heart Attacks	\$605.1	\$421.1	\$798.0	\$636.4	\$598.0	\$1,536.9	\$4,595.4
Respiratory-Related Hospital Admissions	\$35.8	\$24.9	\$47.3	\$37.7	\$35.4	\$91.0	\$272.2
Cardiovascular-Related Hospital Admissions	\$25.4	\$17.7	\$33.5	\$26.7	\$25.1	\$64.5	\$192.9
Acute Bronchitis	\$1.7	\$1.2	\$2.2	\$1.7	\$1.6	\$4.2	\$12.6
Upper Respiratory Symptoms	\$0.5	\$0.3	\$0.6	\$0.5	\$0.5	\$1.2	\$3.6
Lower Respiratory Symptoms	\$1.8	\$1.2	\$2.3	\$1.9	\$1.7	\$4.5	\$13.4
Asthma ER Visits	\$0.9	\$0.6	\$1.2	\$0.9	\$0.9	\$2.3	\$6.8
Minor Restricted Activity Days	\$223.6	\$155.6	\$294.9	\$235.2	\$221.0	\$568.1	\$1,698.5
Work Loss Days	\$67.2	\$46.8	\$88.6	\$70.7	\$66.4	\$170.7	\$510.4
Asthma Exacerbations	\$6.2	\$4.3	\$8.2	\$6.5	\$6.1	\$15.7	\$47.0
Totals	\$52,101.8	\$36,254.6	\$68,706.9	\$54,794.0	\$51,487.5	\$132,336.5	\$395,681.3

Tables 5.10 and 5.11 display the expected regional health outcomes and the associated externalities as a result of the expansion plans proposed in the Simulated Health Damages case. The major driver of human health externalities in the network is Adult Mortality in the Rest of PJM region, which accounts for nearly one-third of the total human health externalities in the network. This is primarily due to the fact that most of the proposed dispatch of combined cycle gas turbine technology is concentrated in this region.

5.5.1 Maryland/Delaware (MDDE) Regional Summary

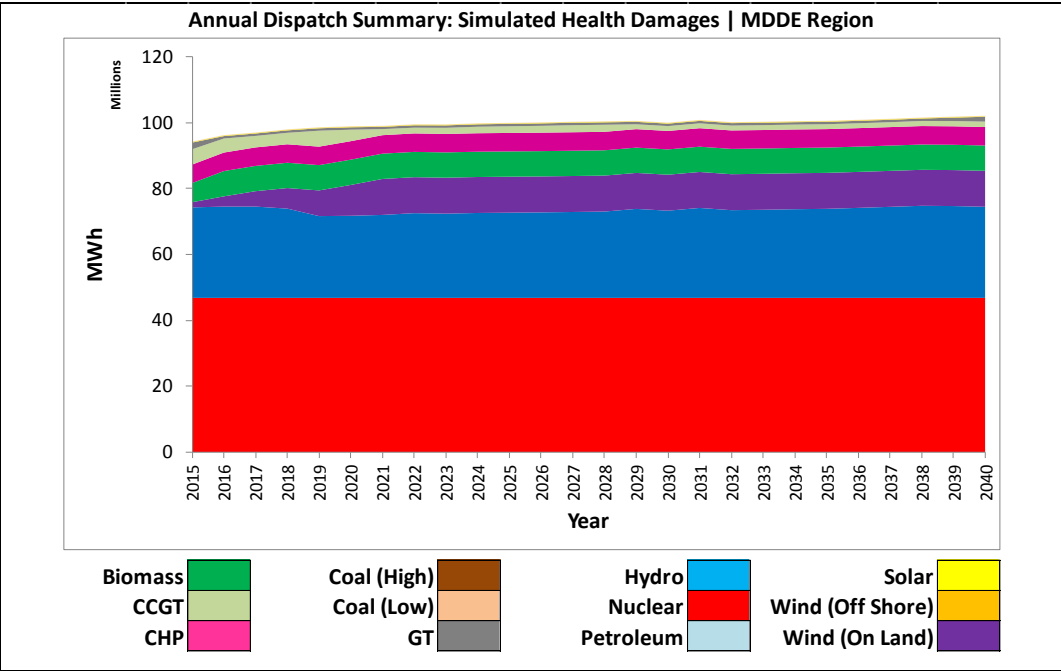


Figure 5.17 – Annual Dispatch Summary – MDDE Region

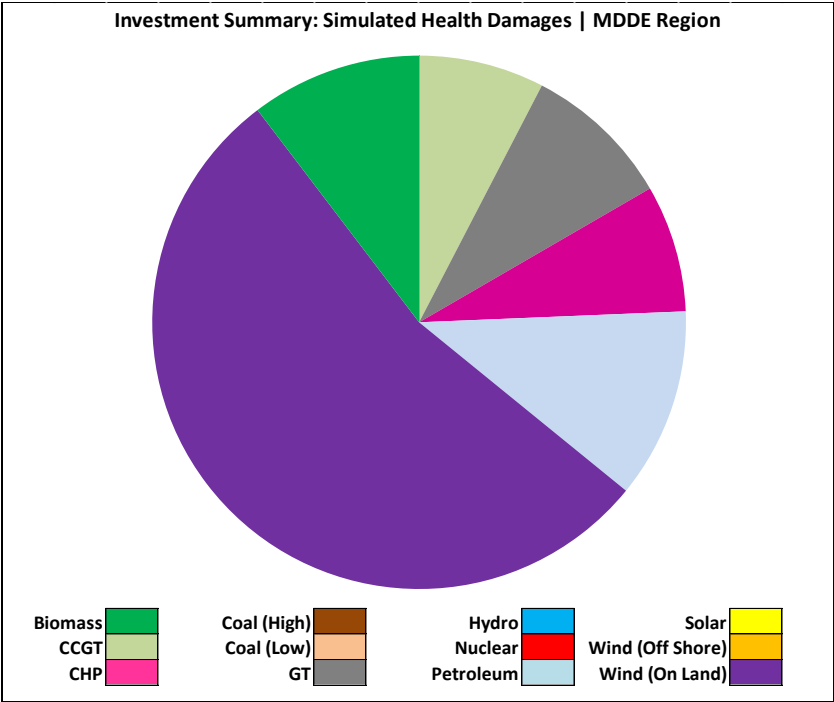


Figure 5.18 – Summary of New Investments – MDDE Region

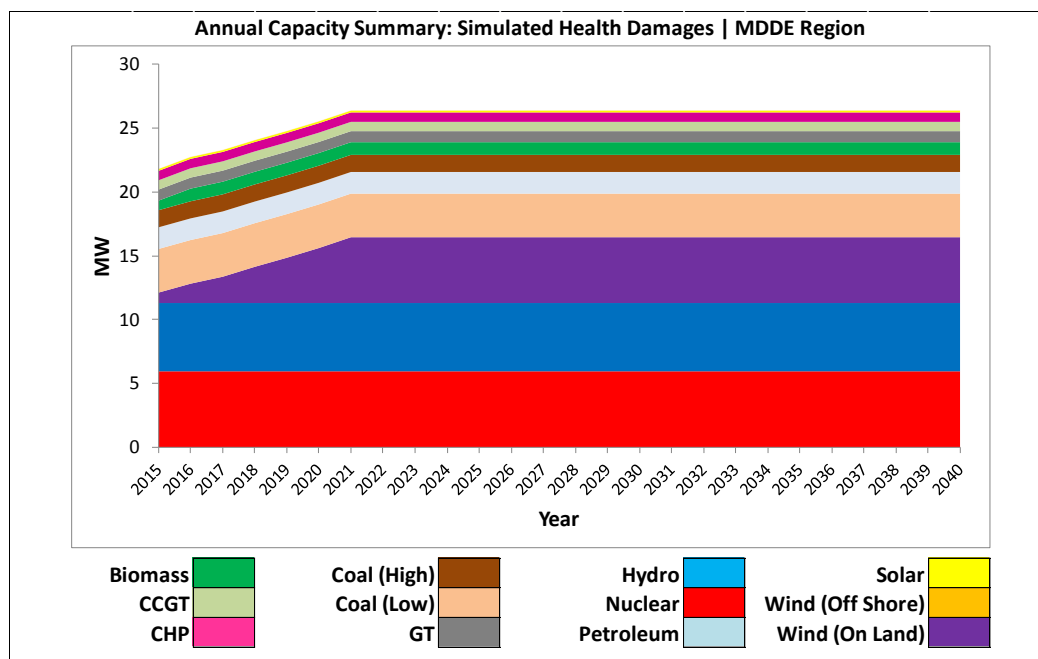


Figure 5.19 – Annual Capacity Summary – MDDE Region

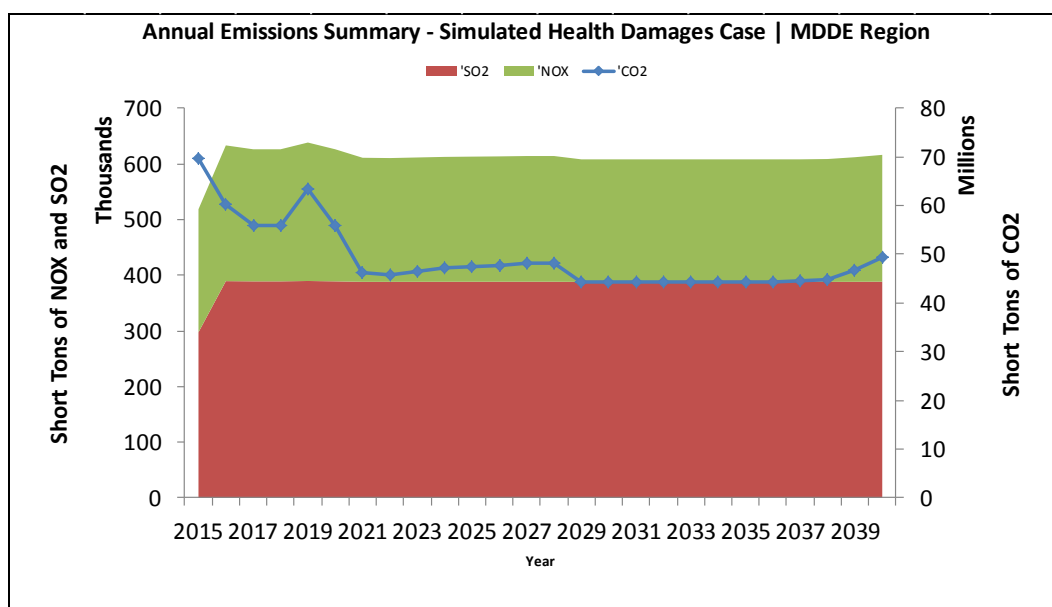


Figure 5.20 – Annual Emissions Summary – MDDE Region

Figures 5.17 through 5.20 display the dispatching, investment, capacity, and emissions results for the MDDE region. In this region, the dispatch plan is heavily dominated with nuclear and hydro generation, with wind (on land) and biomass

generation satisfying an increasing portion of the load over time. The model proposes a diversity of investments in biomass, combined cycle gas turbines, natural gas, petroleum, and combined heat and power, but the investment plan is mainly dominated by wind (on land) investments. From an emissions perspective, CO₂ emissions tend to decrease throughout the model, whereas NO_x and SO₂ emissions are relatively constant in this region.

5.5.2 New England (NE) Regional Summary

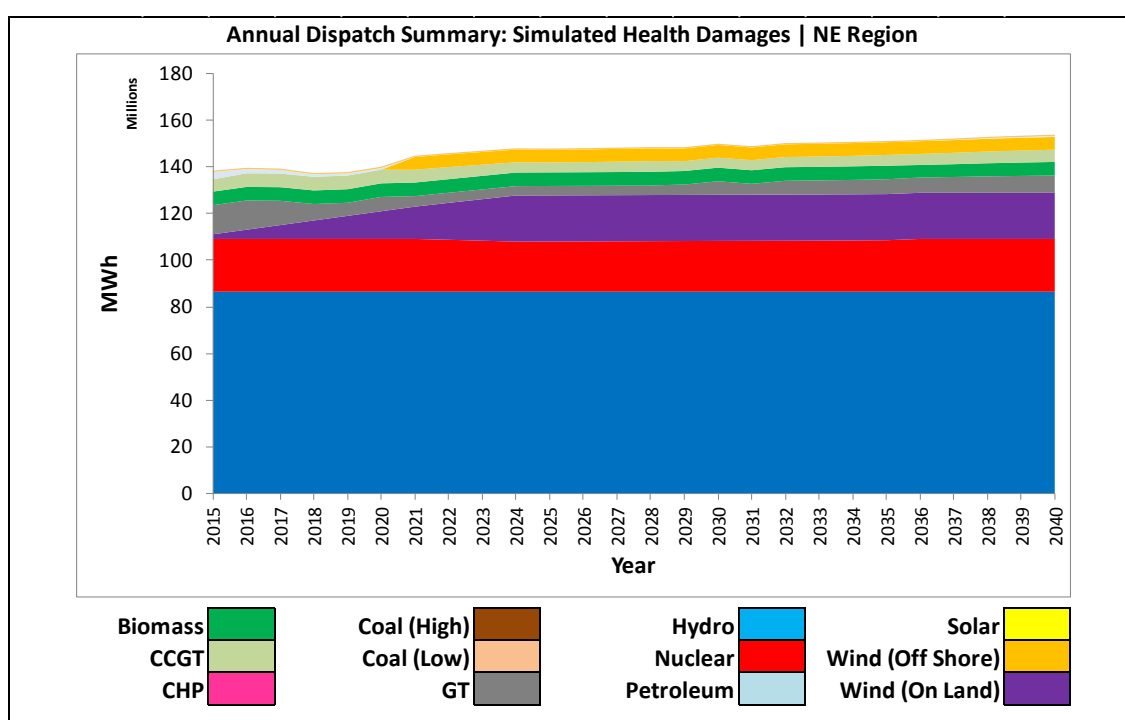


Figure 5.21 – Annual Dispatch Summary – NE Region

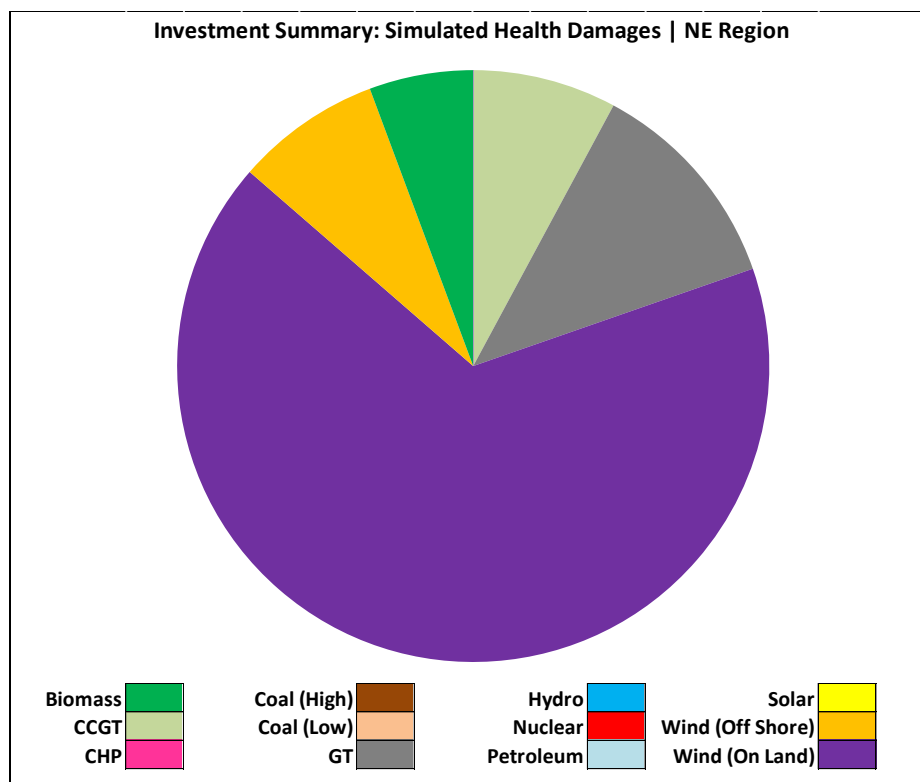


Figure 5.22 – Summary of New Investments – NE Region

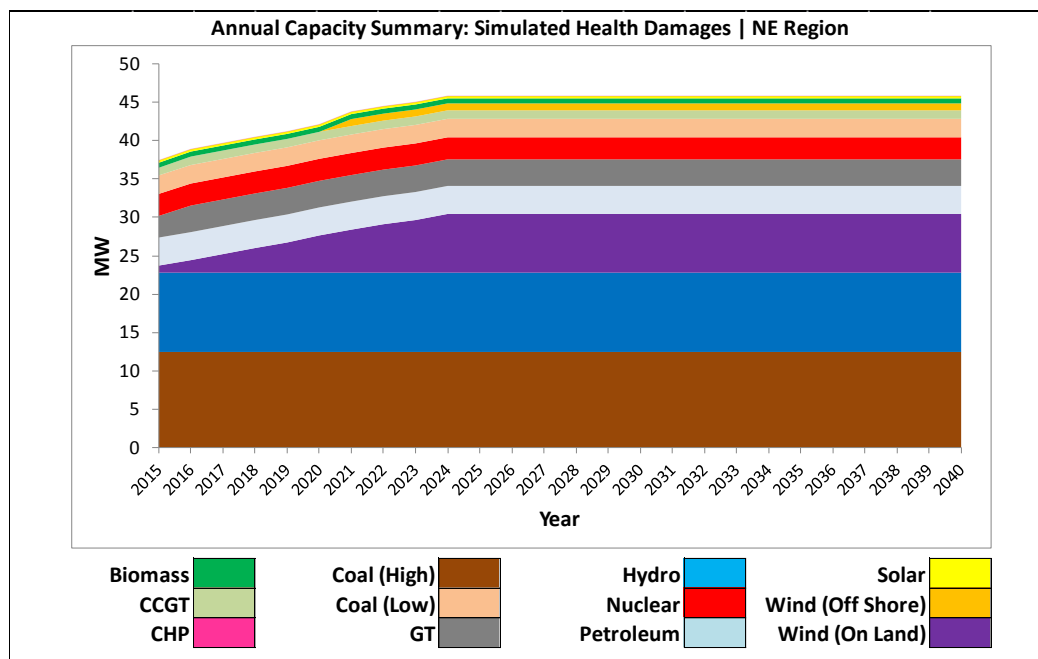


Figure 5.23 – Annual Capacity Summary – NE Region

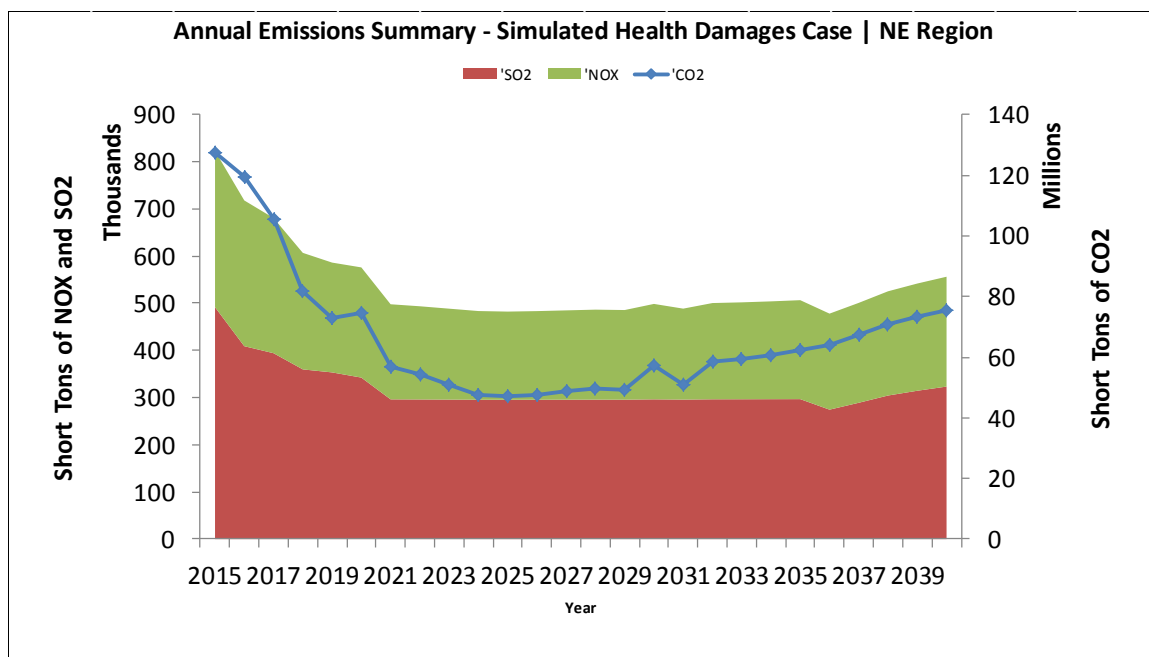


Figure 5.24 – Annual Emissions Summary – NE Region

Figures 5.21 through 5.24 display the dispatching, investment, capacity, and emissions results for the NE region. In this region, the dispatch plan is heavily dominated with hydro and nuclear generation, with increasing amounts of wind (on land) generation satisfying a portion of the load over time. The increase in wind (on land) generation coincides with corresponding capacity expansion investments. From an emissions perspective, CO₂, NO_x and SO₂ emissions tend to decrease throughout the model, with a slight increase toward the end of the time horizon.

5.5.3 New Jersey (NJ) Regional Summary

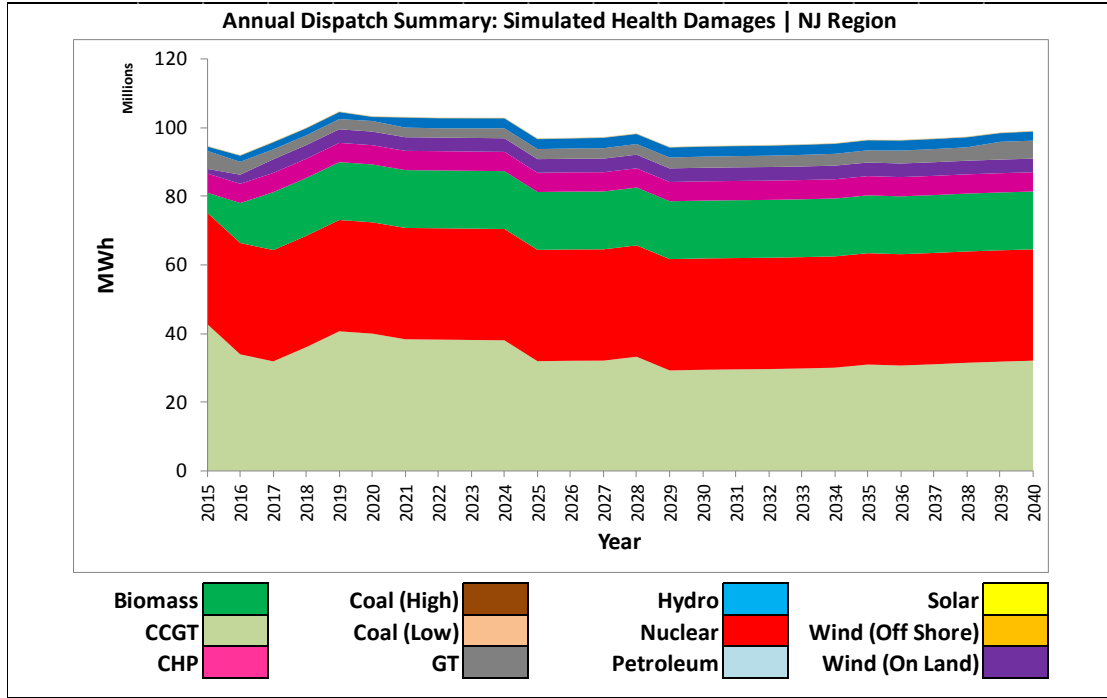


Figure 5.25 – Annual Dispatch Summary – NJ Region

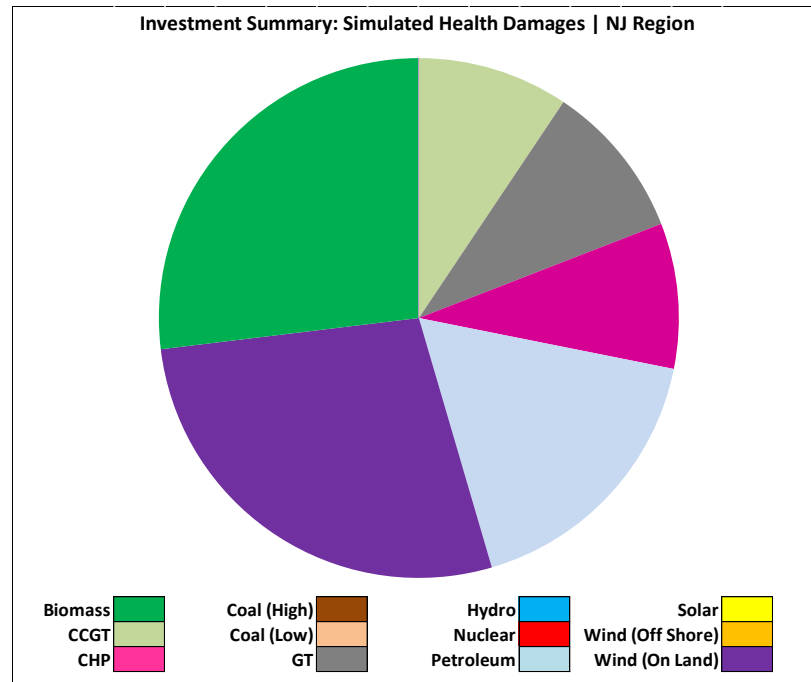


Figure 5.26 – Summary of New Investments – NJ Region

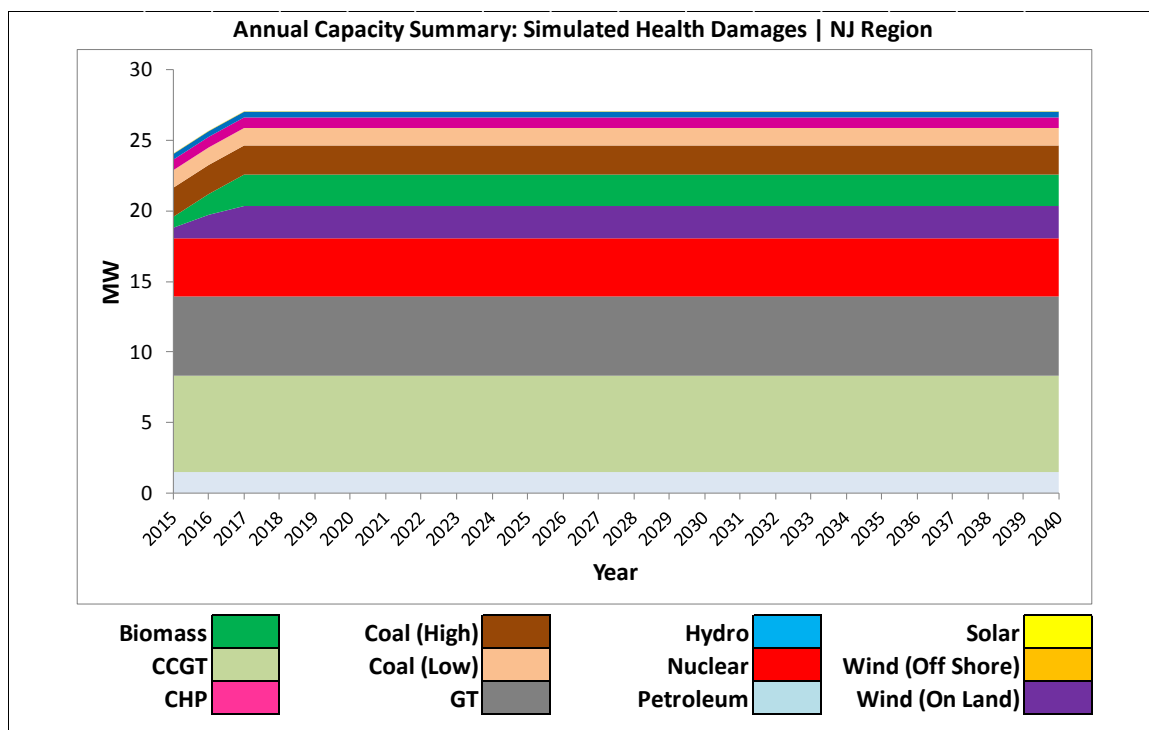


Figure 5.27 – Annual Capacity Summary – NJ Region

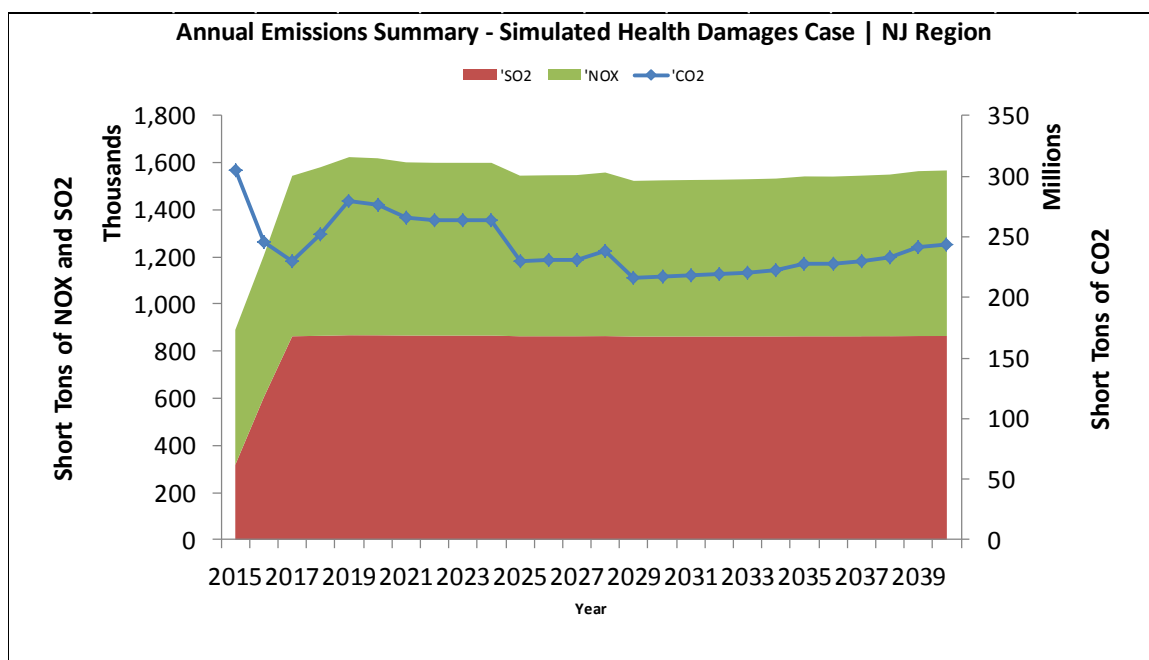


Figure 5.28 – Annual Emissions Summary – NJ Region

Figures 5.25 through 5.28 display the dispatching, investment, capacity, and emissions results for the NJ region. In this region, the dispatch plan is heavily dominated

with combined cycle gas turbine and nuclear generation, with increasing amounts of biomass generation satisfying a portion of the load over time. The increase in biomass generation coincides with corresponding capacity expansion investments. Significant investments are also made in wind (on land) capacity expansions. From an emissions perspective, CO₂ emissions tend to decrease throughout the model, whereas NO_x and SO₂ emissions are relatively constant in this region.

5.5.4 New York (NY) Regional Summary

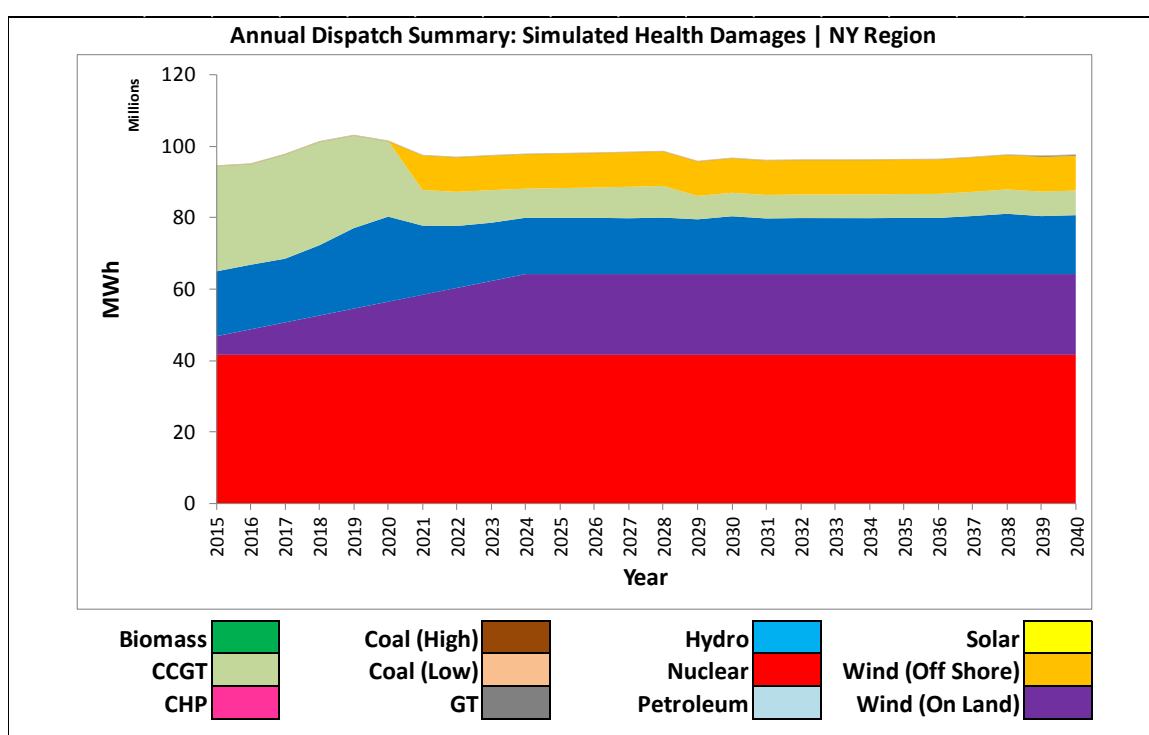


Figure 5.29 – Annual Dispatch Summary – NY Region

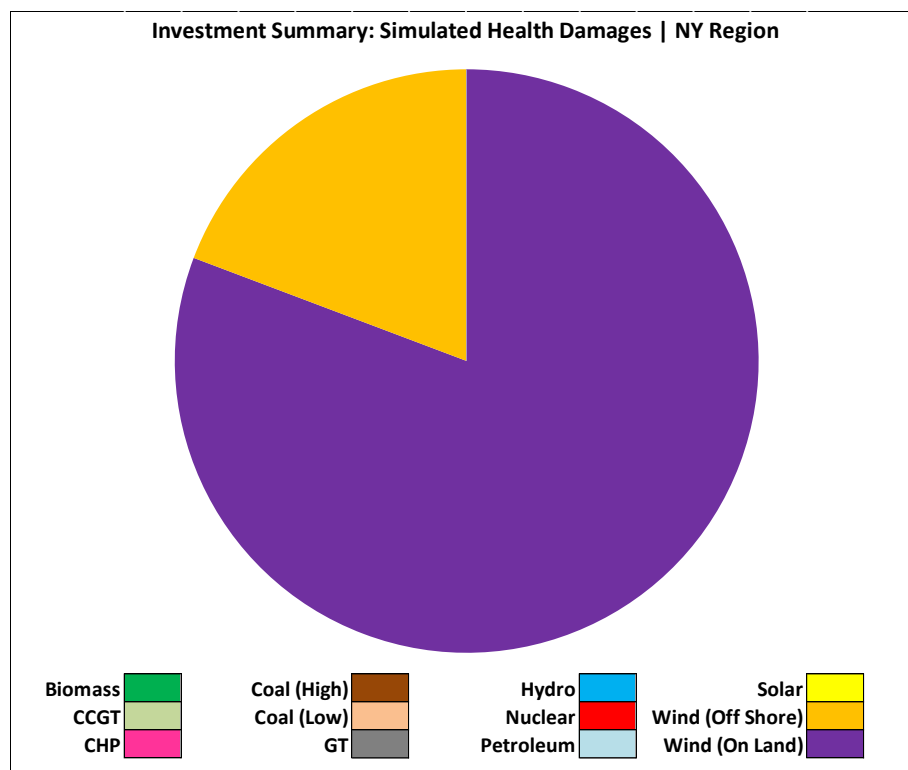


Figure 5.30 – Summary of New Investments – NY Region

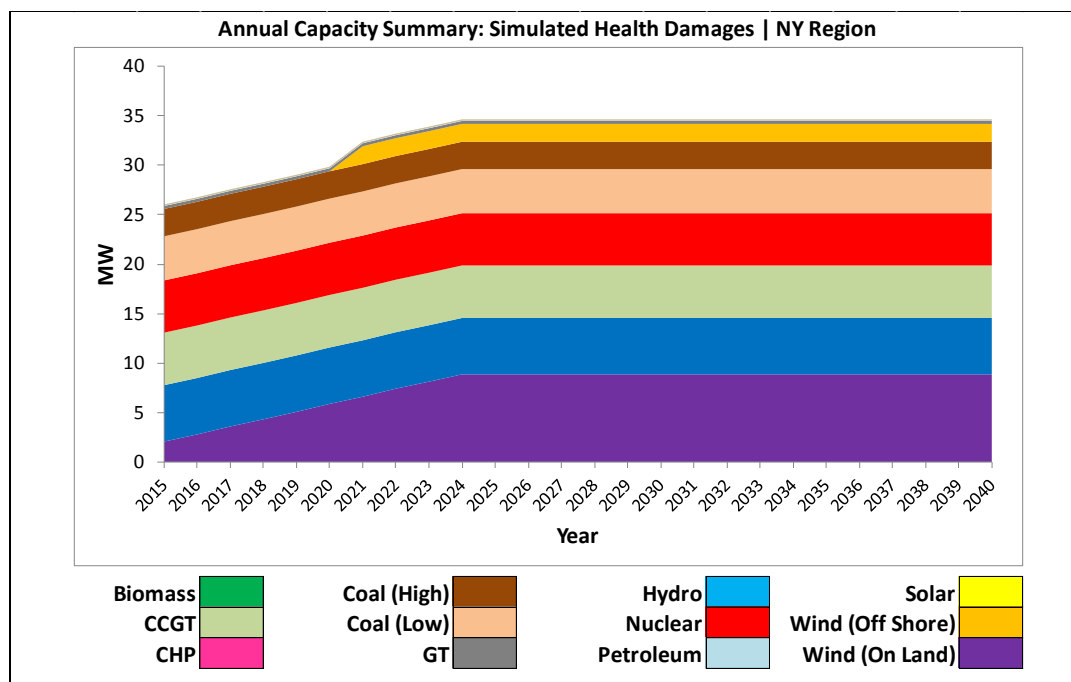


Figure 5.31 – Annual Capacity Summary – NY Region

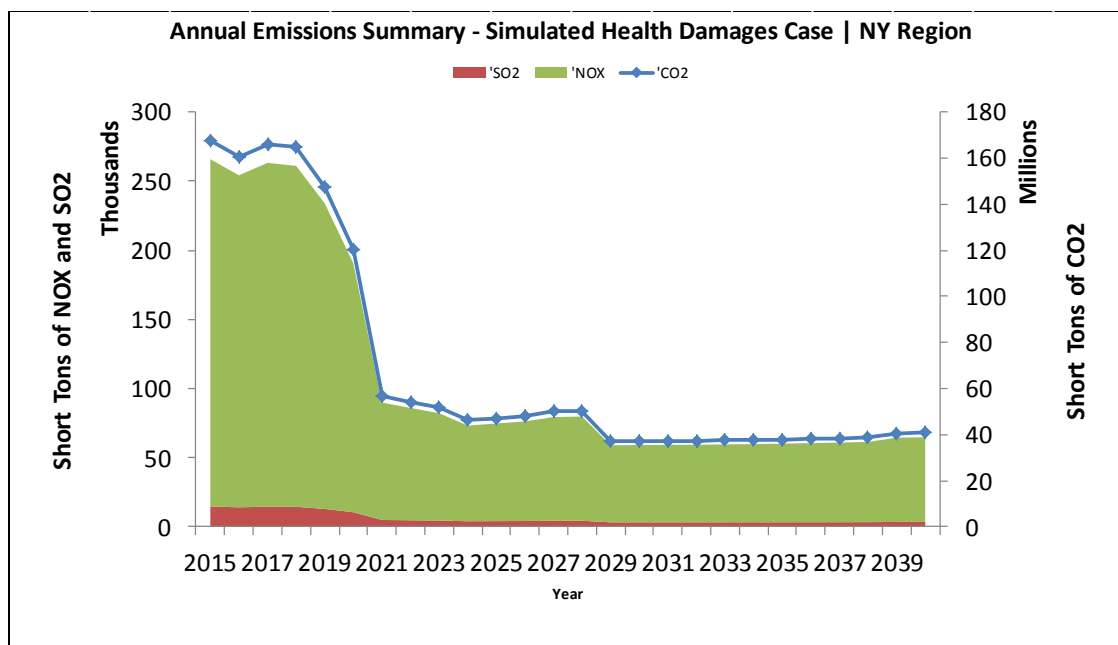


Figure 5.32 – Annual Emissions Summary – NY Region

Figures 5.29 through 5.32 display the dispatching, investment, capacity, and emissions results for the NY region. In this region, the dispatch plan is heavily dominated with nuclear and hydro, with increasing amounts of wind (on land) generation satisfying a portion of the load over time. The increase in wind (on land) generation coincides with corresponding capacity expansion investments. Significant investments are also made in wind (off shore) capacity expansions. From an emissions perspective, CO₂, NO_x and SO₂ emissions tend to decrease throughout the model, with a slight increase toward the end of the time horizon.

5.5.5 New York City (NYC) Regional Summary

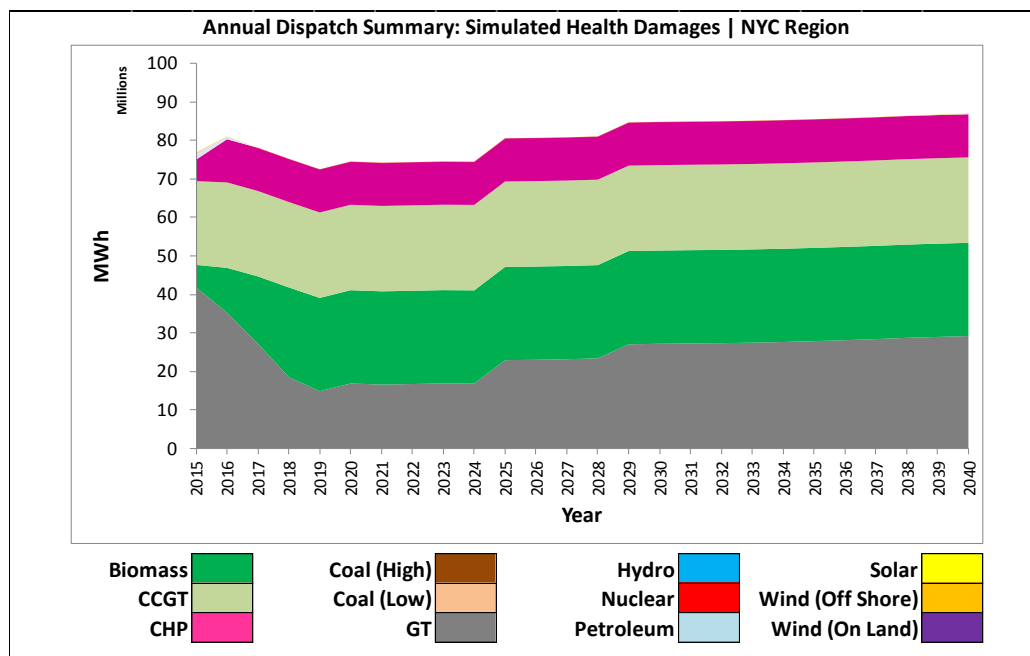


Figure 5.33 – Annual Dispatch Summary – NYC Region

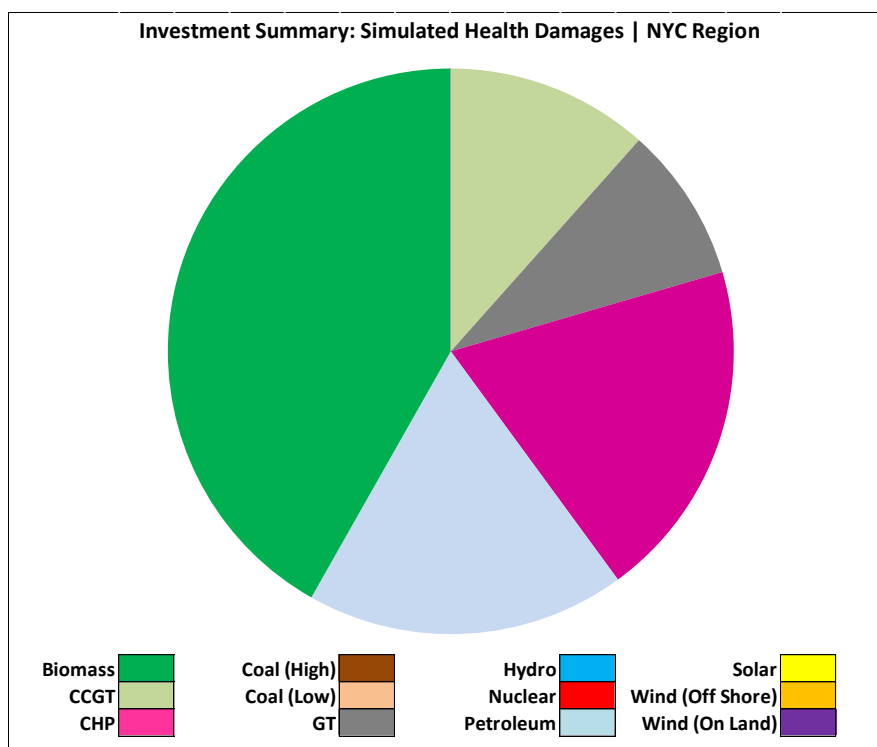


Figure 5.34 – Summary of New Investments – NYC Region

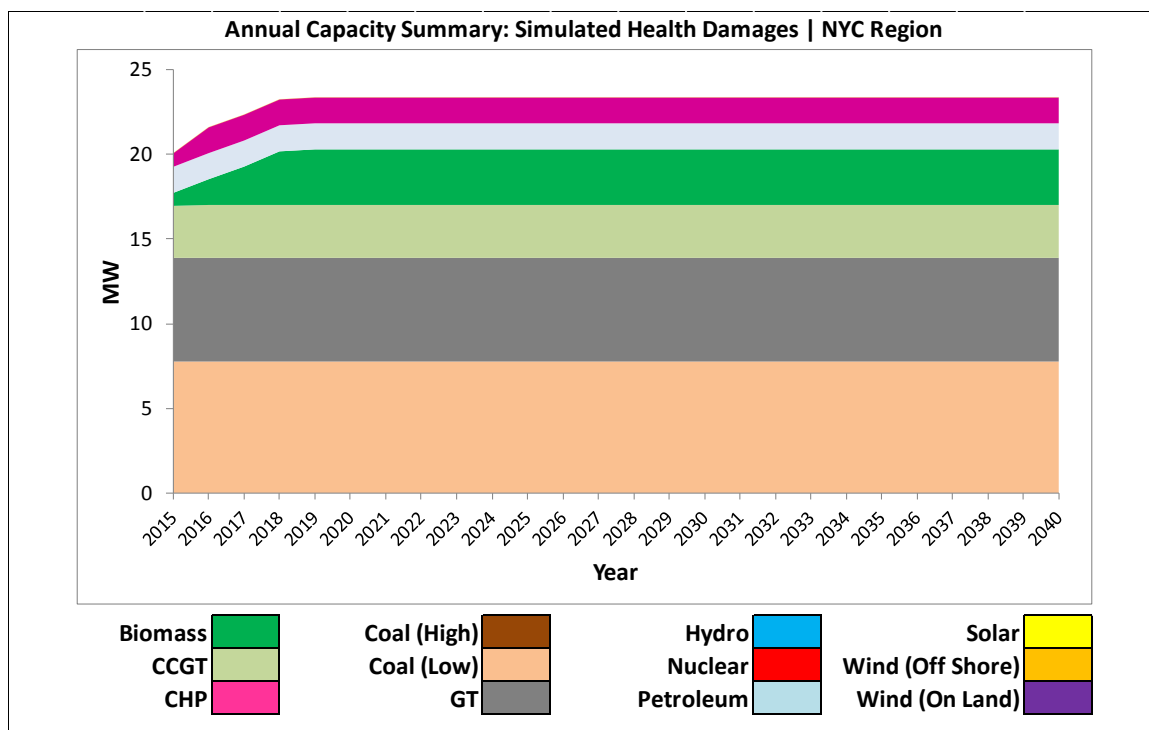


Figure 5.35 – Annual Capacity Summary – NYC Region

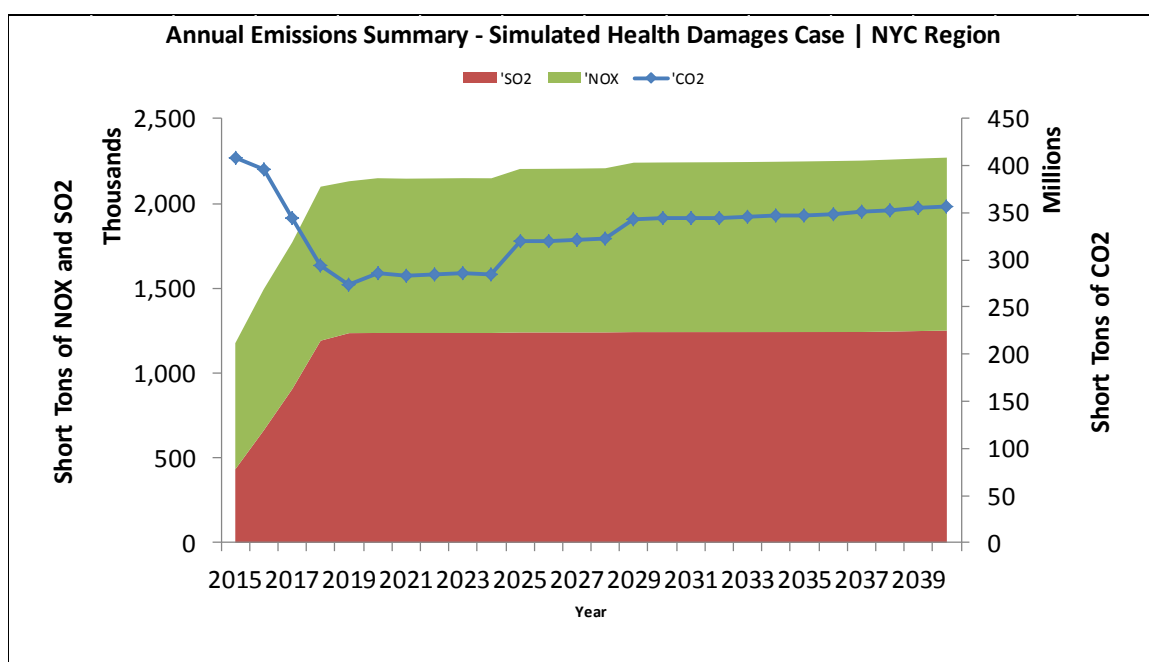


Figure 5.36 – Annual Emissions Summary – NYC Region

Figures 5.33 through 5.36 display the dispatching, investment, capacity, and emissions results for the NYC region. In this region, the dispatch plan is heavily

dominated with natural gas and combined cycle gas turbines, with increasing amounts of biomass generation satisfying a portion of the load over time. The increase in biomass generation coincides with corresponding capacity expansion investments. Additional investments are made in combined heat and power, petroleum, natural gas, and combined cycle gas turbines. From an emissions perspective, CO₂ emissions tend to decrease throughout the model, whereas NO_x and SO₂ emissions are relatively constant in this region.

5.5.6 Rest of PJM (RoPJM) Regional Summary

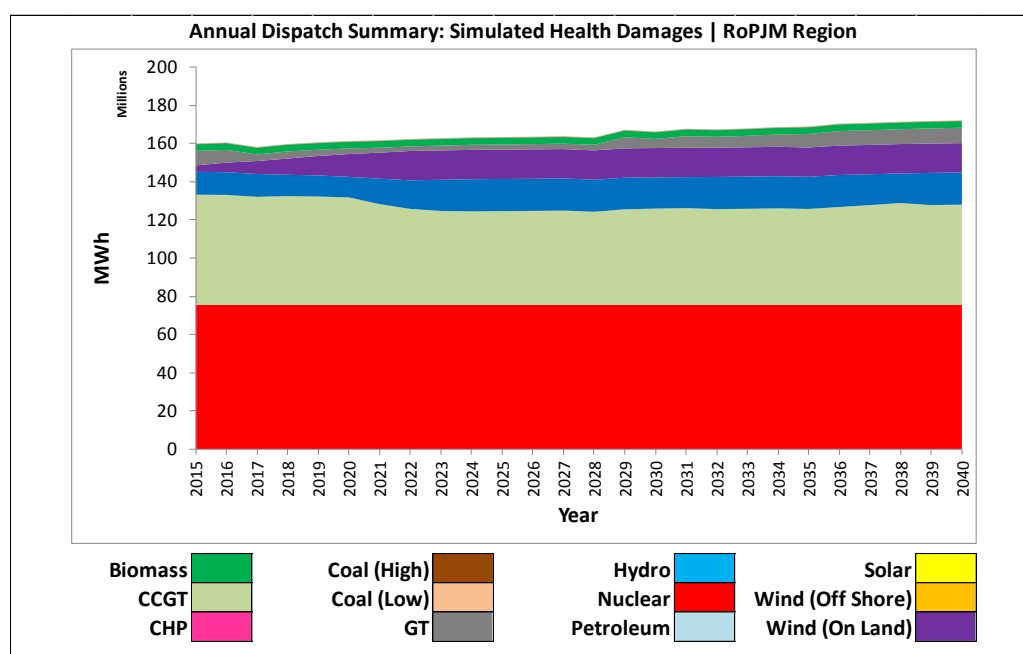


Figure 5.37 – Annual Dispatch Summary – RoPJM Region

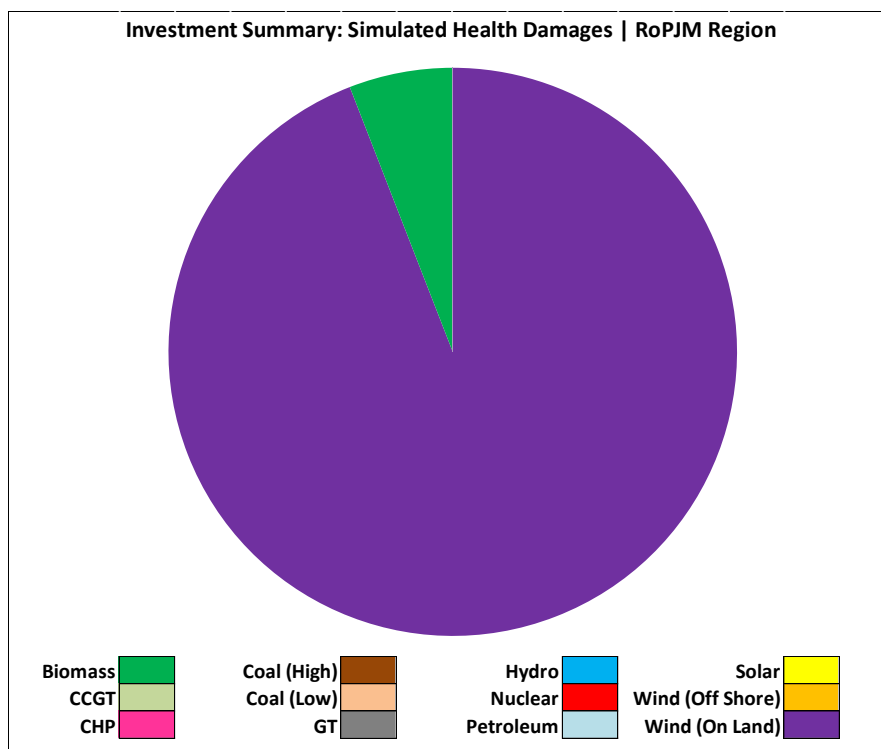


Figure 5.38 – Summary of New Investments – RoPJM Region

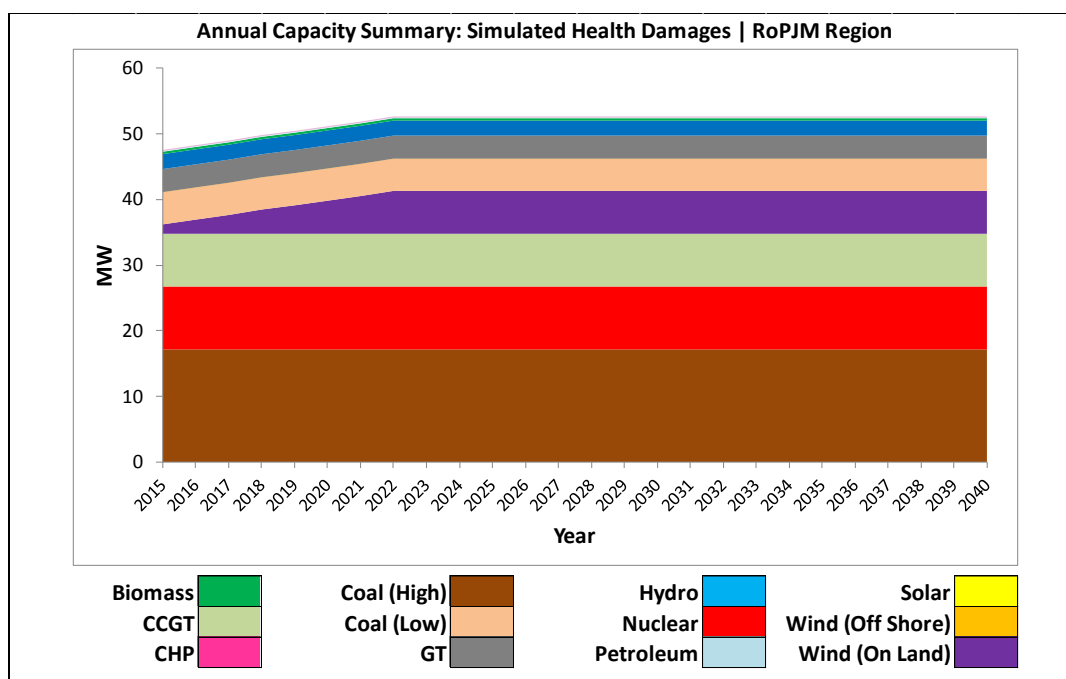


Figure 5.39 – Annual Capacity Summary – RoPJM Region

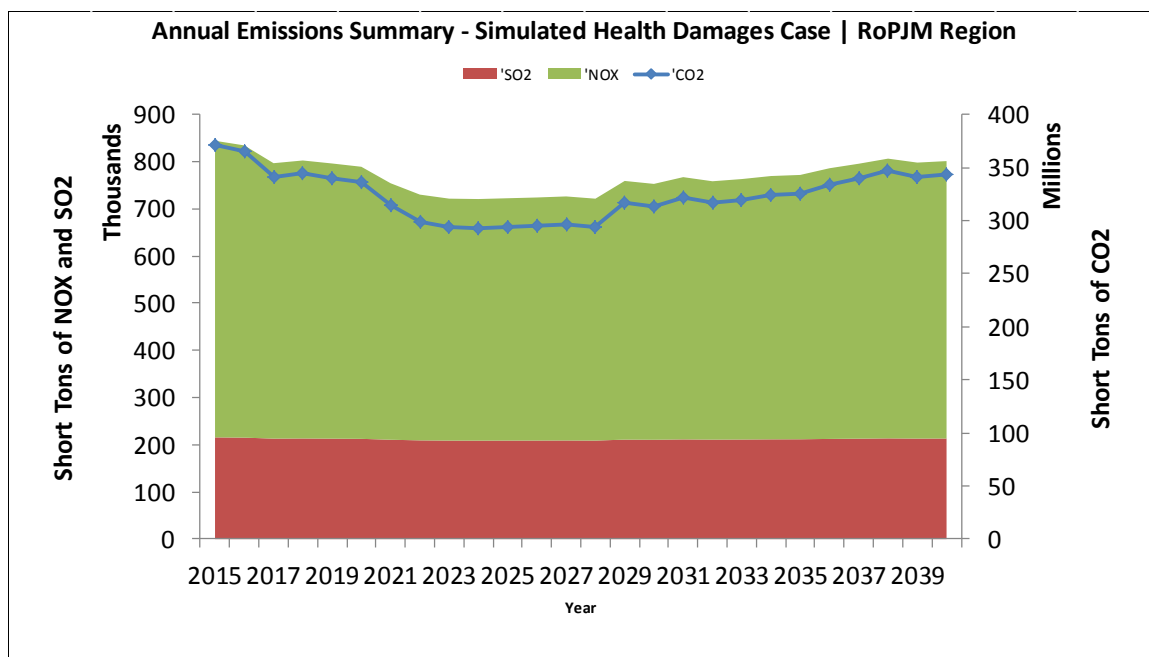


Figure 5.40 – Annual Emissions Summary – RoPJM Region

Figures 5.37 through 5.40 display the dispatching, investment, capacity, and emissions results for the RoPJM region. In this region, the dispatch plan is heavily dominated with nuclear and combined cycle gas turbines, with increasing amounts of wind (on land) generation satisfying a portion of the load over time. The increase in wind (on land) generation coincides with corresponding capacity expansion investments. Additional investments are made in biomass capacity as well. From an emissions perspective, CO₂, NO_x, and SO₂ emissions are relatively constant in this region.

6. Conclusions and Research Extensions

This dissertation presents a modeling framework to solve the generation expansion planning problem that minimizes total costs, inclusive of market costs (including investment costs, fixed and variable operating & maintenance costs, and fuel costs) and health damages (including the social cost of carbon and methane leakage and human health externalities) associated with pollutant emissions from electricity generation via simulation-based optimization methods. Additionally, the proposed expansion plans presented in Chapters 4 and 5 yield better results, from a human health perspective, than traditional generation expansion planning models. Specifically, traditional approaches, such as implementing emissions limits and renewable portfolio standards as constraints (see Cases 1.3 and 2.3 from Chapters 4 and 5 respectively) or applying deterministic multipliers as penalties in the objective function (see Cases 1.5 and 2.5 from Chapters 4 and 5 respectively) are not as effective at reducing the health effects from electricity generation in comparison to the analytical framework presented in this research.

This is the first comprehensive attempt at using advanced statistical and simulation methods to quantify the human health externalities as a function of generation expansion planning decision variables. Furthermore, it is also the first application of a simulation-based optimization in the context of the problem domain. Ultimately, the framework presented in this dissertation allows policy makers to make more informed decisions on expanding power grid capacity. Since existing generation expansion planning models do not include health damages in the objective function, the resulting expansion plans include fossil fuels. While these energy sources can economically (in

terms of market costs) and reliably satisfy the load on the system, they also produce emissions that are potential health hazards. In the context of generation expansion planning, these health hazards are not modeled in sufficient detail to address their highly stochastic, non-linear behavior. Because of this issue, these costs, in the form of externalities, are absorbed by residents. This research addresses this concern from an analytical perspective.

As previously mentioned, this work is the first step in addressing the deficiency in electric power systems planning research of including health damages in the generation expansion planning problem via simulation-based optimization methods. With that said, in order to advance this research further, we must consider extending this work to contribute to the field of power systems planning. Based on our framework, we can assess the impact of additions to our model such as:

- Intermittency of renewable sources of energy;
- Systems reliability;
- Assessments of expansion plans using sophisticated air quality models.

The first research extension addresses the variability of the output of renewable sources of energy, such as wind and solar technologies. In order to fully assess the intermittency of renewable generation, we may consider the following optimization approaches: *(i)* assuming the capacity factors of these units to be random variables and applying stochastic optimization approaches to obtain a solution, *(ii)* applying a simulation procedure to estimate the capacity factors of renewable units and using these values as surrogate capacity factors in a deterministic optimization model, or *(iii)* assuming that capacity factors are seasonal and deterministically incorporating this curve

in an optimization model. This extension is particularly critical to this research since the expansion plan presented in Chapter 5 proposes significant wind capacity expansions within the network. Addressing the intermittency of wind, as well as other renewables, would enable the subsequent dispatching plan to be implemented.

The second research extension is addresses the networks ability to satisfy demand reliably. The reliability of electric power systems is primarily defined by the interruptions to electricity service or the inability to maintain the appropriate reserve margin levels. While there are many approaches to quantifying and evaluating system reliability, across the nation, the utility industry has often used probabilistic metrics. Generally, these metrics rely on the use of Monte Carlo simulation to estimate future system performance under uncertainty in system parameters (such as fuel prices, load, and unit availability). In the literature, the most commonly used reliability metrics are Loss of Load Probability (LOLP), Loss of Load Expectation (LOLE), and Expected Unserved Energy (EUE). Incorporating these metrics as constraints in the generation expansion planning model ensures that reliability requirements are satisfied by a given expansion plan.

Loss of Load Probability (LOLP) quantifies the probability that there is at least one instance where load is not met over the specified time horizon. Since a probability must assume values between 0 and 1, the LOLP can be estimated by taking ratio of the number of simulations in which a shortfall occurs to the total number of simulations. Such an analysis is typically performed at the hourly level over the course of an entire year. A generalization of how LOLP is estimated is given in Equation 6.1,

$$LOLP = \frac{\sum_{i=1}^N S_i}{N} \quad (6.1)$$

where s_i is a simulation trial with at least one simulation in which there is at least one shortfall in the system, and N is the number of Monte Carlo simulations in the time period (Wood & Wollenberg, 1996).

Loss of Load Expectation (LOLE) is defined as the expected number of time units in a given time horizon, where the load exceeds the available system capacity. This metric is not a measure of frequency of individual events, since it's possible for multiple events to occur during a day, and also, a single event can last longer than a day. Similar to the LOLP metric, LOLE can be limited to counting shortfall events that exceed a minimum capacity, as well as energy or duration thresholds. Generally, LOLE is calculated as the number of days in which a shortfall occurs divided by the total number of years simulated. A typical formula for LOLE is given in Equation (6.2),

$$LOLE = \frac{N \times Y \times \sum_{i=1}^D d_i}{N_y} \quad (6.2)$$

where, d_{ijk} is a day in which at least one significant shortfall event occurs, N is the number of simulation trials for the time period (usually at 1 year at the hourly level), Y is the number of years in the study, N_y is the total number of years in the study, and D is the number of days in each year that are simulated (Wood & Wollenberg, 1996).

While LOLP and LOLE are closely associated with the frequency aspect of reliability, neither metric provides any information about the magnitude of shortfalls. Thus, the Expected Unserved Energy (EUE) metric provides a measure of the unserved energy demand in a given time horizon, which quantifies the magnitude of shortfall events. EUE is calculated by summing all unserved energy demand over all hours of the simulation divided by the total number of hours simulated. This calculation is given in Equation (6.3),

$$EUE = \frac{N \times Y \times D \sum_{i=1}^H e_i}{N_h} \quad (6.3)$$

where, e_i the amount of unserved energy in a given hour, H is the number of hours in each day being simulated, D is the number of days in each year that are simulated, Y is the number of years in the study, N is the number of simulation trials for the time period (usually at 1 year at the hourly level), and N_h is the total number of hours in the study (Wood & Wollenberg, 1996).

The third research extension allows for an assessment of air quality and human health externalities by systematically linking sophisticated air quality and economic models used by the EPA to our generation expansion planning model. Specifically, in place of the COBRA model, which is a screening tool used by the EPA, utilizing the EPA's SMOKE (Sparse Matrix Operator Kernel Emissions) model, which is an emissions processing system that allocates emissions both spatially and temporally, and linking the output of this model to the CMAQ (Community Multi-scale Air Quality) model, computes the pollutant concentrations by using the continuity equations (The Institute for the Environment - The University of North Carolina at Chapel Hill, 2012). The economic health implications of the associated pollutant concentrations can be assessed in the EPA's BenMAP (Environmental Benefits Mapping and Analysis Program) tool, which is used to quantify the human health externalities as a function of air quality effects (Abt Associates Inc., 2012). Systematically linking these EPA models to our generation expansion planning model via the framework presented in Chapter 3 not only gives decision makers a more detailed perspective of the health implications of expansion plans, but could further improve optimization results from a human health perspective.

As demonstrated by the results in this dissertation, the inclusion of health damages have a significant impact on the resulting expansion plans. Our proposed framework allows can help decision makers assess the impacts of their power grid expansion decisions based on their system constraints. Furthermore, under the circumstances presented in our framework, including these costs in the objective function proposes significant increases in wind capacity investments within the network. Although, research from the U.S. Department of Energy suggest that such expansion are feasible, additional investigation and research in the context of generation expansion planning would be required to fully address feasibility concerns along other challenges associated with wind capacity expansions.

In conclusion, this research makes research contributions from both an energy systems perspective and an industrial engineering perspective. Firstly, we have developed a method to apply statistical interpolation techniques to express the relationship between generation expansion planning decisions and human health externalities via the utilization of kriging metamodels. Secondly, we solve a large-scale, multi-period optimization model over an extended time horizon with this metamodel as a surrogate objective function. Finally, we develop a new iterative method to search the feasible region of solutions, resample, and refit the metamodel of human health externalities, to ensure the true relationship between decision variables and the metamodel is accurate, and thus, the optimization method yields robust candidate solutions.

References

- Abt Associates Inc. (2012, October 1). *Environmental Benefits Mapping and Analysis Program User's Manual*. Retrieved March 1, 2013, from Environmental Benefits Mapping and Analysis Program (BenMAP): <http://www.epa.gov/airquality/benmap/models/BenMAPManualOct2012.pdf>
- Alnatheer, O. (2006). Environmental benefits of energy efficiency and renewable energy in Saudi Arabia's electric sector. *Energy Policy*, 2-10.
- Alves, A. A., & Uturbey, W. (2010). Environmental degradation costs in electricity generation: The case of the Brazilian electrical matrix. *Energy Policy*, 6204-6214.
- Antunesa, C. H., Martinsa, A. G., & Britob, I. S. (2004). A multiple objective mixed integer linear programming model for power generation expansion planning. *Energy*, 29, 613-627.
- Arsham, H. (1995, February 11). *Systems Simulation: The Shortest Route to Applications*. Retrieved July 9, 2012, from <http://home.ubalt.edu/ntsbarsh/simulation/sim.htm#rrgoalopti>
- Banzhaf, H. S., Desvoisges, W. H., & Johnson, F. R. (1996). Assessing the externalities of electricity generation in the Midwest. *Resource and Energy Economics*, 18, 395-421.
- Bas, D., & Boyaci, I. H. (2007). Modeling and optimization I: Usability of response surface methodology. *Journal of Food Engineering*, 78(3), 836-845.
- Becerra-Lopez, H. R., & Golding, P. (2008). Multi-objective optimization for capacity expansion of regional power-generation systems: Case study of far west Texas. *Energy Conversion and Management*, 49, 1433-1445.
- Becker, N., Soloveitchik, D., & Olshansky, M. (2011). Incorporating environmental externalities into the capacity expansion planning:. *Energy Conversion and Management*, 2489-2494.
- Begovic, M., Novosel, D., & Milisavljevic, M. (2001). Trends in power system protection and control. *Decision Support Systems*, 30, 269-278.
- Bezerra, M. A., Santelli, R. E., Olivera, E. P., Villar, L. S., & Escalera, L. A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, 76(5), 965-977.
- Bloom, J. A., Caramanis, M., & Charny, L. (1984). Long-Range Generation Planning Using Generalized Benders' Decomposition: Implementation and Experience. *Operations Research*, 32(2), 290-313.
- Blume, S. W. (2007). *Electric Power System Basics*. Piscataway, NJ: IEEE Press.

- Bohling, G. (2005, October 19). *Kriging*. Retrieved June 22, 2012, from <http://people.ku.edu/~gbohling/cpe940/Kriging.pdf>
- Botterud, A., Ilic, M. D., & Wangensteen, I. (2005). Optimal Investments in Power Generation Under Centralized and Decentralized Decision Making. *IEEE Transactions on Power Systems*, 20(1), 254-263.
- Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Petron, G., Jordaan, S. M., . . . Harriss, R. (2014). Methane Leaks from North American Natural Gas Systems. *Energy and Environment*, 733-735.
- Burtaw, D., Krupnick, A., Palmer, K., Paul, A., Toman, M., & Bloyd, C. (2003). Ancillary benefits of reduced air pollution in the US from moderate greenhouse gas mitigation policies in the electricity sector. *Journal of Environmental Economics and Management*, 45(3), 650-673.
- Cambridge Energy Solutions. (2011). *DAYZER Day-Ahead Market Analyzer*. Retrieved March 1, 2013, from Cambridge Energy Solutions: http://www.ces-us.com/download/DAYZER_v125_brochure.pdf
- Chen, Q., Kang, C., Xia, Q., & Zhong, J. (2010). Power Generation Expansion Planning Model Towards Low-Carbon Economy and Its Application in China. *IEEE Transactions on Power Systems*, 25(2), 1117-1125.
- Chen, S.-L., Zhan, T.-S., & Tsay, M.-T. (2006). Generation expansion planning of the utility with refined immune algorithm. *Electric Power Systems Research*, 76(4), 251-258.
- Chen, X., Du, W., & Liu, D. (2008). Response surface optimization of biocatalytic biodiesel production with acid oil. *Biochemical Engineering Journal*, 40(3), 423-429.
- Chen, Y.-H., & Hobbs, B. (2005). An Oligopolistic Electricity Market Model with Tradable NO_x Permits. *IEEE Transactions on Power Systems*, 20(1), 119-129.
- Chung, A. S., Wu, F., & Varaiya, P. (2001). A Game-Theoretic Model for Generation Expansion Planning: Problem Formulation and Numerical Comparisons. *IEEE Transactions on Power Systems*, 16(4), 885-891.
- ConvertUnits.com. (2016). *Conversion of Measurement Units*. Retrieved February 11, 2016, from <http://www.convertunits.com/from/hundred+cubic+foot+of+natural+gas/to/MWh>
- Cressie, N. (1990). The Origins of Kriging. *Mathematical Geology*, 22(3), 239-252.
- Day, C. J. (2002). Oligopolistic Competition in Power Networks: A Conjectured Supply Function Approach. *IEEE Transactions on Power Systems*, 17(3), 597-607.

- De Jonghe, C., & Hobbs, B. (2012). Optimal Generation Mix With Short-Term Demand Response and Wind Penetration. *IEEE Transactions on Power Systems*, 27(2), 830-839.
- Delucchi, M. A., Murphy, J. J., & McCubbin, D. R. (2002). The health and visibility cost of air pollution: a comparison of estimation methods. *Journal of Environmental Management*, 64(2), 139–152.
- Elsayed, E. A. (2012). *Reliability Engineering*. Hoboken: John Wiley & Sons, Inc.
- Farghal, S. A., & Abdel Aziz, M. R. (1988). Generation expansion planning including the renewable energy sources. *IEEE Transactions on Power Systems*, 3(3), 816-822.
- Farkas, C. M., Moeller, M. D., Felder, F. A., Baker, K. R., Rodgers, M. D., & Carlton, A. G. (2015). Temporalization of Peak Electric Generation Particulate Matter Emissions during High Energy Demand Days. *Environ. Sci. Technol.*, 49(7), 4696–4704.
- Felder, F. A. (2012). In Depth Introduction to Electricity Markets. Denver: EUCI.
- Firmo, H. T., & Legey, L. F. (2002). Generation Expansion Planning: An Iterative Genetic Algorithm Approach. *IEEE Transactions on Power Systems*, 17(3), 901-906.
- Fukuyama, Y., & Chiang, H.-D. (1996). A parallel genetic algorithm for generation expansion planning. *IEEE Transactions on Power Systems*, 11(2), 955-961.
- GAMS Development Corporation. (2014). *MSNLP and OQNLP*. Retrieved February 12, 2016, from GAMS Documentation Center: <http://www.gams.com/help/index.jsp?topic=%2Fgams.doc%2Fsolvers%2Findex.html>
- Gano, S. E. (2005). *Simulation-Based Design Using Variable Fidelity Optimization*. Notre Dame: University of Notre Dame.
- General Electric International. (2014). *PJM Renewable Integration Study*. Schenectady: PJM Interconnection, LLC.
- Gilmore, E. A., Apt, J., Walawalkar, R., Adams, P. J., & Lave, L. B. (2010). The air quality and human health effects of integrating utility-scale batteries into the New York State electricity grid. *Journal of Power Sources*, 195(8), 2405-2413.
- Gnansounou, E., Jun, D., Pierre, S., & Quintero, A. (2004). Market Oriented Planning of Power Generation Expansion using Agent-based Model. *Power Systems Conference and Exposition, 2004. IEEE PES*. New York.
- Goffman, J. (2010). *Reducing Pollution from Power Plants*. U.S. EPA Office of Air and Radiation. U.S. EPA.

- Gorenstin, B. G., Campodonico, N. M., Costa, J. P., & Pereira, M. V. (1993). Power system expansion planning under uncertainty. *IEEE Transactions on Power Systems*, 8(1), 129-136.
- Gosavi, A. (2003). *Simulation Based Optimization: Parametric Optimization Techniques and Reinforcement Learning* (1st ed.). Norwell: Kluwer Academic Publishers.
- Haines, A., Kovats, R. S., Campbell-Lendrum, D., & Corvalan, C. (2006). Climate change and human health: impacts, vulnerability, and mitigation. *Lancet*, 367, 2101-2109.
- Haines, A., McMichael, A. J., Smith, K. R., Roberts, I., Woodcock, J., Markandya, A., . . . Wilkinson, P. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers. *Lancet*, 374, 2104-2114.
- Han, Z.-H., & Zhang, K.-S. (2012). Surrogate-Based Optimization. *Real-World Applications of Genetic Algorithms*. (D. O. Roeva, Ed.) InTech.
- Hausfather, Z., & Muller, R. (2014). Retrieved February 11, 2016, from <http://static.berkeleyearth.org/memos/epa-report-reveals-lower-methane-leakage-from-natural-gas.pdf>
- He, K., Lei, Y., Pan, X., Zhang, Y., & Chen, D. (2010). Co-benefits from energy policies in China. *Energy*, 35(11), 4265-4272.
- He, Y. Q., & David, A. K. (1997). Time-of-use Electricity Pricing Based on Global Optimization for Generation Expansion Planning. *4th International Conference on Advances in Power System Control, Operation and Management, APSCOM-97*. New York.
- Hemmati, R., Hooshmand, R.-A., & Khodabakhshian, A. (2013). Comprehensive review of generation and transmission expansion planning. *IET Generation, Transmission & Distribution*, 955-964.
- Henshaw, D. L. (2002). Does our electricity distribution system pose a serious risk to public health? *Medical Hypotheses*, 59(1), 39-51.
- Hobbs, B. (2001). Linear Complementarity Models of Nash–Cournot Competition in Bilateral POOLCO Power Markets. *IEEE Transactions on Power Systems*, 16(2), 194-202.
- Huang, D., Allen, T. T., Notz, W. I., & Miller, R. A. (2006). Sequential kriging optimization using multiple-fidelity evaluations. *Struct Multidisc Optim*, 32(5), 369-382.

- Huang, D., Allen, T. T., Notz, W. I., & Zeng, N. (2006). Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. *Journal of Global Optimization*, 34(3), 441-466.
- Humphrey, D. G., & Wilson, J. R. (2000). A Revised Simplex Search Procedure for Stochastic Simulation Response Surface Optimization. *INFORMS Journal on Computing*, 12(4), 272-283.
- Idaho Governor's Office of Energy Resources. (2013). *Baseload Power*. Retrieved March 1, 2013, from Idaho Governor's Office of Energy Resources: <http://www.energy.idaho.gov/baseload.htm>
- Inhaber, H. (1979). Risk with Energy from Conventional and Nonconventional Sources. *Science*, 203(23), 718-723.
- Institute for the Environment, University of North Carolina at Chapel Hill. (2010, February 1). *Operational Guidance for the Community Multiscale Air Quality (CMAQ) Modeling System*. Retrieved March 1, 2013, from Community Multiscale Air Quality: [http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_\(February_2010_release\)_OGD](http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_(February_2010_release)_OGD)
- Intergovernmental Panel on Climate Change. (2016, April). *Climate Change 2007: Working Group III: Mitigation of Climate Change*. Retrieved from http://www.ipcc.ch/publications_and_data/ar4/wg3/en/ch3s3-5-3-3.html
- Jakumeit, J., Herdy, M., & Nitsche, M. (2005). Parameter optimization of the sheet metal forming process using an iterative parallel Kriging algorithm. *Struct Multidisc Optim*, 29(6), 498-507.
- Jeong, S., Murayama, M., & Yamamoto, K. (2005). Efficient Optimization Design Method Using Kriging Model. *Journal of Aircraft*, 42(2), 413-420.
- Jones, D. R. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21, 345-383.
- Jung, J. Y., Blau, G., Pekny, J. F., Reklatis, G. V., & Eversdyk, D. (2004). A simulation based optimization approach to supply chain management under demand uncertainty. *Computers and Chemical Engineering*, 28, 2087-2106.
- Kagiannis, A. G., Askounis, D. T., & Psarras, J. (2004). Power generation planning: a survey from monopoly to competition. *Electrical Power & Energy Systems*, 26(6), 413-421.
- Kalil, S. J., Maugeri, F., & Rodrigues, M. I. (2000). Response surface analysis and simulation as a tool for bioprocess design and optimization. *Process Biochemistry*, 35(6), 539-550.

- Kannan, S., Baskar, S., McCalley, J. D., & Murugan, P. (2009). Application of NSGA-II Algorithm to Generation Expansion Planning. *IEEE Transactions on Power Systems*, 24(1), 454-461.
- Kannan, S., Slochanal, S. M., & Padhy, N. P. (2005). Application and comparison of metaheuristic techniques to generation expansion planning problem. *IEEE Transactions of Power Systems*, 20(1), 466-475.
- Kannan, S., Slochanal, S. M., Baskar, S., & Murugan, P. (2007). Application and comparison of metaheuristic techniques to generation expansion planning in the partially deregulated environment. *IET Generation, Transmission, & Distribution*, 1(1), 111-118.
- Kannan, S., Slochanal, S. M., Subbaraj, P., & Padhy, N. P. (2004). Application of particle swarm optimization technique and its variants to generation expansion planning problem. *Electric Power Systems Research*, 70, 203-210.
- Karaki, S. H., Chaaban, F. B., Al-Nakhl, N., & Tarhini, K. A. (2002). Power generation expansion planning with environmental consideration for Lebanon. *Electrical Power & Energy Systems*, 24(8), 611-619.
- Kaymaz, P., Valenzuela, J., & Park, C. S. (2007). Transmission Congestion and Competition on Power Generation Expansion. *IEEE Transactions on Power Systems*, 22(1), 156-163.
- Kenfack, F., Guinet, A., & Ngundam, J. M. (2001). Investment planning for electricity generation expansion in a hydro dominated environment. *International Journal of Energy Research*, 25(10), 927-937.
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks*. Perth, Western Australia.
- Kerl, P. Y., Zhang, W., Moreno-Cruz, J. B., Nenes, A., Realff, M. J., Russell, A. G., . . . Thomas, V. M. (2015). New approach for optimal electricity planning and dispatching with hourly time-scale air quality and health considerations. *Proceedings of the National Academy of Sciences of the United States of America* (pp. 10884-10889). Washington D.C.: National Academy of Sciences of the United States of America.
- Kim, J.-H., Park, J.-B., & Park, J.-K. (2005). A market-based analysis on the generation expansion planning strategies. *Intelligent Systems Application to Power Systems, 2005*. Washington D.C.
- Klaassen, G., & Riahi, K. (2007). Internalizing externalities of electricity generation: An analysis with MESSAGE-MACRO. *Energy Policy*, 35(2), 815-827.

- Kleijnen, J. P. (2005). An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research*, 164(2), 287-300.
- Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192(3), 707-716.
- Koop, G., & Tole, L. (2004). Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *Journal of Environmental Economics and Management*, 47, 30-54.
- Koop, G., & Tole, L. (2004). Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *Journal of Environmental Economics and Management*, 47(1), 30-54.
- Kougea, E., & Koundouri, P. (n.d.). Air Quality Degradation: Can Economics Help in Measuring its Welfare Effects? A Review of Economic Valuation Studies. *DEOS Working Papers from Athens University of Economics and Business*, 1129.
- Krewitt, W., Hurley, F., Trukenmuller, A., & Friedrich, R. (1998). Health Risks of Energy Systems. *Risk Analysis*, 18(4), 377-383.
- Kumar, E. S., & Sarkar, B. (2012). Proportional Hazards Modeling of Environmental Impacts on Reliability of Photovoltaic Modules . *International Journal of Engineering and Advanced Technology (IJEAT)* , 2(2), 110-115.
- Lang, L., & Xu, K. (2002). A Unified Response for Dual Response Surface Optimization. *Journal of Quality Technology*, 34(4), 437-447.
- Lee, E. T., & Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis*. Hoboken: John Wiley & Sons, Inc.
- Lee, K.-H., & Park, G.-J. (2006). A Global Robust Optimization Using Kriging Based Approximation Model. *JSME International Journal*, 49(3), 779-788.
- Leou, R.-C. (2011). A multi-year transmission planning under a deregulated market. *Electrical Power & Energy Systems*, 33(3), 708-714.
- Li, M., Li, G., & Azarm, S. (2008). A Kriging Metamodel Assisted Multi-Objective Genetic Algorithm for Design Optimization. *Journal of Mechanical Design*, 130(3).
- Lin, D. K., & Tu, W. (1997). Dual Response Surface Optimization. *Journal of Quality Technology*, 27(1), 34-39.
- Linares, P., Santos, F. J., Ventosa, M., & Lapierda, L. (2008). Incorporating oligopoly, CO2 emissions trading and green certificates into a power generation expansion model. *Journal of IFAC*, 44(6), 1608-1620.

- Liu, Z. F., Huang, G. H., & Li, N. (2008). A Dynamic Optimization Approach for Power Generation Planning under Uncertainty. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 30(14-15), 1413-1431.
- Longo, A., Markandya, A., & Petrucci, M. (2008). The internalization of externalities in the production of electricity: Willingness to pay for the attributes of a policy for renewable energy. *Ecological Economics*, 67(1), 140-152.
- Lopez, J. A., Ponnambalam, K., & Quintana, V. H. (2007). Generation and Transmission Expansion Under Risk Using Stochastic Programming. *IEEE Transactions on Power Systems*, 22(3), 1369-1378.
- Machol, B., & Rizk, S. (2013). Economic value of U.S. fossil fuel electricity health impacts. *Environmental International*, 52, 75-80.
- Maghouli, P., Hosseini, S. H., Buygi, M. O., & Shahidehpour, M. (2011). A Scenario-Based Multi-Objective Model for Multi-Stage Transmission Expansion Planning. *IEEE Transactions on Power Systems*, 26(1), 470-478.
- Malcolm, S. A., & Zenios, S. A. (1994). Robust optimization for power systems capacity expansion under uncertainty. *The Journal of the Operational Research Society*, 45(9), 1040-1049.
- Markandya, M., & Wilkinson, P. (2007). Electricity generation and health. *Lancet*, 370, 979-990.
- Marneris, I. G., Biskas, P. N., & Bakirtzis, E. A. (2016). An Integrated Scheduling Approach to Underpin Flexibility in European Power Systems. *IEEE Transactions on Sustainable Energy*, 647-657.
- Massachusetts Institute of Technology. (2011). The Future of the Electric Grid: An Interdisciplinary MIT Study. Massachusetts Institute of Technology.
- Mazadi, M., Rosehart, W. D., Malik, O. P., & Aguado, J. A. (2009). Modified Chance-Constrained Optimization Applied to the Generation Expansion Problem. *IEEE Transactions on Power Systems*, 24(3), 1635-1636.
- Mazumdar, M., & Kapoor, A. (1995). Stochastic models for power generation system production costs. *Electric Power Systems Research*, 35, 93-100.
- Meza, J. L., Yildirim, M. B., & Masud, A. S. (2007). A Model for the Multiperiod Multiobjective Power Generation Expansion Problem. *IEEE Transactions on Power Systems*, 22(2), 871-878.
- Meza, J. L., Yildirim, M. B., & Masud, A. S. (2009). A Multiobjective Evolutionary Programming Algorithm and Its Applications to Power Generation Expansion Planning. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 39(5), 1086-1096.

- Min, K. J., & Subramanian, P. S. (2002). A generation expansion model for electric utilities with stochastic stranded cost. *Electrical Power and Energy Systems*, 24, 875-885.
- Mo, B., Hegge, J., & Wangenstein, I. (1991). Stochastic generation expansion planning by means of stochastic dynamic programming. *IEEE Transactions on Power Systems*, 6(2), 662-668.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to Linear Regression Analysis* (4th ed.). Hoboken: John Wiley & Sons, Inc.
- Moura, J. (2016, April 29). Reliability and the Future of the Electricity Grid: A North American Bulk Power System Perspective. North American Electric Reliability Corporation.
- Murphy, F. H., & Smeers, Y. (2005). Generation Capacity Expansion in Imperfectly Competitive Restructured Electricity Markets. *Operations Research*, 53(4), 646-661.
- Nanduri, V., Das, T. K., & Rocha, P. (2009). Generation Capacity Expansion in Energy Markets using a Two-Level Game-Theoretic Model. *IEEE Transactions on Power Systems*, 24(3), 1165-1172.
- National Research Council. (2010). *Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use*. Washington, DC: National Academies Press.
- National Research Council of the National Academies. (2010). *Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use*. Washington, DC: The National Academies Press.
- Nguyen, K. Q. (2008). Internalizing externalities into capacity expansion planning: The case of electricity in Vietnam. *Energy*, 33(5), 740-746.
- Nordlund, P., Sjelvgren, D., Pereira, M. V., & Bubenko, J. A. (1987). Generation expansion planning for systems with a high share of hydro power. *IEEE Transactions on Power Systems*, 2(1), 161-167.
- North American Electric Reliability Corporation. (2016). *About NERC*. Retrieved from NERC: North American Electric Reliability Corporation: <http://www.nerc.com/AboutNERC/Pages/default.aspx>
- Northwest Power and Conservation Council. (2011). *A Probabilistic Method to Assess Power Supply Adequacy for the Pacific Northwest*. Northwest Power and Conservation Council.
- Nualhong, D., Chusanapiputt, S., Jantarang, S., & Pungprasert, V. (2005). Generation Expansion Planning Including Biomass Energy Sources with Global

Environmental Consideration Using Improved Tabu Search. *TENCON 2005 IEEE Region 10*. Melbourne, Australia.

- Nuclear Energy Institute. (2012, June). *nei.org*. Retrieved 2013, from U.S. Capacity Factors by Fuel Type:
<http://www.nei.org/resourcesandstats/documentlibrary/reliableandaffordableenergy/graphicsandcharts/uscapacityfactorsbyfueltype/>
- Park, J.-B., Kim, J.-H., & Lee, K. Y. (2002). Generation Expansion Planning in a Competitive Environment Using a Genetic Algorithm. Chicago: IEEE Power Engineering Society.
- Park, J.-B., Park, Y.-M., Won, J.-R., & Lee, K. Y. (2000). An Improved Genetic Algorithm for Generation Expansion Planning. *IEEE Transactions on Power Systems*, 15(3), 916-922.
- Park, Y. M., Park, J. B., & Won, J. R. (1998). A hybrid genetic algorithm/dynamic programming approach to optimal long-term generation expansion planning. *Electrical Power & Energy Systems*, 20(4), 295-303.
- Pereira, A. J., & Saraiva, J. T. (2010). A decision support system for generation expansion planning in competitive electricity markets. *Electric Power Systems Research*, 80(7), 778-787.
- Pervin, T., Gerdtham, U.-G., & Lyttkens, C. H. (2008). Societal costs of air pollution-related health hazards: A review of methods and results. *Cost Effectiveness and Resource Allocation*, 6(19), 1-22.
- Pokharel, S., & Ponnambalam, K. (1997). Investment planning for electricity generation expansion. *International Journal of Energy Research*, 21(2), 185-194.
- Poulin, A., Dostie, A., Fournier, M., & Sansregret, S. (2008). Load duration curve: A tool for technico-economic analysis of energy solutions. *Energy and Buildings*, 40, 29-35.
- Quanhong, L., & Caili, F. (2005). Application of response surface methodology for extraction optimization of germinant pumpkin seeds protein. *Food Chemistry*, 92(4), 701-706.
- Rashad, S. M., & Hammad, F. H. (2000). Nuclear power and the environment: comparative assessment of environmental and health impacts of electricity-generating systems. *Applied Energy*, 65, 211-229.
- Rasmussen, C. E. (2004). Gaussian Processes in Machine Learning. *Lecture Notes in Computer Science*, 3176.

- Ratle, A. (2001). Kriging as a surrogate fitness landscape in evolutionary optimization. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 15(1), 37-49.
- Roux, W. J., Stander, N., & Haftka, R. T. (1998). Response Surface Approximations for Structural Optimization. *International Journal for Numerical Methods in Engineering*, 42(3), 517-534.
- Rowe, R. D., Lang, C. M., & Chestnut, L. G. (1996). Critical factors in computing externalities for electricity resources. *Resource and Energy Economics*, 18(4), 363-394.
- Sadegheih, A., & Drake, P. R. (2008). System network planning expansion using mathematical programming, genetic algorithms and tabu search. *Energy Conversion and Management*, 49, 1557-1566.
- Sakata, S., Ashida, F., & Zako, M. (2003). Structural optimization using Kriging approximation. *Computer Methods in Applied Mechanics and Engineering*, 192(7-8), 923-939.
- Sakata, S., Ashida, F., & Zako, M. (2004). An efficient algorithm for Kriging approximation and optimization with large-scale sampling data. *Computer Methods in Applied Mechanics and Engineering*, 193(3-5), 385-404.
- Sallaberrya, C., Helton, J., & Hora, S. (2008). Extension of Latin hypercube samples with correlated variables. *Reliability Engineering and System Safety*, 93(7), 1047-1059.
- Schenk, K. F., & Chan, S. (1981). Incorporation and impact of a wind energy conversion system in generation expansion planning. *IEEE Transactions on Power Apparatus and Systems*, 100(12), 4710-4718.
- Scientific American. (2016). *Scientific American*. Retrieved from How is electricity from different generators synchronized so that it can be combined to service the same grid?: <http://www.scientificamerican.com/article/how-is-electricity-from-d/>
- Selcuklu, S. B., Coit, D. W., Felder, F., & Rodgers, M. (2013). A new methodology for solving multi-objective stochastic optimization problems with independent objective functions. *2013 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 101-105). Bangkok: IEEE.
- Shayanfar, H. A., Lahiji, A. S., Aghaei, J., & Rabiee, A. (2009). Generation Expansion Planning in pool market- A hybrid modified game theory and improved genetic algorithm. *Energy Conversion and Management*, 50(5), 1149-1156.
- Simpson, T. W., Mauery, T. M., Korte, J. J., & Mistree, F. (2001). Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *AIAA Journal*, 39(12), 2233-2241.

- Sinha, N., Chakrabarti, R., & Chattopadhyay, P. K. (2003). Evolutionary Programming Techniques for Economic Load Dispatch. *IEEE Transactions on Evolutionary Computation*, 7(1), 83-94.
- Sirikum, J., & Techanitisawad, A. (2007). Power generation expansion planning with emission control: a nonlinear model and a GA-based heuristic approach. *International Journal of Energy Research*, 30(2), 81-99.
- Sirikum, J., Techanitisawad, A., & Kachitvichyanukul, V. (2007). A New Efficient GA-Benders' Decomposition Method: For Power Generation Expansion Planning With Emission Controls. *IEEE Transactions on Power Systems*, 22(3).
- Slochanal, S. M., Kannan, S., & Rengaraj, R. (2004). Generation Expansion Planning in the Competitive Environment. *2004 International Conference on Power System Technology - POWERCON 2004*. Singapore.
- Su, C.-T., Lii, G.-R., & Chen, J.-J. (2000). Long-term generation expansion planning employing dynamic programming and fuzzy techniques. *Proceedings of IEEE International Conference on Industrial Technology*. Goa, India.
- Tekiner, H. (2010). *Multi-Objective Stochastic Models for Electricity Generation Expansion Planning Problems Considering Risk*. Ann Arbor: ProQuest LLC.
- Tekiner, H., Coit, D. W., & Felder, F. A. (2009). Effects of Smart Grid Technologies on Generation Expansion Plans. *3rd Annual Trans-Atlantic Infraday Conference on Applied Infrastructure Modeling and Policy Analysis*. Washington, DC.
- Tekiner, H., Coit, D. W., & Felder, F. A. (2012). Multi-period multi-objective electricity generation expansion planning problem with Monte-Carlo simulation. *Electric Power Systems Research*, 80(12), 1394-1405.
- Thanh, B. D., & Lefevre, T. (2000). Assessing health impacts of air pollution from electricity generation: the case of Thailand. *Environmental Impact Assessment Review*, 20(2), 137-158.
- The Institute for the Environment - The University of North Carolina at Chapel Hill. (2012, August 3). *SMOKE v3.1 User's Manual*. Retrieved March 1, 2013, from SMOKE-Model.org: <http://www.smoke-model.org/version3.1/html/>
- The White House. (2015, November 23). *whitehouse.gov*. Retrieved from Obama Administration Announces 2016 Greenhouse Gas Targets and Sustainability Plans; Highlights Federal Leadership on Climate Action: <https://www.whitehouse.gov/the-press-office/2015/11/23/obama-administration-announces-2016-greenhouse-gas-targets-and>
- Truong, T. H., & Azadivar, F. (2003). Simulation Based Optimization for Supply Chain Configuration Design. *Proceedings of the 2003 Winter Simulation Conference*. New Orleans.

- U.S. Department of Energy. (2008). *20% Wind Energy by 2030: Increasing Wind Energy's Contribution to U.S. Electricity Supply*. Oak Ridge: U.S. Department of Energy.
- U.S. Energy Information Administration. (2012, June). *www.eia.gov*. Retrieved 2013, from Assumptions to the Annual Energy Outlook 2012: [http://www.eia.gov/forecasts/aeo/assumptions/pdf/0554\(2012\).pdf](http://www.eia.gov/forecasts/aeo/assumptions/pdf/0554(2012).pdf)
- U.S. Energy Information Administration. (2015). *Electricity Data Browser*. Retrieved January 6, 2016, from <http://www.eia.gov/electricity/data/browser/>
- U.S. Environmental Protection Agency. (2013, July). *www.epa.gov*. Retrieved 2013, from Clean Energy: <http://www.epa.gov/cleanenergy/energy-and-you/affect/air-emissions.html>
- U.S. Environmental Protection Agency. (2015). *Regulatory Impact Analysis for the Clean Power Plan Final Rule*. Washington, D.C.: U.S. Environmental Protection Agency.
- U.S. Environmental Protection Agency. (2016). *FACT SHEET: Clean Power Plan Benefits*. Retrieved from U.S. Environmental Protection Agency: <https://www.epa.gov/cleanpowerplan/fact-sheet-clean-power-plan-benefits#benefits>
- U.S. Environmental Protection Agency. (2016, April). *Sources of Greenhouse Gas Emissions*. Retrieved from United States Environmental Protection Agency: <https://www3.epa.gov/climatechange/ghgemissions/sources.html>
- U.S. Environmental Protection Agency. (2016). *The Social Cost of Carbon*. Retrieved January 17, 2016, from <https://www3.epa.gov/climatechange/EPAactivities/economics/scc.html>
- U.S.-Canada Power System Outage Task Force. (2004, April). *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. Retrieved from Energy.gov: <http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf>
- Union of Concerned Scientists. (2013). *The Hidden Cost of Fossil Fuels*. Retrieved from Union of Concerned Scientists: http://www.ucsusa.org/clean_energy/our-energy-choices/coal-and-other-fossil-fuels/the-hidden-cost-of-fossil.html#.VzkEo5GDGko
- United States Department of Energy. (2012, April 17). *Energy.gov*. Retrieved from Powering Up America's Waterways: <http://energy.gov/articles/powering-americas-waterways>

- United States Environmental Protection Agency. (2015). *User's Manual for the Co-Benefits Risk Assessment (COBRA) Screening Model - Version 2.7*. Washington, DC: United States Environmental Protection Agency.
- United States Environmental Protection Agency. (2016, April). *Future Climate Change*. Retrieved from United States Environmental Protection Agency: <https://www3.epa.gov/climatechange/science/future.html#Temperature>
- Unsihuay-Vila, C., Marangon-Lima, J. W., Zambroni de Souza, A. C., & Perez-Arriaga, I. J. (2011). Multistage expansion planning of generation and interconnections with sustainable energy development criteria: A multiobjective model. *Electrical Power and Energy Systems*, 33, 258-270.
- UW Electrical Engineering. (2010). *IEEE Power Systems Test Case Archive*. Retrieved 2013, from UW Electrical Engineering: U.S. Energy Information Administration
- Valenzuela, J., & Mazumdar, M. (2005). A Probability Model for the Electricity Price Duration Curve Under an Oligopoly Market. *IEEE Transactions on Power Systems*, 20(3), 1250-1256.
- Valenzuela, J., & Mazumdar, M. (2005). A Probability Model for the Electricity Price Duration Curve Under an Oligopoly Market. *IEEE Transactions on Power Systems*, 20(3), 1250-1256.
- von Winterfeldt, D., Eppel, T., Adams, J., Neutra, R., & DelPizzo, V. (2004). Managing Potential Health Risks from Electric Powerlines: A Decision Analysis Caught in Controversy. *Risk Analysis*, 24(6), 1487-1502.
- Voorhees, A. S., Fann, N., Fulcher, C., Dolwick, P., Hubbell, B., Bierwagen, B., & Morefield, P. (2011). Climate Change-Related Temperature Impacts on Warm Season Heat Mortality: A Proof-of-Concept Methodology Using BenMAP. *Environmental Science and Technology*, 45(4), 1450-1457.
- Voropai, N. I., & Ivanova, E. Y. (2003). Hierarchical Game Theoretical Problem of Electric Power System Expansion Planning. *Proceedings, 2003 IEEE Bologna PowerTech Conference*. Bologna.
- Voropai, N. I., & Ivanova, E. Y. (2006). Shapley Game for Expansion Planning of Generating Companies at Many Non-Coincident Criteria. *IEEE Transactions on Power Systems*, 21(4), 1630-1637.
- Winston, W. L., & Venkataramanan, M. (2003). *Operations Research: Volume One - Introduction to Mathematical Programming*. Pacific Grove: Brooks/Cole - Thomson Learning.
- Wong, E. (2013, July 8). Pollution Leads to Drop in Life Span in Northern China, Research Finds. *The New York Times*, p. A6.

- Wood, A., & Wollenberg, B. (1996). *Power Generation, Operation, and Control* (Second Edition ed.). New York: John Wiley & Sons, Inc.
- Xija, N., Yokoyama, R., Zhou, Y. C., & Kozu, A. (2000). An effective DP solution for optimal generation expansion planning under new environment. Perth, Australia: IEEE Power System Technology.
- Xija, N., Yokoyama, R., Zhou, Y. C., & Kozu, A. (2000). An effective DP solution for optimal generation expansion planning under new environment. *Proceedings PowerCon 2000*. Perth, Australia.
- Yildirim, M., Erkan, K., & Ozturk, S. (2006). Power generation expansion planning with adaptive simulated annealing genetic algorithm. *International Journal of Energy Research*, 30(14), 1188-1199.
- Youn, B. D., & Choi, K. K. (2004). A new response surface methodology for reliability-based design optimization. *Computers & Structures*, 82(2-3), 241-256.
- Yu, Y., Gan, D., Wu, H., & Han, Z. (2010). Frequency induced risk assessment for a power system accounting uncertainties in operation of protective equipments. *Electrical Power and Energy Systems*, 32(6), 688-696.
- Zhu, J., & Chow, M.-y. (1997). A Review of Emerging Techniques on Generation Expansion Planning. *IEEE Transactions on Power Systems*, 12(4), 1722-1728.

Appendix A: Additional Model Assumptions

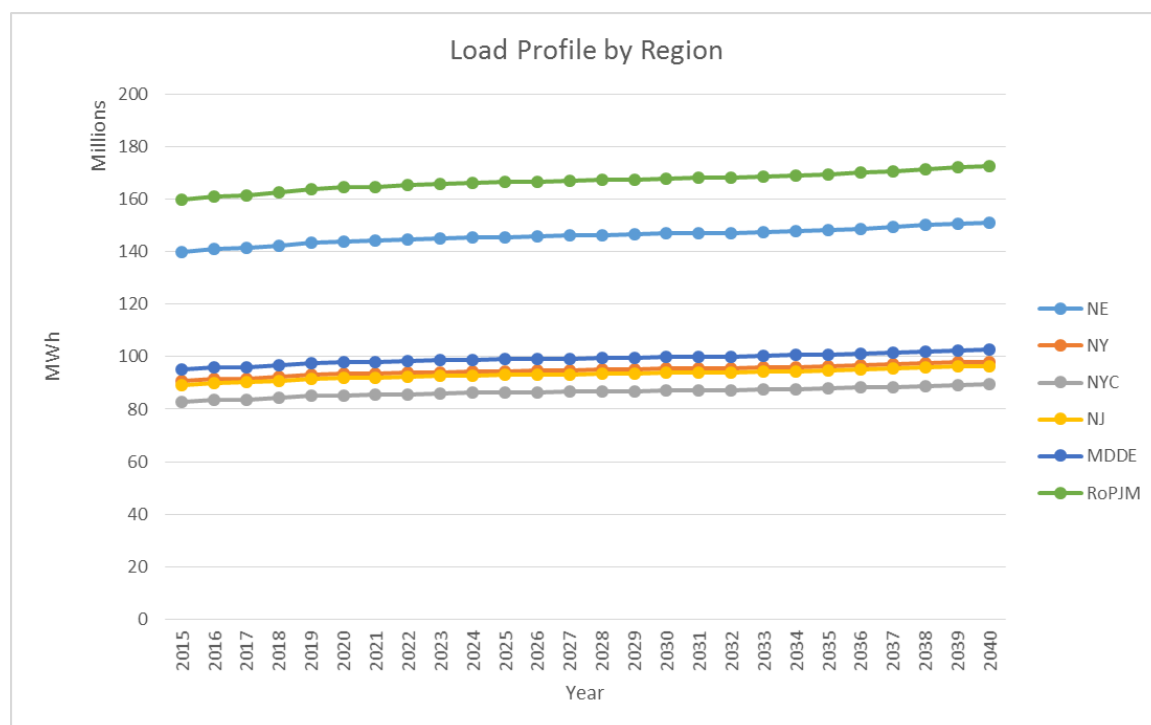


Figure A.1 – Load Profile by Region (U.S. Energy Information Administration, 2015)

Table A.1 – Initial Capacity by Region (in MW) (Cambridge Energy Solutions, 2011)

	Nuclear	Combined Cycle Gas Turbines	Coal (Low)	Coal (High)	Gas Turbines	Wind (On Land)	Petroleum	Hydro	Solar
NE	2,873	203	2,414	12,447	2,110	0	3,656	10,342	303
NY	5,282	5,313	4,463	2,763	305	1,274	99	5,705	-
NYC	-	2,216	7,765	-	5,426	-	107	-	-
NJ	4,119	6,055	1,241	2,063	4,816	8	47	406	13
MDDE	5,943	-	3,419	1,340	-	-	590	5,348	128
RoPJM	9,583	8,041	4,911	17,143	3,515	704	204	2,281	2
Totals	27,799	21,828	24,212	35,756	16,172	1,985	4,703	24,082	446

Table A.2 – Transmission Limits by Region (in MWh) (Cambridge Energy Solutions, 2011)

		Receiving Region					
		NE	NY	NYC	NJ	MDDE	RoPJM
Transmitting Region	NE	-	1,420	-	-	-	-
	NY	-	-	-	950	-	-
	NYC	430	-	-	1,685	-	-
	NJ	-	-	-	-	-	-
	MDDE	-	-	-	-	-	5,150
	RoPJM	-	2,000		9,268	-	-

Table A.3 – Unit Derating and Capacity Value Percentages (Cambridge Energy Solutions, 2011; U.S. Energy Information Administration, 2015)

	Derating Factor (%)	Capacity Value (%)
Nuclear	91%	100%
Combined Cycle Gas Turbines	84%	100%
Gas Turbines	86%	100%
Wind (On Land)	97%	32%
Gas Turbines	97%	100%
Combined Heat & Power	86%	100%
Biomass	89%	100%
Coal (High)	86%	100%
Coal (Low)	86%	100%
Solar	95%	14%
Petroleum	86%	100%
Hydro	95%	100%

**Table A.4 – Unit Capacity Factors by Region (Cambridge Energy Solutions, 2011;
U.S. Energy Information Administration, 2015)**

	Wind (On Land)	Wind (Off Shore)	Solar
NE	31%	59%	13%
NY	31%	57%	12%
NYC	31%	57%	14%
NJ	21%	39%	14%
MDDE	25%	39%	14%
RoPJM	27%	43%	13%

Table A.5 – Reserve Margin Percentages

	Reserve Margin Capacity (% of Peak Demand)
NE	116%
NY	117%
NYC	117%
NJ	115%
MDDE	115%
RoPJM	115%

Table A.6 – RGGI CO₂ Emissions Limits

	Total Annual CO₂ Emissions Limits in RGGI Regions (in lbs, Billions)
2015	366.6
2016	357.4
2017	348.5
2018 through 2040 annually	339.8

**Table A.7 – Regional Emissions Limits (U.S. Environmental Protection Agency,
2013)**

	NO_x Emissions Limits (in lbs, Millions)	SO₂ Emissions Limits (in lbs, Millions)
NE	120.00	0.54
NY	120.00	0.26
NYC	120.00	0.26
NJ	120.00	0.13
MDDE	120.00	0.36
RoPJM	120.00	1.07

Table A.8 – Annual Emissions Limits (U.S. Environmental Protection Agency, 2013)

	Total Annual SO₂ Emissions Limits (in lbs, Billions)
2015 through 2040 annually	17.9

Table A.9 – Transmission Losses

	Transmission Losses by Period (%)
Summer-Peak	91%
Summer-Off-Peak	91%
Winter-Peak	95%
Winter-off-Peak	95%
Spring/Fall-Peak	93%
Spring/Fall-Off-Peak	93%

Table A.10 – Renewables Trading Network (General Electric International, 2014)

		Receiving Region
Transmitting Region	NE	NE
	NY	NY
		NJ
		MDDE
		RoPJM
	NYC	NYC
		NE
	NJ	NJ
		RoPJM
	MDDE	MDDE
		NJ
		NY
		RoPJM
	RoPJM	RoPJM

Table A.11 – Available Renewables by Region (*General Electric International, 2014*)

	Renewables Available by Region
MDDE	Solar
	Biomass
	Wind (On Land)
	Wind (Off Shore)
NE	Biomass
	Wind (On Land)
	Wind (Off Shore)
	Solar
NJ	Biomass
	Wind (On Land)
	Wind (Off Shore)
	Solar
NY	Wind (On Land)
	Wind (Off Shore)
	Biomass
	Solar
NYC	Biomass
	Wind (Off Shore)
	Solar
RoPJM	Biomass
	Wind (On Land)
	Solar

Table A.12 – Minimum Percentage of Total Annual Dispatch from Renewables by Region (General Electric International, 2014)

	NE	NY	NYC	NJ	MDDE	RoPJM
2015	11%	6%	6%	12%	12%	11%
2016	12%	6%	6%	13%	14%	14%
2017	13%	6%	6%	14%	15%	14%
2018	14%	6%	6%	16%	17%	15%
2019	15%	6%	6%	18%	18%	15%
2020	16%	6%	6%	20%	19%	16%
2021	16%	6%	6%	23%	19%	18%
2022	16%	6%	6%	23%	20%	18%
2023	16%	6%	6%	23%	20%	18%
2024	17%	6%	6%	23%	20%	18%
2025	17%	6%	6%	23%	20%	18%
2026	17%	6%	6%	23%	20%	18%
2027	17%	6%	6%	23%	20%	18%
2028	17%	6%	6%	23%	20%	18%
2029	17%	6%	6%	23%	20%	18%
2030	17%	6%	6%	23%	20%	18%
2031	17%	6%	6%	23%	20%	18%
2032	17%	6%	6%	23%	20%	18%
2033	17%	6%	6%	23%	20%	18%
2034	17%	6%	6%	23%	20%	18%
2035	17%	6%	6%	23%	20%	18%
2036	17%	6%	6%	23%	20%	18%
2037	17%	6%	6%	23%	20%	18%
2038	17%	6%	6%	23%	20%	18%
2039	17%	6%	6%	23%	20%	18%
2040	17%	6%	6%	23%	20%	18%

Table A.13 – Minimum Percentage of Total Regional Dispatch from Renewable Energy Sources (General Electric International, 2014)

	Biomass	Wind (On Land)	Solar
MDDE	-	-	2%
NE	-	-	-
NJ	-	-	2%
NY	-	-	-
NYC	-	-	-
RoPJM	1%	8%	10%

Table A.14 – Annual Construction Limits by Unit (in MW per Region)

[illegible]

Table A.15 –Overall Construction Limits by Unit

	Maximum Construction Limit (MW)
Wind (On Land)	150,000
Combined Heat & Power	150,000
Nuclear	150,000
Wind (Off Shore)	150,000
Solar	150,000
Petroleum	150,000