Solar, Wind, and Storage: Optimizing for Least Cost Configurations of Renewable Energy Generation in the PJM Grid

Luke L. Cheng

Advisor: Professor Warren B. Powell

June 2014
Princeton, New Jersey

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Engineering

Department of Operations Research and Financial Engineering

Princeton University
I HEREBY DECLARE that I am the sole author of this thesis.

I AUTHORIZE Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Luke L. Cheng

I FURTHER AUTHORIZE Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Luke L. Cheng
The integration of solar and wind power into the grid poses many challenges due to the intermittent nature of weather conditions. This thesis models the hourly generation, storage, and consumption of solar, offshore wind, onshore wind, and fossil fuel energy such that demand is met every hour. For a given fossil fuel penalty, the least cost renewable energy build-out is determined through the use of a finite-difference stochastic approximation algorithm. The algorithm optimizes over five decision variables: solar power, offshore wind, onshore wind, battery inverter power, and battery storage capacity. The relationship between fossil fuel penalties and energy outcomes is explored for four different scenarios. This thesis finds that as fossil fuel energy costs rise, onshore wind and lithium-titanate grid-level storage become cost-effective for meeting demand.
ACKNOWLEDGEMENTS

This thesis is the culmination of not only many hours of my own work, but also the contributions of an entire community. Here, I would like to express my appreciation for a select few who made an significant impact on my time at Princeton.

Thank you to:

...Professor Warren Powell, for introducing me to the world of energy systems and for being an incredible advisor.

...Dr. Hugo Simao and Henry Chai for their time and patience in helping me obtain and process my data.

...Dr. Cory Budischak and Dr. Willet Kempton at the University of Delaware for sharing their data and code.

...Professor Eric Larson and Professor Dan Steingart for taking the time to help me with energy economics and battery technology costs.

...Professor Christopher Kuenne and Professor Ed Zschau for continuously inspiring me and guiding me.

...Malcolm Reid Jr., Victor Pomary, and Thomas Truongchau for being the best roommates a guy can ask for.

...Su Fen Goh and friends in Tower, for providing life-sustaining food and encouragement.

...the ORFE class of 2014, for their friendship and for working long nights alongside me.

...and my family, for their unconditional love.
# Table of Contents

Abstract ........................................................................................................... iii  
Acknowledgements ....................................................................................... iv 
Table of Contents ............................................................................................ v 
List of Figures ................................................................................................... vii 
List of Tables .................................................................................................... ix

## 1 Introduction ................................................................. 1 
1.1 PJM Interconnection ................................................................. 1 
1.2 Prior Studies .............................................................................. 4 
1.3 This Study .................................................................................... 7 

## 2 The Simulation Model .................................................. 8 
2.1 Model ......................................................................................... 12 
  2.1.1 State Variable ........................................................................ 13 
  2.1.2 Decision Variables ................................................................. 14 
  2.1.3 Exogenous Information ............................................................ 15 
  2.1.4 Transition Function ................................................................. 16 
  2.1.5 Objective Function ................................................................. 17 
2.2 Storage Policy Optimality ............................................................ 19 
2.3 Capacity Factor And Load Data ................................................... 23 
  2.3.1 Data From Budischak et al ....................................................... 25 
  2.3.2 Historical Power Output Data .................................................. 29 
2.4 Parameters And Costs ................................................................. 33 
  2.4.1 Calculating Levelized Cost Of Electricity .................................. 35 
  2.4.2 Storage Technology ................................................................. 37 
  2.4.3 Bid Stack Cost Function .......................................................... 40

## 3 Optimization ................................................................. 41 
3.1 Finite-Difference Stochastic Approximation .............................. 44 
3.2 Convergence .............................................................................. 46 
  3.2.1 Optimization As A Linear Programming Problem ..................... 48 
  3.2.2 Empirical Convexity Of The Loss Function ............................. 52 
3.3 Optimization Parameters ............................................................... 58 

## 4 Algorithmic Testing ...................................................... 59 
4.1 Stepsizes: alpha ....................................................................... 59 
4.2 Perturbation Size: delta ............................................................... 62 
4.3 Initial Estimates ......................................................................... 64 

## 5 Results ............................................................................. 67 
5.1 Simulation Model Output ........................................................... 68 
  5.1.1 Budischak Dataset ................................................................. 68 
  5.1.2 Historical Generation Dataset ................................................. 70 

V
5.2 FDSA Algorithm Performance ..............................................................73
5.3 Constant Fossil Fuel Energy Cost ......................................................77
  5.3.1 Budischak Dataset .........................................................................81
  5.3.2 Historical Generation Dataset ......................................................86
5.4 Bid Stack .........................................................................................91

6 DISCUSSION .......................................................................................93
  6.1 Opportunities For Further Research ...............................................95

7 CONCLUSION .....................................................................................97

8 REFERENCES ......................................................................................99

Colophon .................................................................................................103
LIST OF FIGURES

Figure 2-1: A typical PJM bid stack.........................................................11
Figure 2-2: PJM hourly load, Apr. 1998 to Dec. 2002 ..........................26
Figure 2-3: Locations of sites used for wind data..................................27
Figure 2-4: Hourly capacity factors for first week of Apr. 1998 .............28
Figure 2-5: Hourly capacity factors for first week of Sep. 2012 ............29
Figure 2-6: PJM onshore wind farm locations .....................................30
Figure 2-7: PSE&G solar panel sites .......................................................32
Figure 2-8: Bid stack used in simulation model .....................................41
Figure 3-1: Heat maps of battery inverter power and battery capacity ......53
Figure 3-2: Heat maps of solar and offshore wind .................................54
Figure 3-3: Offshore and onshore wind vs LCOE .................................55
Figure 3-4: Onshore wind and battery capacity vs LCOE .....................56
Figure 3-5: Offshore wind and battery capacity vs LCOE .....................57
Figure 3-6: Heat maps of solar and battery capacity (historical dataset) ...58
Figure 4-1: Effect of alpha value on FDSA results ..................................60
Figure 4-2: FDSA runs for alpha = 1, 5 .................................................60
Figure 4-3: FDSA behavior for alpha = 25 .............................................61
Figure 4-4: Effect of perturbation delta on FDSA results .......................62
Figure 4-5: Random initial estimates at fossil fuel cost of $500/MWh .......64
Figure 4-6: Random initial estimates at fossil fuel cost of $1750/MWh ....65
Figure 4-7: Initial decisions that converge to the true solution ...............66
Figure 4-8: Initial decisions that converge to an incorrect solution ........67
Figure 5-1: Power generated, Budischak dataset ..................................69
Figure 5-2: Energy in storage, Budischak dataset ..................................69
Figure 5-3: Fossil fuel generation, Budischak dataset ............................70
Figure 5-4: Power generation, historical generation dataset ....................71
Figure 5-5: Energy in storage, historical generation dataset ........................................ 72
Figure 5-6: Fossil fuel generation, historical generation dataset ............................. 73
Figure 5-7: Percent coverage and LCOE over a single run of FDSA ....................... 75
Figure 5-8: Decision variables over a single run of FDSA .................................. 76
Figure 5-9: Fossil fuel cost vs % coverage and LCOE for various scenarios ...... 78
Figure 5-10: Fossil fuel cost vs decision variables for various scenarios .......... 80
Figure 5-11: Results for Budischak dataset, 2008 costs ..................................... 82
Figure 5-12: Decision variables for Budischak dataset, 2008 costs ................. 83
Figure 5-13: Results for Budischak dataset, 2030 costs .................................... 84
Figure 5-14: Decision variables for Budischak dataset, 2030 costs ............... 85
Figure 5-15: Results for historical generation dataset, 2008 costs ............... 87
Figure 5-16: Decision variables for historical generation dataset, 2008 costs .... 88
Figure 5-17: Results for historical generation dataset, 2030 costs ................. 89
Figure 5-18: Decision variables for historical generation dataset, 2030 costs .... 90
Figure 5-19: Bid stack percent coverage and LCOE ........................................ 92
Figure 5-20: Bid stack optimal decisions ......................................................... 92
LIST OF TABLES

Table 2-1: Summary statistics for Budischak dataset .............................................. 28
Table 2-2: Correlation coefficients, Budischak dataset ........................................ 28
Table 2-3: Summary statistics for historical generation dataset .......................... 32
Table 2-4: Correlation coefficients, historical generation dataset ....................... 33
Table 2-5: 2008 cost parameters ........................................................................ 34
Table 2-6: 2030 cost parameters ........................................................................ 34
Table 3-1: Decision variable limits ..................................................................... 59
Table 5-1: Bid stack percent coverage and LCOE ............................................. 92
This thesis is dedicated to my parents and to my sisters Selina and Julia.
I INTRODUCTION

Renewable energy sources offer a number of benefits over conventional methods of generating power. Wind and solar power require no fuels and emit no greenhouse gases during daily operation and thus are sustainable in the long term. We live on an Earth that is on track to grow warmer by up to 6 degrees Centigrade over the next century, and fossil fuel combustion is one of the leading causes. Furthermore, in a world with an ever-diminishing supply of fossil fuels, moving toward renewable sources is a matter of energy security and national security for the United States.

However, several key obstacles lie ahead in our journey to fossil-fuel-independence. Apart from social and political issues, the high costs associated with renewable energy generation are an important barrier. Furthermore, the intermittent nature of wind and solar power generation provides a daunting challenge; if energy is often generated when it is not needed, and lacking when it is needed, how can we ensure that the lights are always on?

1.1 PJM INTERCONNECTION

We take the PJM electric grid as an example. PJM acts as an RTO (regional transmission organization) that coordinates an electricity market across state boundaries in the Mid-Atlantic region of the United States. Within such a market, demand varies stochastically and in response, LMPs (locational marginal
pricing) can spike to many times their average levels. To further complicate the problem of guaranteeing constant access to electricity, the input of solar and wind energy into the grid is intermittent and varies dramatically. Solar generation exhibits volatile behavior especially on days with scattered clouds; the presence of wind at any given location is also difficult to predict, regardless of whether the wind turbines are located on-shore or off-shore.

PJM runs two wholesale electricity spot markets: a Day-Ahead market and a Real-Time market. The Day-Ahead market determines hour-by-hour prices a day in advance of when the power will be generated and delivered. Generators, including wind and solar farms, power plants, and turbines, submit a bid indicating the minimum price at which they will sell their power the next day, and the maximum amount of energy (in kWh) they can supply. Wholesale buyers submit asks indicating the amount of electricity they will buy the next day. The next day, the buyers and sellers are then obligated to buy and sell the amounts they indicated at the prices they indicated. However, market actors are unable to accurately predict conditions a full day in advance. To allow generators and wholesale buyers to make trades throughout the day based on real operating conditions, the PJM Real-Time spot market determines prices every five minutes based on supply and demand. In this way, demand can be met even on short time scales.

Nuclear and coal-burning power plants are able to provide power deterministically to meet base load, but have long ramp-up times. Though they
can reliably supply power at low cost, they must know a day in advance whether they will be running the next day, and how much power to supply during the day. Though power plants have the ability to ramp up and down, they are unable to do so quickly enough to meet sharp decreases and increases in demand. Thus, peak load must often be met by other power generation technologies, such as turbines burning natural gas, which can ramp up and down relatively quickly (within a few minutes).

The presence of photovoltaic cells and wind turbines on the grid exacerbates the volatility in the system; RTOs must not only account for the stochastic nature of demand, but also that of the supply of energy. When wind and solar comprise a large portion of the energy generation on a grid, situations can easily arise in which both renewable sources are unable to produce power and baseload power plants cannot ramp up quickly enough to meet demand. In this case, gas turbines and other peaking power plants must kick in to meet demand, which is costly. In other cases, there may be an oversupply, and base load power plants will not sell enough electricity to cover the operating cost for that day, while much of the energy generated by solar and wind will go to waste.

The obvious solution is to use storage technologies for excess energy generated by wind and solar, which can smooth out the supply of energy coming from those renewable sources. Batteries are on the edge of becoming a financially viable option for acting as a load-balancing player on the market but may still be more expensive than utilizing a gas turbine to meet peak demand. A central part
of this thesis attempts to answer the following question: with so many options for managing the supply of power, which combinations of technologies allow for a least cost solution? Furthermore, how do fossil fuel costs and renewable energy costs affect the least-cost solution?

1.2 PRIOR STUDIES

Many studies prior to this one aimed to understand how current energy infrastructure could be replaced with renewable energy technologies. Reports such as those published by the National Renewable Energy Laboratory offer suggestions for how to coevolve energy systems through qualitative analysis and case studies [1]. Many studies also addressed cost and feasibility, sometimes with wildly optimistic results. The NREL’s Renewable Electricity Futures Study concluded that 80% of all generation could be performed by renewables by 2050, while still meeting load every hour [2]. Jacobson and Delucchi ambitiously conducted a two-part study to show that the entire world’s energy needs could be met with wind, water, and solar power (WWS) at reasonable cost [3][4]. This conclusion was achieved in two parts. The first part of the Jacobson study is an exercise in accounting. The study considers various costs, supplies, demands, and resource limits to understand how energy needs might theoretically be met given the availability of land, minerals, and bodies of water.

The second part of the study suggested several different approaches to mitigating for the intermittent nature of some renewable energy sources by connecting geographically dispersed sites of generation, diversifying over many
different power sources, storing excess off-peak energy, and implementing
demand-side management of load. Notably, part of the study involves an hour-by-
hour Monte Carlo simulation of load and generation over the course of two years
using data from California. The energy is generated by five different renewable
energy sources, and in Delucchi’s model, this diversity of sources alone was
enough to meet 99.8% of load. However, there was no mention of cost with
regard to the simulation.

The Delucchi paper references many other studies to show that a bevvy
of other strategies can be feasibly implemented to overcome challenges related to
stochasticity in energy markets. The study concludes by stating that converting
about 1.16% of global land area for renewable energy generation will allow us to
supply the entire world’s energy demand in excess. In addition, the cost of doing
so in 2030 would be similar to the total cost (including externalities) of fossil fuel
generation today. However, since much of the ‘total’ cost of fossil fuel generation
is in externalities, the cost of fossil fuel generation to consumers would be much
lower than the cost of supplying the world’s energy through wind, water, and
solar power.

The Jacobson and Delucchi study does not analytically consider the
economics of transitioning from a fossil-fuel economy to a renewable energy one,
and neither does it try to tackle the issue of how to find least-cost solutions.
Would the market, for example, be motivated to transition to renewable energy
sources at today’s fossil fuel costs? Would storage technology or forecasting
technology need to improve greatly in order to allow renewables to be more cost-effective than fossil fuel or nuclear power? How do we decide whether to spend on storage or more diversified generation? The only simulation in the Jacobson paper did not consider costs, and there is very little documentation in the paper about how the simulation was accomplished. The result is that while the Jacobson study does provide us with a goal to strive toward, it seems at times to be somewhat removed from the reality that we live in.

A study done by Budischak et al. sought to introduce a more sophisticated model of the energy grid to understand what a least-cost solution to the renewable energy problem would look like at a regional level. The Budischak study asked the question: if we were to mandate that a certain percentage of load had to be covered by renewables in the PJM electrical grid, what would be the least-cost solution? His conclusion was that in 2030, a least-cost configuration of wind power, solar power, and battery storage could power the PJM grid 99.9% of the time at costs that are comparable to today’s fossil fuel costs. Budischak used an exhaustive search to simulate and evaluate 28 billion different combinations of wind, solar, and storage over the course of four years of historical data. Budischak’s least-cost configurations would power the grid 30%, 90%, and 99.9% of the time. Notably, each of the levelized costs of electricity (LCOEs) found in the Budischak study were below 50 cents per kWh, for both 2008 and 2030 cost parameters. It is important to note that the Budischak study optimized over cost of renewable infrastructure only; the optimization model did not consider fossil fuel energy costs, although they were included in calculating the
LCOEs. As a result, the Budischak model does not fully consider the role of fossil fuel energy costs in the market-driven energy grid.

1.3 THIS STUDY

This study seeks to use the Budischak simulation model to solve a set of slightly different problems that consider fossil fuel generation more fully. In essence, we vary the cost of fossil fuel energy from $50/MWh, which is the current wholesale price, up to $2500/MWh, which is an extraordinarily high cost for energy and is experienced very infrequently. At each fossil fuel energy cost, we ask what the optimal mix of renewable technology build-outs are, with the assumption that all energy not met by renewables and battery are met using fossil fuel generation. Since this problem is much broader and computationally intensive than the problem posed by Budischak, we also propose a new way to solve for the least cost solution: the finite-difference stochastic approximation algorithm instead of an exhaustive search. Along the way, we will consider new ways to calculate the cost parameters that better represent the cost of renewable energy generation. In addition, we introduce a newer, more recent dataset to use as an input to the model, and finally, we introduce the use of a bid stack to represent fossil fuel costs instead of a constant fee.

The problem that we are trying to solve is, in effect, a variant of the classical newsvendor problem, albeit with a five-dimensional decision variable. Instead of looking at newspaper sales over the course of a day, we examine the consumption of electricity over the course of multiple years. However, our
problem is still a one-period model, because the decision variable is decided at the beginning of the time period and remains unchanged throughout. Similarly to the newsvendor problem, we must decide how much ‘inventory’ of power generation capacity to supply, given unpredictable demand. The inclusion of a fossil fuel backup fee in our model penalizes us for not supplying enough power, while at the same time, we incur costs for each unit of inventory that we do supply. This is analogous to the opportunity cost of not stocking enough newspapers to meet demand, while also incurring a fixed cost per newspaper that is stocked. The solution to the newspaper problem is to look at the expected value of the contribution function; its maximum occurs when the expected revenue from each unit of increased inventory equals the per-unit cost of supplying that inventory. Our problem is slightly different in that we know what demand will be like, however the relationship between ‘inventory’ levels and the objective function can be determined only through simulation, due to the complexity of the model and the large number of decision variables to be considered. However, we can expect that the behavior of the objective function will be similar to that of a newsvendor problem. We simply need to supply the amount of power generation capacity that is not too much or too little; like Goldilocks’ porridge, it needs to be just right.

2 THE SIMULATION MODEL

This chapter establishes a mathematical model to simulate the generation and consumption of energy within the PJM grid on an hour-by-hour basis. This model closely matches the one used by Budischak, though it offers additional
flexibility in how fossil fuel costs are calculated. The simulation determines the amount of renewable energy generated during each hour and matches the generated energy to the load during that hour. Generation in excess of load is stored in a central battery, whereas if energy generated by renewables combined with energy in storage is lower than the load for that hour, conventional energy sources must be used to meet demand for electricity. We impose a strict constraint that all load is met every hour, with the assumption that no demand-side management occurs, in order to simply the model. In instances where supply exceeds demand, excess generation is simply spilled or discarded at zero value. To simplify the computational complexity of the model, it is assumed that there will always be enough fossil-fuel energy to meet demand when renewables fail to meet load. In addition, the battery greatly simplified. Charge curves are not considered; the battery is assumed to have a certain fixed capacity throughout its entire lifetime, and the C-rate of the battery is assumed to be 1C or faster—enough to match the power rating of the battery inverter. The battery inverter is assumed to be the limiting factor to how much energy can flow in and out of the battery at any given time. The battery is assumed to have a certain round trip efficiency (RTE) and hourly self-discharge rate, which are included in the model because they affect the quantity of energy available to use during any given hour.

Unlike the model used by Budischak et al., this simulation model accounts for fossil fuel costs that are incurred when renewable energy sources and battery storage are unable to meet demand; furthermore, these fossil fuel costs are included when the model is used to cost-optimize the generation/storage
technology mix. This pragmatic perspective on the relationship between build-out of renewable generation and total cost of energy acknowledges that a market-driven system would balance the high cost of renewable energy infrastructure with the cost of purchasing fossil fuel energy on-demand in order to arrive at an optimal solution. Optimizing for total cost, including cost of ‘fast’ fossil fuel generation, is a holistic way to understand the costs associated with the question of energy generation, compared to setting arbitrary benchmarks for the percent of load that must be covered by renewable sources as is done by Budischak et al.

The underlying assumption of this method of modeling costs is that fast-ramping fossil fuel generation (such as natural gas turbines) must be ramped up during any hour in which renewable sources of energy are less than the total PJM load. Ramp-up rates for these peaking power plants are presumed to be fast enough to supply energy on-demand on an hour-by-hour basis. Alternatively, as is pointed out by Budischak, if the RTO (PJM in this case) is able to forecast generation and load to any extent, slow-ramping base load generation technologies can be used to charge batteries, which can then be discharged when renewable generation fails to meet load. These ‘slower’ sources of power are primarily coal-burning plants or nuclear plants. In addition, the model can be used to understand scenarios in which a carbon tax or a cap-and-trade system adds additional cost to the burning of fossil fuels; the dollar cost per MWh of buying fossil fuel generation simply increases if a carbon tax is present.
Within the model, the cost per MWh of fossil fuel energy is a function of the unmet load. The simplest way to model this fossil fuel cost is to assume a fixed dollar value per MWh of load that is met via non-renewable sources. A more realistic cost function would be to use a bid stack (also called a generation stack), in which the marginal price of each successive MW of generation would be strictly non-decreasing [5]. With a bid stack, energy prices differ for different technologies; renewable generation has the lowest cost, whereas oil-burning turbines generally have the highest cost. The high cost of energy from peaking power plants is not only due to the cost of fuel, though the upper portions of the bid stack are highly correlated with gas and oil prices [5]. The high cost also accounts for the fact that peaking power plants only sell energy when base load plants are unable to meet load, which may be a few times a day or a few times a year. Thus, the revenue made when selling during those few days must cover the cost of maintaining the plant during the rest of the year.

![Figure 2-1: A typical PJM bid stack](image-url)

Figure 2-1: A typical PJM bid stack
Furthermore, the model assumes perfect transmission and zero transmission cost, which would unnecessarily complicate the simulation. In addition, we model the PJM grid as an isolated entity, so spillover to adjacent systems is not considered. As noted by Budischak, transmission costs would increase the cost of power, whereas sale of excess power to adjacent markets would decrease cost; all things considered, this study’s model likely overestimates cost, since sale to adjacent markets would more than offset transmission costs [6].

For this thesis, the simulation model is run with two different sets of load and capacity factor data. One dataset is the one used by Budischak, which uses weather data to extrapolate capacity factors; the other is compiled from actual power output data from PJM and PSE&G. When using the first dataset, the model runs for four years and nine months; however, the first nine months are run only to charge the battery ahead of the four-year period being analyzed and optimized over. The second dataset contains data for only 347 days; therefore, none of the data is allocated for a ‘pre-charging’ period and the optimization covers the entire range of days.

2.1 Model

A mathematical formulation of the simulation model is framed similarly to a dynamic program, using a notational style described in Approximate Dynamic Programming by Powell [7]. The model takes a vector of decisions variables as inputs, which the operator of the model controls; the model generates
a succession of state variable vectors, one per hour of simulated time. The model also requires the input of exogenous information, which describes data that is random and/or beyond the control of the model. A transition function dictates how the state evolves from one hour to the next and is dependent on the decision variables, the exogenous information, and the previous hour’s state variable. Finally, the objective function is the quantity that will be optimized. Note that for all variables, time indices begin at zero and end at one less than the total number of hours simulated.

2.1.1 STATE VARIABLE

Within the simulation, the state variable at any given hour consists only of the level of charge of the battery at the beginning of the period. The state variable is defined as follows:

\[ S_i = \{ R_i \} \]

Where \( R_i \) = amount of energy stored in batteries at time \( t \)

for \( t_0 < t < T \)

A note on time indexing: the evolution of the system occurs in discrete time steps \( t \), with each time step representing one hour. The state variable for time \( t \) describes the state of the system at the beginning of the \( t^{th} \) hour, whereas exogenous information becomes known during that hour. Thus, the \( 0^{th} \) state
variable is before the first hour of generation, and the 0th hour is the hour directly after the 0th state variable is recorded.

2.1.2 DECISION VARIABLES

The decision variables consist of the build-outs of each of five different technologies for generation and storage: solar photovoltaic, offshore wind, onshore wind, battery inverter power, and battery storage capacity. Each of the decision variables is measured in megawatts or megawatt-hours to indicate the total nameplate power rating or storage capacity for that technology. A unique vector of decision variables represents a particular configuration of renewable technology build-outs connected to the PJM grid. For example, $x_{pv}^p$ represents the sum of all nameplate capacities of solar farms connected to the PJM grid under a hypothetical scenario. Storage technology is modeled as if it were one larger central battery connected to the grid. The amount of energy that can be stored in the battery is determined by its storage capacity (in MWh), whereas the rate of charge and discharge is determined by the power rating of the inverter connected to the central battery, which is required in order to supply AC energy to the grid and prevent islanding. A decision variable set at zero would indicate an absence of that technology within the PJM grid. The decision variables are as follows:

$$x = (x_{pv}^p, x_{off}, x_{on}^p, x_{st}, x_{sc})$$

$$x_{pv}^p = \text{photovoltaic generation capacity (MW)}$$
\[ x^{\text{off}} = \text{offshore wind generation capacity (MW)} \]

\[ x^{\text{on}} = \text{onshore wind generation capacity (MW)} \]

\[ x^{m} = \text{battery inverter power rating (MW)} \]

\[ x^{sc} = \text{battery storage capacity (MWh)} \]

\[ x^{pv}, x^{\text{off}}, x^{\text{on}}, x^{m}, x^{sc} > 0 \]

\subsection*{2.1.3 Exogenous Information}

The exogenous information \( W_t \) used in the model consists of the load at each hour and the hourly capacity factors for each of the three generation technologies. For this thesis, \( W_t \) is given by historical data and is known in advance of running the optimization, making the optimization a deterministic problem. However, we will model the exogenous information \textit{as if} it is stochastic, since the deterministic problem is a sub-problem of the stochastic one. Modeling the exogenous information as stochastic provides flexibility for the model to take in real-time weather conditions, or random capacity factors and load generated by a statistical model, e.g. for a Monte Carlo simulation. The exogenous information is as follows:

\[ W_t = (L_t, W_t^{pv}, W_t^{\text{off}}, W_t^{\text{on}}) \]

where \( L_t = \text{total load in PJM at time } t \), and
\( W_{t}^{pv}, W_{t}^{off}, W_{t}^{on} \) represent hourly capacity factors for solar photovoltaic generation, offshore wind generation, and onshore wind generation, measured in MWh per MW. These are derived either from weather data (solar insolation and wind speeds) or from actual generation data in conjunction with nameplate capacities.

\[ L_{t} > 0 \text{ and } W_{t}^{pv}, W_{t}^{off}, W_{t}^{on} \in [0,1] \]

### 2.1.4 Transition Function

The transition function below determines new values of the state variable for each successive time step. During each hour of simulation, the model determines the amount of energy generated, compares that to load, and then determines the Energy Into Storage (EIS), Energy Out of Storage (EOS), and the final level of storage in the battery, \( R_{t} \). The transition function is as follows:

\[
S_{t+1} = S^{M}(S_{t}, x, W_{t})
\]

Let \( G_{t} \) equal the total energy generated during hour \( t \).

\[
G_{t} = x^{pv}W_{t}^{pv} + x^{off}W_{t}^{off} + x^{on}W_{t}^{on}
\]

Let \( \eta \) be the round-trip efficiency of the battery, and let \( \lambda \) be the battery’s fraction of leakage per hour.
Energy Into Storage (EIS) = $E_i^+ = \inf \left\{ \left( (G_i - L_i)_{+}, x^+, \lambda^+ \right) \right\}$

Energy Out of Storage (EOS) = $E_i^- = \inf \left\{ \frac{(L_i - G_i)}{\eta}, x^-, (1 - \lambda) R_i \right\}$

where $f^+$ indicates the positive part of $f$,
equivalent to writing $\sup \{ f, 0 \}$

$$R_{i+1} = \inf \left\{ (1 - \lambda) R_i - E_i^- + E_i^+, x^\prime \right\}$$

2.1.5 **OBJECTIVE FUNCTION**

The objective function for this problem is a loss function that represents the total cost of supplying power to load, given a vector of decision variables. This model accounts for the following components of total cost: the initial capital costs associated with purchasing and installing equipment, annual operation and maintenance (O&M) costs, and the cost of using non-renewable sources to cover load that is not met by solar, wind, and storage. This last component differentiates this study from that of Budischak. By including the cost of fossil fuel generation in the objective function of our optimization problem, we find the true least cost solution for any given cost scenario, whereas Budischak’s problem imposes arbitrary requirements for how many hours of load must be met by renewable sources. The objective function is formulated as follows:

Let $U_i(G_i, E_i^-, L_i)$ represent the load at each hour that is not met by renewables and energy in storage.
The objective function is:

\[
\min_x \{ F(x,U) \}
\]

for

\[ F(x,U) = n \left( c^{pv} x^{pv} + c^{off} x^{off} + c^{om} x^{om} + c^{mw} x^{mw} + c^{sc} x^{sc} \right) + \sum_{t=t_0}^{T} C^u(U_t) \]

s.t. \( x \in X \)

where:

\( n \) is the number of years being considered in the optimization. We take 365.242 to be the number of days in a year. For the Budischak dataset, \( n = 4 \), whereas for the historical generation dataset, \( n = 0.95 \).

\( c = \{ c^{pv}, c^{off}, c^{om}, c^{mw}, c^{sc} \} \) is the annualized cost per MW (or per MWh) of solar, offshore wind, onshore wind, battery inverter, and battery storage technology. It is the sum of an annualized cost of capital plus annual O&M costs. For more information about how these cost parameters are calculated, please refer to Chapter 2.4 (Parameters and Costs).

\( C^u(U_t) \) is the cost function for supplying power on-demand to meet load that is not met by renewables and storage (recall that this load is represented by \( U_t \)). Within this thesis, we use two different types of cost functions. The first is a constant \$/MWh fossil fuel penalty,

\[ C^u(U_t) = kU_t \]

for some positive \( k \) and a bidstack \( C^u_B(U_t) \) for which the
marginal cost of electricity depends on the total amount of load that must be serviced that hour.

and finally, $\mathbf{x}$ is the domain of all possible decision variables. Note that $x$ must be non-negative—it makes no sense to have negative build-out of a technology. In addition, an upper bound for each dimension of $x$ is set according to the upper power limits for solar, offshore wind, and onshore wind, as defined in Budischak et al. Battery inverter power and battery storage are assumed to not have an upper bound.

Note that in order to convert the total cost (represented by the objective function) into a levelized cost of energy (LCOE), we must divide the total cost by the total load over the course of the $n$ years being studied. Since all costs are recorded as non-discounted numbers, this calculation results in the average cost per kWh over the course of those $n$ years.

2.2 STORAGE POLICY OPTIMALITY

Prior research has been done on storage policies in the presence of stochastic load and power generation. They tackle problems such as: in a model of the energy market in which grid-connected storage can sell (discharge) and buy (charge) from the market at any time, what policy should the battery operator employ in order to maximize profit through energy price arbitrage? Or, as with Sami Yabroudi’s thesis, when multiple small timescale energy storage devices are connected to a wind turbine and a building in a standalone system, what storage control policies maximizes energy delivered to the building [8]?
Like our current study, these problems deal with storage control when supply and demand vary stochastically.

Our simulation model uses a myopic policy for storage control. During each hour, if there is excess generation, as much of it as possible is stored, given the battery’s storage capacity; if there is any unmet load, the battery will be discharged to satisfy as much of that load as possible. Our policy \( \pi_0 \) is formulated as follows:

For \( t = t_0, t_0 + 1, \ldots, T \)

Let \( X_t^\pi \) be the energy discharged from storage at time \( t \) under policy \( \pi \).

Negative values indicate charging. Recall that \( \eta \) is the Round Trip Efficiency of the battery, which is modeled as the fraction of the discharged energy that is available to be used to meet load.

\[
X_t^{\pi_0}(S_t) = \begin{cases} 
-\inf \left[ G_t - L_t, x_t^u, x_t^{sc} - (1 - \lambda) R_t \right] & \text{if } G_t > L_t \\
0 & \text{if } L_t = G_t \\
\inf \left[ \frac{L_t - G_t}{\eta}, x_t^u, (1 - \lambda) R_t \right] & \text{if } G_t < L_t 
\end{cases}
\]

Recall that we want to minimize the following objective function:

\[
\min_{\pi} \left\{ h \left( c^{pv} x^{pv} + c^{off} x^{off} + c^{on} x^{on} + c^{st} x^{st} + c^{sc} x^{sc} \right) + \sum_{t=t_0}^{T} C^{\pi}(U_t) \right\}
\]
However, since $x$ is decided before $t_0$, the first term is effectively a constant when considering this sub-problem that occurs during each hour. In addition, recall that $U_t = L_t - \left( G_t + \left[ X^\pi_t \right]^+ \eta \right)$. Thus the objective function above is equivalent to:

$$\min_{\pi} \left\{ \sum_{t=t_0}^T C^\pi \left( L_t - \left( G_t + \left[ X^\pi_t \right]^+ \eta \right) \right) \right\}$$

Does our policy $\pi_0$ minimize this objective function? An alternate policy for controlling Energy Out of Storage (EOS) may be to discharge the battery only if the unmet load exceeds a certain level, in order to smooth out the unmet load over time. It turns out that optimality of the battery control policy depends on $C^u(U_t)$, the fossil fuel cost function applied to unmet load. Below, I will show that when $C^u(U_t) = kU_t$, for some positive constant $k$, our policy $\pi_0$ is optimal. However, when the fossil fuel cost function is dictated by a bid stack, the myopic policy fails to minimize the loss function. Because the bid stack cost function is an increasing function, smoothing out the unmet load over all time periods results in a lower-cost solution.

We tackle the case when $C^u(U_t) = kU_t$. We will deal primarily with the battery discharge policy, since it is trivial to prove that the myopic charging policy is optimal. (Our model assumes that spilled energy has zero value, and there is no cost associated with storing excess generation. Any energy that is
stored has a non-zero probability of deducting a positive amount from the loss function, thus it is optimal to maximize Energy Into Storage (EIS) at each hour.

We offer a proof by contradiction to show that the discharge policy is optimal. Suppose that an alternate policy $\gamma$ is optimal. Since this new policy is different from $\pi_0$, we know that for at least one hour $t'$, this alternate policy will choose to not discharge the battery to its full extent. We know that $X_{t'}^{\pi_0}$ (the energy discharged under our original policy) is the maximum amount of energy that could be discharged to satisfy unmet load during time $t'$. We know that:

$$X_{t'}^{\pi_0} - X_{t'}^{\gamma} = d, \ d > 0$$

Let $U_{t'}^{\pi_0}$ be the load unmet during time $t'$ under the original policy. The contribution to the loss function from that unmet load is: $kU_{t'}^{\pi_0}$.

Let $U_{t'}^{\gamma}$ be the unmet load during time $t'$ under the alternate policy.

$$U_{t'}^{\gamma} = U_{t'}^{\pi_0} + d, \ \text{since} \ X_{t'} + U_{t'} = L_{t'} + E_{t'} - G_{t'}, \ \text{for any} \ X_{t'}.$$

Thus we know that the contribution to the loss function from the unmet load during time $t'$ under policy $\gamma$ is: $k(U_{t'}^{\gamma}) = k(U_{t'}^{\pi_0} + d)$.

However, because policy $\gamma$ keeps an additional $d$ units of energy in the battery, this extra stored energy will go on to deduct from the loss function at a future time $t''$. By the time these $d$ units of energy are discharged from the battery for use, they will have degraded due to self-discharge of the battery. Thus the deduction from the loss function due
to future use of $d$ units of energy is: $-kd(1-\lambda)^{t'-t}$, where lambda is the self-discharge rate.

Thus the net contribution to the loss function under policy $\gamma$ is:

$$ k(U_{t'}^{\pi_0} + d) - kd(1-\lambda)^{t'-t} $$

$$ = k\left[U_{t'}^{\pi_0} + d - d(1-\lambda)^{t'-t}\right] $$

$$ = kU_{t'}^{\pi_0} + kd(1-(1-\lambda)^{t'-t}) $$

The term $kd(1-(1-\lambda)^{t'-t})$ is positive, so the net contribution to the loss function under policy $\gamma$ is greater than the net contribution to the loss function under policy $\pi_0$, which was $kU_{t'}^{\pi_0}$. This is in direct contradiction to our initial assumption that policy $\gamma$ is optimal, since an optimal policy should not result in a larger value of the loss function. Thus, we know that our initial assumption was not true; there exists no alternate policy that performs better than our original myopic policy $\pi_0$.

2.3 **CAPACITY FACTOR AND LOAD DATA**

This study makes use of two separate sets of input data for the simulation, the first of which is the data used in the Budischak study. The
Budischak dataset includes over 41,500 hours worth of hourly historical data from 1998 to 2002 [6]. This includes load data from PJM, as well as hourly capacity factors (in MWh per MW of capacity installed) for photovoltaics, offshore wind, and onshore wind, which are calculated using meteorological datasets combined with a certain level of extrapolation. The second dataset draws from actual generation data in 2012 and 2013, obtained from PJM and Public Service Electric and Gas Company (PSE&G); this data comes from up-and-running wind farms in the PJM region in addition to photovoltaic arrays in New Jersey. In addition, the dataset includes hourly integrated load data from PJM. In the figures to follow, note the intermittent nature of each of the three sources of energy, as well as the unpredictable, stochastic nature of the hourly capacity factors and load. In order to create a model that would be computationally efficient, a single RTO-wide capacity factor was calculated for each renewable energy source per hour, aggregating the capacity factors for multiple generation sites. This simplifies the problem of optimizing for least cost and assumes an RTO with perfect transmission properties.

The computation model used in this thesis allow for easy input of alternative datasets. The model, which is written in MATLAB, accepts a comma-separated values file with hourly load, solar CF, offshore wind CF, and onshore wind CF for any number of hours.
2.3.1 DATA FROM BUDISCHAK ET AL

The Budischak study uses data from the period between April 1st, 1998 and December 31st, 2002 (inclusive) in order to model the PJM grid as it was before it started a phase of growth and expansion in 2002. The PJM Interconnection served Pennsylvania, New Jersey, and Maryland until Allegheny Power joined in April of 2002 as PJM’s first external market participant. During the 1998 to 2002 period, the average hourly load of the PJM grid was a mere 31 GW, less than a third of the 101 GW average load served in 2014. The load data used in the Budischak study comes directly from PJM’s records of historical integrated hourly load, collected from raw telemetry data.

Though the Budischak dataset spans from April 1998 to December 2002, the optimization only looks at the four years of time between January 1999 and December 2002. The simulation is run for 9 months preceding the start of the period under scrutiny in order to start the battery off with a realistic amount of energy in storage. Thus within the simulation model, \( t_0 \) is the first hour of April 1, 1998, and \( T \) is the last hour of Dec 31, 2002. However, the objective function only considers time periods from 6601 to \( T \) (there are 6600 hours between April 1\(^{st}\), 1998 and January 1\(^{st}\), 1999).
The wind and solar data used in the Budischak study comes from the same time period as the load data. In order to create hourly capacity factors for onshore wind generation, wind speeds were collected from various meteorology stations, the speeds were extrapolated from measurement height to a turbine hub height of 80 meters, and then speeds were converted to power output via a Repower 5M commercial turbine power curve. An initial 135 locations were screened and then winnowed down to 23 sites; a cutoff of 30% annual capacity factor was used to select these locations (see figure below). The wind speed data were obtained from the National Climate Data Center. The process of screening potential sites for developing wind farms resulted in an overall capacity factor of 40%, which is somewhat high, as Budischak acknowledges in his paper. Offshore wind data followed an identical approach; however, the offshore wind speed data originated from NOAA buoys, the locations of which are shown below [6][9].

Figure 2-2: PJM hourly load, Apr. 1998 to Dec. 2002
Hourly capacity factors for photovoltaic generation were calculated using irradiation data from the National Renewable Energy Laboratory’s (NREL). The irradiation values were then converted to output power using NREL’s PVWatts program. Notably, the Budischak paper uses irradiation data from a single city within the PJM system—Wilmington, Delaware. Within the PJM region, Wilmington is situated at a mid-latitude location; however, it may not be realistic to use irradiation data from a point source to model situations in which photovoltaic build-out is extensive. One would expect the volatility of solar generation to decrease when panels are located across a diverse geographic range, though the primary source of volatility is the periodic 24-hour fluctuation of solar irradiation.
**Figure 2-4: Hourly capacity factors for first week of Apr. 1998**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load</td>
<td>31,032.83</td>
<td>6812.51</td>
<td>17461.00</td>
<td>64127.00</td>
</tr>
<tr>
<td>Solar CF</td>
<td>0.16</td>
<td>0.25</td>
<td>0.00</td>
<td>0.94</td>
</tr>
<tr>
<td>Offshore Wind CF</td>
<td>0.41</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Onshore Wind CF</td>
<td>0.40</td>
<td>0.21</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Table 2-1: Summary statistics for Budischak dataset**

<table>
<thead>
<tr>
<th>Load</th>
<th>Solar CF</th>
<th>Offshore Wind CF</th>
<th>Onshore Wind CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load</td>
<td>100%</td>
<td>-7.83%</td>
<td>-0.07%</td>
</tr>
<tr>
<td>Solar CF</td>
<td>100%</td>
<td>-0.84%</td>
<td>-17.58%</td>
</tr>
<tr>
<td>Offshore Wind CF</td>
<td>100%</td>
<td></td>
<td>46.11%</td>
</tr>
<tr>
<td>Onshore Wind CF</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2-2: Correlation coefficients, Budischak dataset**
2.3.2 Historical Power Output Data

The second dataset used in this thesis seeks to complement the one used in the Budischak study by providing a more recent dataset of hourly capacity factors derived from already-installed wind farms and photovoltaic arrays. The PJM interconnection is currently connected to over 6 GW of onshore wind generation, as determined from generation data provided by PJM; the total wind nameplate capacity of the PJM grid is 19 GW when including projects that are both installed and underway [10]. See the figure below for the locations of each of PJM’s up-and-running onshore wind farms, as of April 2014. In addition, about 2 GW of solar power is active and under construction [10]. PJM currently does not have any offshore wind generation, despite the potential to supply close to 100 GW of power on average if offshore wind is fully exploited [11][12], so this dataset does not include offshore wind capacity factors.

![Hourly capacity factors for first week of Sep. 2012](image)

The dataset covers a period of 347 days, from September 1st, 2012 to August 13, 2013. The hourly load data within the second dataset comes PJM’s
online records [13]. The onshore wind data comes from PJM’s records of 5-minute power output from each of its wind farms. This study uses data from 45 such sites, with certain wind farms excluded because of poor quality of data or because they were not operational during the entire period studied. To calculate hour capacity factors, 5-minute power outputs were converted to energy generated per hour, which were then summed up across all 45 farms, and then divided by the total nameplate capacity for the entire set of wind farms. The use of actual PJM wind farms removes several speculative aspects found in the Budischak data: choosing sites for wind generation, extrapolating measurement height wind speeds to hub height, and deriving power output from wind speed data. In addition, the use of actual pre-installed sites may reduce the upward bias present in the Budischak onshore wind capacity factor data, which resulted from their selection bias of only choosing sites with high capacity factors.

![Figure 2-6: PJM onshore wind farm locations](image)
The solar hourly capacity factors come from records of solar power output at 20 different photovoltaic panel arrays at 19 different locations managed by PSE&G as part of their Solar 4 All program (see figure below for a map of all 19 sites). The Solar 4 All program entails the construction and operation of 45MW of planned solar photovoltaic facilities on landfills and brownfield sites [14], however this study included only about 27 MW of capacity that was fully operation during the studied time period. Hourly capacity factors were derived from power output data collected at 5-minute intervals in the same manner as with the onshore wind data. The use of a geographically diverse set of solar arrays is an important difference between this dataset and the one used in the Budischak study. Using data from various locations throughout New Jersey provides a more representative time series of solar irradiation. In addition, as with the PJM wind data, the use of actual up-and-running generation technology allows our model to use data that more realistically captures the availability of solar power over time.
**Table 2-3: Summary statistics for historical generation dataset**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load</td>
<td>88,549.33</td>
<td>15,250.75</td>
<td>56,814.00</td>
<td>155,334.00</td>
</tr>
<tr>
<td>Solar CF</td>
<td>0.16</td>
<td>0.23</td>
<td>0.00</td>
<td>0.90</td>
</tr>
<tr>
<td>Onshore Wind CF</td>
<td>0.29</td>
<td>0.19</td>
<td>0.00</td>
<td>0.84</td>
</tr>
</tbody>
</table>

*Figure 2-7: PSEG solar panel sites*
<table>
<thead>
<tr>
<th></th>
<th>Load</th>
<th>Solar CF</th>
<th>Onshore Wind CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load</td>
<td>100%</td>
<td>29.68%</td>
<td>-10.40%</td>
</tr>
<tr>
<td>Solar CF</td>
<td>100%</td>
<td></td>
<td>-10.10%</td>
</tr>
<tr>
<td>Onshore Wind CF</td>
<td></td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2-4: Correlation coefficients, historical generation dataset*

2.4 PARAMETERS AND COSTS

The simulation model uses several parameters to determine costs and battery behavior. These parameters have a dramatic effect on the outcome of the optimization, especially with regard to the relationship between cost and the number of simulation hours covered solely by renewable energy. The parameters used in this model include: the fossil fuel cost function, annualized capital costs, yearly operation and maintenance (O&M) costs, round trip efficiency (RTE) of the battery, and self-discharge rate of the battery. For rapidly developing technologies, estimating costs is a tricky matter. The challenge is even greater for choosing a technology for aggregate grid-level storage. For this reason, the computation model implemented for this thesis allows for easy input of alternative cost parameters. The MATLAB model is able to read any properly formatted comma-separate values file that contains the relevant parameters. The following two tables show the two sets of cost parameters used for this thesis: historical costs from the year 2008 as well as projected costs for 2030.
### Table 2-5: 2008 cost parameters

<table>
<thead>
<tr>
<th>Technology</th>
<th>Overnight Capital Cost ($/MW or $/MWh)</th>
<th>Lifetime (yrs)</th>
<th>Capital Cost Annuity ($/yr)</th>
<th>O&amp;M ($/MW or $/MWh per yr)</th>
<th>Total Annual Cost ($/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV (MW)</td>
<td>6,762,560</td>
<td>20</td>
<td>905,363</td>
<td>13,082</td>
<td>918,445</td>
</tr>
<tr>
<td>Offshore wind (MW)</td>
<td>4,313,120</td>
<td>20</td>
<td>577,435</td>
<td>100,218</td>
<td>677,653</td>
</tr>
<tr>
<td>Onshore wind (MW)</td>
<td>2,153,760</td>
<td>20</td>
<td>288,343</td>
<td>33,936</td>
<td>322,279</td>
</tr>
<tr>
<td>Battery inverter (MW)</td>
<td>748,720</td>
<td>10</td>
<td>132,512</td>
<td>13,087</td>
<td>145,599</td>
</tr>
<tr>
<td>Li-titanate storage (MWh)</td>
<td>336,784</td>
<td>10</td>
<td>59,605</td>
<td>-</td>
<td>59,605</td>
</tr>
</tbody>
</table>

### Table 2-6: 2030 cost parameters

Cost data was obtained from a broad body of pre-existing literature seeking to understand the cost of implementing renewable energy sources at high penetrations. Costs for solar and wind technologies come from Delucchi and
Jacobson’s study on hypothetically meeting global demand for energy through wind, water, and solar power [4]. Battery inverter costs come from a National Renewable Energy Laboratory study on equivalent PV power inverters [15]. Battery costs are derived from Burke & Miller’s Emerging Lithium Battery Test Project and an Argonne National Laboratory study on Li-ion batteries for EVs [16] [17]. All costs were adjusted using the average Consumer Price Index for the relevant years [18].

2.4.1 CALCULATING LEVELIZED COST OF ELECTRICITY

One of the main outputs of the simulation model is a Levelized Cost of Electricity (LCOE), which represents the price at which electricity must be sold from a specific source in order to break even over the lifetime of the project. In other words, the present value of the LCOE multiplied by the load served is equal to the present value of the cost of the project. The LCOE takes into account capital costs, O&M, and fuel, but it does not consider financing fees and future replacement. We use LCOE as an indicator of true cost because it considers the time value of money and can be compared to locational marginal pricing of electricity in PJM’s market today.

Our model calculates levelized cost of electricity (LCOE) using a novel approach. Ordinarily, LCOE is calculated using the following formula [19]:

\[
C_{LCOE} = \frac{C_K \theta_{CRF} + C_{O&M}}{8760 \kappa}
\]
\[ \theta_{CRF} = \frac{i}{1-(1+i)^{-M}} \]

where \( C_K \) is the overnight capital cost for the project (the one-time cost of equipment and installation incurred at the beginning of the project), measured in \$/MW.

\( \theta_{CRF} \) is the capital recovery factor, which is calculated using \( i \), the discount rate, and \( M \), the lifetime of the project. The term \( C_K \theta_{CRF} \) represents an annualized version of the overnight cost – as if the cost of the project is turned into an annuity whose present value at \( t_0 \) is equal to the overnight capital cost.

\( C_{O&M} \) is the annual operations and maintenance cost, measured in \$/MW per yr.

And \( \kappa \) is the capacity factor associated with the project—the average percent utilization of the facility’s power generation capacity over the lifetime of the project.

For our simulation, there are two complicating factors that force us to use a different calculation of LCOE. First, it is bothersome to measure the empirical capacity factors for each of the technologies in our model. Second, we must include the cost of fossil-fuel generation into our calculation of LCOE for any given configuration of renewable technology build-outs. Thus we use a different calculation of LCOE:
\[ C_{LCOE} = \frac{n \sum_{j \in \Omega} (C_k^j \theta_{CRF}^j + C_{O&M}^j) x^j + \sum_{t=t_0}^T C_t^u (U_t)}{\sum_{t=t_0}^T L_t} \]

where \( \Omega = \{pv, off, on, st, sc\} \), and \( n \) is the number of years being simulated.

The numerator is the total sum of costs over the duration of the simulation, undiscounted. \( \theta_{CRF} \) is calculated the exact same way as is the previous LCOE formula, but it is calculated for each technology separately. (Each technology uses the same discount rate, 12%, but each of them have different lifetimes.) The first term in the numerator represents costs due to annualized capital costs and O&M multiplied by the number of years being simulated. The second term in the numerator is the total cost due to fossil fuel usage. The denominator is total load served, undiscounted.

This formula allows us to avoid calculating average capacity factors explicitly, and it allows the ability to account for fossil fuel costs. The formula calculates total cost divided by total load while taking the time value of money into account to arrive at a levelized cost of electricity that is equivalent to the conventional formulation.

2.4.2 STORAGE TECHNOLOGY

For our simulation model, we chose to use a specific type of lithium-ion battery for grid-level storage. It is useful to understand that among rechargeable
batteries, lithium-ion chemistries are known for their high energy density and low rate of self-discharge (for example, in comparison to Ni-MH batteries or lead-acid batteries [20]), leading to their widespread use in consumer electronics and electric vehicles (EVs) [17]. Furthermore, lithium-ion batteries exhibit efficiencies of 85% or greater [21], and the quantity of lithium on earth is almost unlimited given the high concentration of lithium ions in seawater [22]. However, most commercially available lithium-ion chemistries are designed to last 3-4 years and would not be suitable for grid-level storage [23], though PJM does currently employ a 64 MW array of lithium-ion batteries in tandem with a wind farm in West Virginia [24]. However, the amounts of storage modeled here could be hundreds of GW.

The properties of a lithium-ion battery that affect its performance as a grid-level storage solution depend largely on the particular electrode material that is used. For our simulation model, we have chosen to use lithium-titanate batteries. Though lithium-titanate cells have never been used for grid-level storage, the technology is already commercially available and can be easily priced [25][26][27]. Lithium-titanate batteries are a type of lithium-ion battery with lithium-titanate nanocrystals on the surface of the anode, which otherwise would be coated with carbon [16]. The nanocrystals dramatically increase the surface area of the anode and allow for faster charging and discharging; in practice, this chemistry can allow for complete charging in 10 minutes (a 6C rating) [16]. In addition, this particular chemistry mix has a relatively long cycle life and calendar life, and lifetimes are relatively unaffected by fast-charging [16]. Lithium-titanate
batteries are rated for over 5,000 cycles [16]; some manufacturers even claim that the battery life will exceed 10,000 cycles [26]. Compare this to normal Lithium-ion batteries, which are rated for less than 1000 cycles [28]. As a very conservative estimate, this thesis assumes that the lithium-titanate batteries in the model will last for 10 years; note that the Budischak study assumed a 15 year lifespan [6]. This thesis assumes a round trip efficiency of 81% (90% each way) and a self-discharge rate of $8.33 \times 10^{-5}$ [29][30].

Pumped hydro was also considered as a storage technology for use in this model; in certain regions, pumped-storage hydroelectric power stations are used at a large scale to store energy when there is excess and release energy when there is a deficit. Pumped-storage hydroelectricity is implemented by building two connected reservoirs. Excess energy is stored as gravitational potential by pumping water into the higher-altitude reservoir; it is harvested when water is allowed to flow back down into the lower reservoir. Pumped hydro can also be used to store low-cost off-peak energy to be sold during peak hours. PJM in fact contains the largest pumped hydro facility in the U.S. [24]. The Bath County Pumped Storage Station has been operational since 1985 and is rated for 3,003 MW of power and has an estimated 23.1 GW of storage capacity [31] [32]. However, the geography of the region is not well-suited for scaling the amount of pumped-hydro. This is apparent when one considers the fact that electricity prices have gone into the negatives due to excess wind power during off-peak hours [24]. The Bath County Station is currently PJM’s only grid-level storage
facility, and the RTO is currently considering alternatives as the need for storage grows [24], including electro-chemical battery units.

2.4.3 BID STACK COST FUNCTION

The final cost parameter that our model requires is a cost function for on-demand fossil fuel generation, \( C^u(U_i) \). As noted in Chapter 2.1.5, we look at two different families of cost functions for this thesis. We primarily examine a range of different constant cost functions, \( C^u(U_i) = kU_i \). Additionally, we consider how the use of a bid stack, which is more realistic, affects total costs. The generation stack that is used comes directly from PJM and gives an estimate of the cost of generation for each marginal MW of power over the course of one hour, excluding startup costs. (PJM also makes bid data available to the public on a four-month lag [33].) For the purposes of this study, only the ‘fast’ generation technologies were included: gas turbines and internal combustion engines that do not use biofuel. Nuclear, coal, hydro, waste, and biofuels are excluded for simplicity. The assumption is that on-demand power can only be provided by sources with sufficiently high ramp-rates. In the figure below, note that there is a large increase in marginal cost at around 20 GW, presumably due to the switch from natural gas to oil. As noted before, the actual marginal costs are highly dependent on fuel prices; the bid stack used in this study is simply meant to be a representative sample.
3 OPTIMIZATION

This thesis examines two different sets of cost parameters and two separate input datasets for load and hourly capacity factors, amounting to four different optimizations that must be performed in order find the least cost solution for each scenario. In addition, one optimization is performed to understand how a bid stack cost function for fossil fuels changes the problem. Each of these optimizations finds the decision set that minimizes the total cost of electricity, including the cost of burning fossil fuels to meet unmet load. This is equivalent to minimizing the levelized cost of electricity over the time period studied, since total cost is levelized cost multiplied by total load. The optimization problem is framed as such:

$$\min_{\pi} \left\{ n(c_{pv}x_{pv} + c_{off}x_{off} + c_{on}x_{on} + c_{st}x_{st} + c_{sc}x_{sc}) + \sum_{t=t_0}^{T} C^U(U_t) \right\}$$

s.t. \( x \in \mathcal{X} \)

\( \mathcal{X} = \left\{ x^i \mid 0 \leq x^i \leq x^i_{\text{max}} \right\} \) for \( i = \{pv, \text{off}, \text{on}, \text{st}, \text{sc}\} \)
There are several different methods that may be used to solve this optimization problem. The study done by Budischak used an exhaustive grid search; for any given set of input cost parameters, 1.8 billion scenarios were simulated. Given that the Budischak dataset spanned 41,664 hours, the grid search amounted to simulating 75 trillion hours of energy generation. In order to complete this computationally intensive task, Budischak employed the use of 3,000 processors in parallel; even so, it took over 11 days to process the 18 combinations of cost parameters being studied.

This thesis uses a method called finite difference stochastic approximation (FDSA) to optimize over the domain of all feasible renewable technology build-outs. This algorithm is used to minimize loss functions when direct measurements of the gradient are not possible; thus the algorithm is ‘gradient-free’. The method relies on perturbing components of the decision variable one at a time to approximate the gradient of the loss function and has many of the same convergence properties as stochastic gradient optimization [34]. The FDSA algorithm is a local optimizer, thus it works best to find global minima when the objective function is convex.

Though the ability of the algorithm to find the global minimum does depend on the nature of the loss function, the advantage to using such a search algorithm is its speed and precision. For this thesis, we found that an FDSA algorithm can converge to the optimal solution within 500 iterations of the simulation model, whereas the number of simulations necessary for a grid search
would increase at a rate of the fifth power of the resolution of the search. For example, the Budischak study took the feasible range for each dimension of the decision variable and linearly sampled that range 70 times; thus $70^5$ combinations were simulated for each optimization. Because the computational requirements are greatly reduced when using a gradient descent algorithm, this thesis was able to perform optimizations for an entire range of fossil fuel costs in order to build a curve that shows the relationship between fossil fuel costs and optimal build-outs of renewable energy technology for use in the grid. Cost-optimized results can also be calculated for any arbitrary percent-coverage of renewables between 0% and 100%. For this thesis, we were able to solve 50 different optimization problems at a range of fossil fuel penalties for each of four different sets of input parameters. This would not have been possible with a grid search algorithm.

In addition to benefits relating to speed and efficiency, the gradient descent method provides improved resolution of the final solution. The optimal decision variable can be determined down to any number of significant digits, depending on how many iterations of the algorithm are run and what the parameters of the algorithm are set to. Meanwhile, the computational requirements for a grid-search increase at an undesirable rate for higher resolutions.

One may ask why a stochastic search algorithm is applied to our problem here, which is deterministic in nature. The theoretical basis for such a tactic is that using an FDSA algorithm allows us to treat the deterministic historical data
as a sample path of an underlying random distribution for hourly capacity factors and loads. The problem at hand involves using noisy measurements of the exogenous processes in order to simulate various cost, load, and weather scenarios. In real life, these exogenous processes are stochastic in nature; however, the use of historical data gives us only one sample realization. Framing the problem as a stochastic one allows us to translate our methods directly to a truly stochastic problem with little or no modification. This would allow, for example, randomly generated exogenous information to be used in the place of historical data. The underlying problem here is a newsvendor problem, in which we ask how much generation capacity to supply given uncertain demand. Though we use historical data in this thesis, it is a proxy for the much more relevant problem of understanding how to meet future demand for energy, which by nature is stochastic.

Note that Introduction to Stochastic Search and Optimization by James C. Spall was consulted heavily for this chapter [34].

3.1 FINITE-DIFFERENCE STOCHASTIC APPROXIMATION

We use a one-sided finite-difference algorithm with decreasing stepsizes to solve the optimization problem associated with each set of input parameters and historical data. As noted above, the algorithm is a variant of the stochastic gradient algorithm but only requires us to obtain noisy measurements of the loss
function, \( L(x) + \varepsilon(x) \). This recursive procedure determines successive iterations of the decision variable:

\[
\hat{x}_{k+1} = \hat{x}_k - a_k \hat{g}_k(\hat{x}_k) \quad \text{for } k = 1, 2, \ldots, K
\]

where \( \hat{g}_k(\hat{x}_k) \) is an estimate of the gradient of the loss function, \( \nabla F(x) \).

and \( a_k \) is the stepsize, also called the gain, or learning rate. For reasons related to convergence, we set \( a_k \) to a general harmonic sequence:

\[
a_k = \frac{\alpha}{\alpha + k}
\]

An estimate of the gradient is calculated by measuring the difference in the loss function when each dimension of the decision variable is perturbed by some small perturbation \( \delta > 0 \). In other words, the gradient estimate is a numerical directional derivative of the objective function. The gradient is calculated as a vector of five elements:

\[
\hat{g}_k(\hat{x}_k) = \begin{bmatrix}
\frac{F(\hat{x}_k + \delta \xi_1) - F(\hat{x}_k)}{\delta} \\
\vdots \\
\frac{F(\hat{x}_k + \delta \xi_5) - F(\hat{x}_k)}{\delta}
\end{bmatrix}
\]

where \( \xi_i \) denotes a vector of all zeros, but with a 1 in the \( i \)th place.

This process is repeated for \( K \) iterations. Note that for each iteration, the gradient is estimated using runs along the same sample path. Each iteration
requires six runs of the simulation model; once for the unperturbed decision variable and once for each of the five perturbed variants. In practice, we found that each iteration of the algorithm took approximately 1/3rd of a second, with no parallel processing implemented.

### 3.2 Convergence

The rate of convergence for FDSA methods is \( \frac{1}{k^{1/3}} \), where \( k \) is the number of iterations run. Though stochastic gradient algorithms converge at a faster rate of \( \frac{1}{\sqrt{k}} \), they also require direct measurement of the gradient, which is not possible for our problem. There are several conditions for the convergence of the FDSA algorithm, as stated by Spall and further supported by George and Powell [34][35]. The most relevant conditions are imposed on the choice of stepsize and the uniqueness of the minimum. (A condition is imposed on decreasing \( \delta \) as well; however, we choose to use a constant delta in order to maintain numerical stability, and the iterations still converge to the optimal solution within a certain margin of error.) The stepsize conditions are:

\[
\sum_{k=0}^{\infty} a_k = \infty \quad \text{and} \quad \lim_{k \to \infty} a_k = 0
\]

We use a stepsize function \( a_k = \frac{\alpha}{\alpha + k} \) that is a version of the generalized harmonic sequence, which is cited in Goerge and Powell as being convergent when the alpha parameter is tuned [35]. As for satisfying the conditions of
convergence, it is trivial to show that the sequence converges to zero, thus satisfying the second condition. With regard to the first condition, it is easy to show via comparison test that the generalized harmonic series diverges:

$$\frac{\alpha}{\alpha + k} \geq \alpha \left( \frac{1}{k + \lfloor \alpha \rfloor} \right)$$

We sum over all terms and the inequality holds:

$$\sum_{k=0}^{\infty} \frac{\alpha}{\alpha + k} \geq \sum_{k=0}^{\infty} \alpha \left( \frac{1}{k + \lfloor \alpha \rfloor} \right)$$

On the right hand side, we take the coefficient out of the summation and re-index the summation with \( j = k + \lfloor \alpha \rfloor \):

$$\sum_{k=0}^{\infty} \frac{\alpha}{\alpha + k} \geq \alpha \sum_{j=0}^{\infty} \frac{1}{j}$$

and

$$\sum_{k=0}^{\infty} \frac{\alpha}{\alpha + k} \geq \alpha \left[ \sum_{j=0}^{\infty} \frac{1}{j} - \sum_{j=0}^{\lfloor \alpha \rfloor} \frac{1}{j} \right]$$

The right hand side goes to infinity, since one of its terms is the harmonic series, and since the left hand side is greater than or equal to the RHS, it goes to infinity as well.

However, the most important condition for convergence to a global minimum is that there must exist one unique minimum. For our problem, this is equivalent to asserting that the objective function and domain are convex.
However, because of the nature of our problem, it is very difficult to show that
the loss function is convex throughout the entire domain. As Spall states,

The convergence conditions above provide an abstract ideal. In
practice one will rarely be able to check all of the conditions ... due
to a lack of knowledge about [the loss function]. In fact, the
conditions may not be verifiable for the very reason that one is
using the gradient-free FDSA algorithm!

Regardless, we will show that our problem’s loss function is convex, from
both a theoretical perspective and an empirical one.

3.2.1 OPTIMIZATION AS A LINEAR PROGRAMMING PROBLEM

One useful property of linear programs is that they are convex when
expressed as a minimization problem [36]. Below, we will show that our overall
optimization problem is in fact a linear program, which is a sufficient condition
for convexity. However, we present a proof only for the situation in which the
fossil fuel cost function is a constant for two reasons: first, our storage control
policy is optimal only for the constant fossil fuel penalty scenario; and second,
using a bid stack for the fossil fuel cost function causes the problem to be non-
linear; therefore the problem cannot be expressed as a linear program. Because of
the complexity of the problem, the number of decision variables in the linear
programming problem is equal to $5 + 4N$, where $N$ is the number of hours being
simulated. The linear programming formulation of the problem is as follows:
\[
\min \ c^T y
\]

where \( c = \begin{bmatrix} c^{pv} \\ \vdots \\ c^{sc} \\ k \\ \vdots \\ k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \) and \( y = \begin{bmatrix} x^{pv}_0 \\ \vdots \\ x^{sc}_0 \\ U_0 \\ \vdots \\ U_T \\ E^{+}_0 \\ \vdots \\ E^{+}_T \\ E^{-}_0 \\ \vdots \\ E^{-}_T \\ R_0 \\ \vdots \\ R_T \end{bmatrix} \)

Note that the vector of decision variables \( y \) includes variables not only for the decisions related to the technology mix \( x \), but also: \( U_i \) (the load that is not met by renewables during each hour), \( E_i^+ \) (energy into storage), \( E_i^- \) (energy out of storage), and \( R_i \) (the level of charge in the battery). The costs associated with each of the decision variables are \( c^{pv} \) \( \ldots \) \( c^{sc} \) for \( x \); \( k \) (the constant fossil fuel penalty) for \( U_i \); and zero for all other decision variables, since it costs nothing to operate the battery. We continue the linear programming problem:
\[
\min c^T y
\]
such that:
\[
W_t^{pv} x^{pv} + W_t^{off} x^{off} + W_t^{on} x^{on} + U_t + E_t^- - E_t^+ \geq L_t \quad \text{for } t_0 \leq t \leq T
\]

This condition specifies that the energy generated plus fossil fuel energy used plus energy out of storage minus energy into storage must be greater than or equal to the load during that hour, for all hours. Next, we specify the behavior of the battery:

\[
x_t^- - E_t^- \geq 0 \quad \text{for } t_0 \leq t \leq T
\]

This ensures that the energy coming out of the battery is limited by the battery inverter power rating.

\[
R_t - E_t^- \geq 0 \quad \text{for } t_0 \leq t \leq T
\]

This ensures that the energy coming out of the battery is limited by the level of charge of the battery.

\[
x_t^+ - E_t^+ \geq 0 \quad \text{for } t_0 \leq t \leq T
\]

This specifies that the energy coming into the battery is also limited by the battery inverter power rating.

\[
x_t^+ - R_t - E_t^+ \geq 0 \quad \text{for } t_0 \leq t \leq T
\]

This specifies that the energy going into the battery at any given hour is no greater than the amount of available battery storage capacity.
\[ R_0 = 0 \]

\[ R_t (1 - \lambda) - R_{t+1} - E_t^- + E_t^+ \geq 0 \quad \text{for} \quad t_0 + 1 \leq t \leq T \]

These two conditions specify that the battery starts off uncharged and that the battery self-discharges at a rate \( \lambda \). In addition, the next hour’s battery charge must be equal to this hour’s battery charge minus EOS plus EIS.

There are three notable conditions that are not specified in the linear programming problem. First, there is no condition that states that if excess energy is generated, all of it should be stored in the battery. Second, there is no condition that energy generated during an hour must be used toward satisfying load instead of being stored in the battery. Lastly, there is no condition that energy in storage must be used immediately to satisfy load (instead of being stored for later hours). The reason for the apparent under-specification of the problem is that these strategies are, by nature, optimal (as proven in Chapter 2.2), and solving the linear programming problem will automatically select for these strategies.

Thus we have successfully transformed our model into a linear programming problem. The problem as is outlined above is guaranteed to be translatable into the canonical form \( \min \left\{ c^T y \mid Ay \geq b, y \geq 0 \right\} \), since the objective function is a linear function of the decisions variables, and each condition is a
linear inequality. Because the problem can be expressed as a linear programming problem, we also know that it is convex.

3.2.2 **EMPIRICAL CONVEXITY OF THE LOSS FUNCTION**

Below, we use several heat map plots to understand the ‘shape’ of the loss function by measuring its value at lattice points within a two-dimensional plane. Heat maps are only able to show three dimensions, (the z-axis is represented by color), so the plots are effectively ‘cross-sections’ of the 6-dimensional space of the objective function. Nonetheless, they allow us to visualize how the loss function changes in response to variations in the decisions variable. All heat maps below use 2008 costs; the first five examine the Budischak dataset and the last one looks at the historical generation dataset, which we can expect to have a somewhat differently shaped loss function.

Our first heat map explores battery inverter power and battery storage capacity at a fossil fuel cost of $1000/MWh for the Budischak dataset. Solar and offshore wind were set to zero and onshore wind was at its maximum value; this configuration of solar and wind happens to be the optimal solution for a wide range of fossil fuel costs, from approximately $1000/MWh to $1750/MWh. From the heat map of LCOE on the right, we note that with respect to these two dimensions, the loss function is clearly convex. The local minimum here occurs at an interior point in the plane and no alternative local minima seem to exist within the range. Intuitively, this makes sense. Using small amounts of battery storage and inverter power allows excess generation to be used later, which decreases the
load that must be serviced by expensive fossil fuel generation; however, at a
certain point, the marginal cost of increasing the battery capacity and inverter
power rating becomes greater than the marginal benefit of displacing fossil fuel
generation. Note that the heat map of percent coverage shows that increasing
battery capacity without the necessary inverter power does very little to increase
the supply of renewable power, and vice versa. However, when both dimensions
increase in tandem, the percentage of hours covered by renewables increases
significantly.

**Figure 3-1: Heat maps of battery inverter power and battery capacity**

The second heat map shows the relationship between solar power,
offshore wind, and the objective function. For this heat map, the fossil fuel cost
was set to $2500/MWh, and onshore wind was set to its maximum value; the
battery capacity was set to 200 GWh, and the inverter was set to 25GW of
power. One can tell that the loss function is convex surrounding a minimum at
approximately 6 GW of offshore wind and 0 MW of solar. These two plots also
show that solar power is not cost-efficient compared to offshore wind. It seems that both variables linearly increase percent coverage in this localized part of the domain; however, the high price of photovoltaic arrays does not justify the power that it provides.

![Solar & offshore wind vs percent coverage](image1)

**Figure 3-2: Heat maps of solar and offshore wind**

Our third heat map looks at onshore and offshore wind and is accompanied by a contour plot to better illustrate the shape of the loss function. For this heat map, the fossil fuel penalty was set to $2500/MWh, solar was set to zero, and the battery size and inverter power were set to the same values as in the previous plot. As with the previous heat map, we can tell here (through the contour plot) that a minimum occurs in the lower right-hand corner, at a point where onshore wind is maximized and offshore wind takes a small value. It is worth noting that onshore wind here is the clear winner in terms of cost effectiveness. However, in the area where both offshore and onshore wind are

---

[Image of Solar & offshore wind vs percent coverage]

[Image of Solar & offshore wind vs LCOE]
close to zero, the use of a very small amount of wind generation decreases electricity costs dramatically, regardless of type of wind farm.

The fourth heat map for the Budischak dataset examines the relationship between LCOE, wind generation, and storage. The fossil fuel cost is set to $750/MWh, and 50 GW of battery inverter power is assumed to be available. This heat map does not assume that there is any solar power available. At first glance, the heat map seems to indicate that battery storage almost has no effect on cost. Cost seems to be entirely dependent on whether or not onshore wind is supplied. However, the contour plot shows that the loss function minimum occurs at an interior point, and in fact, there is some variation in the loss function with regard to storage capacity. It is clear that for low levels of onshore wind generation, storage does not matter. However, once onshore wind passes the 80 GW mark, it seems that a small amount of storage capacity can help decrease

\[ \text{Figure 3-3: Offshore and onshore wind vs LCOE} \]
costs. As with the results of the Budischak paper, maximizing onshore wind seems to be a dominant strategy for reducing LCOE.

The fifth and final heat map for the Budischak dataset shows that when onshore wind is already maximized, increasing storage capacity decreases cost more so than diversifying generation by adding offshore wind. This heat map was generated with a $2500/MWh fossil fuel cost, no solar power, and the maximum amount of onshore wind. From the contour plot, we can tell that a minimum of the objective function occurs at around 6 GW of offshore wind and 200 GW of storage, which is consistent with the previous heat maps. However, note that the variation in the loss function with regard to offshore wind is negligible compared to the effect that storage has.

**Figure 3-4: Onshore wind and battery capacity vs LCOE**
The last heat map plots gives us a peek at the shape of the objective function when the historical generation dataset is used instead of the Budischak dataset. The heat maps show how solar power build-out and storage capacity change the percent coverage and cost outcomes. Here, the fossil fuel cost was set to $1100/MWh, onshore wind was set to its maximum value, and battery inverter power was 25 GW. We note that the convexity of the objective function is supported by the presence of only one minimum, which is at an interior point. In addition, it seems that unlike for the Budischak dataset, LCOE is most sensitive to solar power build-out, probably because of the second dataset’s increased load and inability to access offshore wind.
Several parameters must be set in order to carry out the optimization effectively and in a way that makes sense for the problem at hand. In Chapter 4, we will discuss how certain parameters for the FDSA algorithm were determined. However, for the overall optimization problem, we must also determine what the upper bounds are for each technology. In particular, we ask what limits to impose on the build-outs of solar power, offshore wind, onshore wind, battery storage, and inverter power. In order to simplify the problem, we specify the domain of the decision variable $x$ to simply be bounded by zero and a set of five upper bounds, which were taken from the Budischak study [6].
4. ALGORITHM TESTING

The success of the FDSA algorithm is dependent on not only the convexity of the objective function but also the choice of various parameters involved in the algorithm, namely: the stepsizes \( a_k \) by which the gradient estimate is scaled, the perturbation \( \delta \) that is used to calculate the numerical directional derivatives, and the initial vector of decision variables, \( x_0 \).

4.1 STEPSIZES: ALPHA

Although the generic form \( a_k = \frac{\alpha}{\alpha + k} \) was used as a ‘gain’ coefficient for each iteration of the FDSA algorithm, the \( \alpha \) parameter that was used for each run differed. It turned out that the ability of the algorithm to find the least cost solution depended heavily on the choice of alpha. The ideal alpha value for a given run depended on which dataset was used, which cost parameters were used, and what the fossil fuel penalty was. Thus, in practice, the FDSA algorithm was run multiple times at various alphas, and the best solution was taken. In

<table>
<thead>
<tr>
<th>Decision variable</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar (MW)</td>
<td>0</td>
<td>186,000</td>
</tr>
<tr>
<td>Offshore Wind (MW)</td>
<td>0</td>
<td>157,809</td>
</tr>
<tr>
<td>Onshore Wind (MW)</td>
<td>0</td>
<td>131,834</td>
</tr>
<tr>
<td>Battery Inverter (MW)</td>
<td>0</td>
<td>( \infty )</td>
</tr>
<tr>
<td>Battery storage (MWh)</td>
<td>0</td>
<td>( \infty )</td>
</tr>
</tbody>
</table>

*Table 3-1: Decision variable limits*
order to guess which value of alpha to start at for a given dataset and cost parameter set, we chose a representative fossil fuel cost and ran the algorithm with a gamut of alpha values. Below is an example of one of these tests, where a range of alpha values from 0.05 to 50 were tested. A total of 500 iterations were run on the Budischak dataset at 2008 costs, with delta = 500 and a constant fossil fuel penalty of 1750 $/MWh.

*Figure 4-1: Effect of alpha value on FDSA results*

*Figure 4-2: FDSA runs for alpha = 1, 5*
The plots above show how the results of the FDSA algorithm varied in relation to the choice of alpha value. Looking at the graph of final LCOE vs. alpha, it is clear that alpha values of 2.5 and 5 allow the algorithm to converge for this specific set of input parameters. The lack of convergence for other parameters is evident when looking at how the objective function changes over the course of the FDSA run for alpha = 1. We notice that in the very beginning of the run, the objective function spikes up and for alpha = 1, the function never ‘bottoms out’ to a final value. Meanwhile, the graph of the objective function for alpha = 5 shows that the objective function peaks then falls rapidly to a final minimum value, which is an ideal behavior for the algorithm. The graph below shows how the LCOE of the iterates vary over the course of the algorithm when alpha = 25, and it is clear that for too-high values of alpha, the algorithm also does not converge.

![Figure 4-3: FDSA behavior for alpha = 25](image)
4.2 PERTURBATION SIZE: DELTA

The behavior of the FDSA algorithm is also dependent on the choice of the perturbation $\delta$, which affects the calculation of the numerical partial derivatives that determine the path of steepest descent. To understand the effect of various choices of delta on the results of the algorithm, we executed many runs with a range of different deltas and then plotted the solutions found by each run. The other parameters were held constant: the runs were done on the Budischak dataset at 2008 costs, with a fossil fuel energy cost of $1750/MWh, with 500 iterations, $\alpha = 2.5$, and the same starting decision $x_0$. In the figure below, we graph delta on the x-axis. On the left-hand graph, the y-axis represents the percentage of hours covered entirely by renewables in the solution found by the FDSA algorithm. On the right-hand graph, the y-axis represents the levelized cost of electricity of the solution.

**Figure 4-4: Effect of perturbation delta on FDSA results**
It seems that up to a point, smaller perturbations allow the algorithm to detect solutions at a better ‘resolution’. This makes sense if one considers a similar optimization problem with a two-dimensional decision variable. The loss function is then a convex 3D surface, perhaps a paraboloid. When iterating from a decision variable that is close to, but not quite at the global minimum, the gradient estimate will indicate that the function increases in every direction if the perturbation is too large. With a smaller perturbation, the gradient estimate will be more precise, and will lead the next iteration of the decision variable closer to the true minimum. Note that in the graph above, the LCOE actually increases with delta less than 500. According to Spall’s *Introduction to Stochastic Search and Optimization*, a delta converging to zero is desirable, which is contradictory to what is found here. I can only speculate that this is because the loss function employed in this thesis is very noisy due to the use of historical data, and a too-small delta will be ‘caught’ in the noise and find local minima close to the global minimum.

From the graph above, it is clear that the error associated with choosing a too-large delta is minute; the difference between the minimum and the maximum LCOE values on the graph above is about 0.4 cents. For this thesis, a delta of $\delta = 500$ was chosen, since the minimum LCOE was achieved at that value of delta in the graph above and in other similar tests.
4.3 INITIAL ESTIMATES

To test the FDFA algorithm’s sensitivity to the initiation estimate of the decision variable $x_0$, we conducted trials in which the initial estimate was randomly selected but all other parameters were kept the same. The result showed that depending on the input parameters, the algorithm could converge to multiple locations or only one. Two sets of trials were conducted, one at a fossil fuel cost of $500/MWh and another at a fossil fuel cost of $1750/MWh. Both used the Budischak dataset at 2008 costs; with $\delta = 500$ and alpha tuned to the cost parameters. Initial estimates for each dimension of the decision variable were chosen using a uniform random distribution over the entire domain of $x$, except for battery inverter power and battery storage capacity, whose domains are unbounded. The results are below:

![Graphs showing percent coverage and LCOE over 50 random trials.]

*Figure 4.5: Random initial estimates at fossil fuel cost of $500/MWh*
It is clear that the first test converged to the same optimal solution for every trial out of the 50, whereas the second test, conducted with a fossil fuel cost of $1750/MWh, converged to one of two different solutions. Out of the 50 trials performed for that test, 17 converged to the true solution, whereas 33 converged to another point. The true solution was:

\[ \hat{x} = \left\{ 0 \ 0 \ 131834.17 \ 20392.27 \ 152469.11 \right\} \]

And the false solution was:

\[ x' = \left\{ 0 \ 8297.51 \ 131834.17 \ 0 \ 0 \right\} \]

It is evident that the two points of convergence represent very different ‘strategies’ for reducing cost; the true solution maxes out onshore wind and uses a battery to make excess generated energy available at other times, whereas the false solution opts for offshore wind generation to supplement the onshore wind.
When looking at the initial estimates from the two different outcome groups, we find that there is no clear pattern that predicts which point an initial estimate will converge to. This leads us to consider whether or not the noisiness of the historical data (which can also be thought of as a sample path of a random time series) is causing the FDSA algorithm to find local minima around the true optimal solution. For this reason, the final optimizations that were executed for this thesis use various initial estimates for the same input parameters, in order to ensure that one of them will find the true least cost solution.

*Figure 4-7: Initial decisions that converge to the true solution*
5 RESULTS

The results from the simulation model and FDSA algorithm allow us to understand the relationships between various cost parameters, time periods, fossil fuel penalties, and energy outcomes. Below, we examine the output from a single simulation without optimization and the output from a single run of the FDSA algorithm, in order to get a sense of how the model behaves when processing real data. The core of this thesis’s results is a collection of curves showing the relationship between fossil fuel penalties and energy outcomes. We cover four different cost and time-frame scenarios, and we examine results from two different types of fossil fuel cost functions.
5.1 SIMULATION MODEL OUTPUT

We start by examining the output of the simulator without optimization in order to gain an understanding of how battery levels, generation, and fossil fuel backup change over the course of a simulation. In the two sections that follow, we look at simulations run over our two different datasets. In each, we use the optimal technology mix for the given scenario. For both simulations, fossil fuel costs are a constant $2500/MWh and all other costs parameters are taken from 2008.

5.1.1 BUDISCHAK DATASET

When using the Budischak dataset, the FDSA algorithm outputs the following optimal mix of technologies: 0 MW solar, 6,076 MW offshore wind, 131,834 MW onshore wind, 23,591 MW battery inverter power, and 192,671 MWh of storage capacity. For this technology mix, 91.2% of simulated hours were fully covered by renewable energy sources and storage. The overall LCOE was 32.41 cents, which is high compared to current-day prices, but this is understandable given the high cost of fossil fuel energy in this simulation.
The average generated power over the simulated four years is 56 GW, which is almost 1.8 times the average load of 31.5 GW. Note that the power generation follows yearly cycles and seems to be a stationary time series; however, at smaller time scales, the level of generated power is extremely unpredictable.

Figure 5-2: Energy in storage, Budischak dataset
The amount of energy in storage is extremely volatile over the course of the simulation. It is clear that during periods of high power output from offshore and onshore wind turbines, the battery was able to maintain a state of charge close to 100%; however, during parts of the year when hourly capacity factors for wind were low, the battery storage levels fluctuated wildly between empty and full. The figure for fossil fuel generation below matches the battery storage and generation plots above; times of low generation and low battery storage correlated with increased reliance on ‘backup’ fossil fuel energy.

![Figure 5.3: Fossil fuel generation, Budischak dataset](image)

### 5.1.2 Historical Generation Dataset

When using the historical generation dataset, the FDSA algorithm outputs the following optimal mix of technologies: 186,000 MW solar, 131,834 MW onshore wind, 77,688 MW battery inverter power, and 556,613 MWh of storage capacity. Note that solar and onshore wind build-outs are at their upper
bounds. When this technology mix scenario is run through the simulator, we find that the percent coverage is much lower than for the Budischak dataset simulation that covered 1999 to 2000; meanwhile the cost is significantly higher. For this technology mix and dataset, 45.1% of simulated hours were fully covered by renewable energy sources and battery. The overall LCOE was $1.04. This is almost entirely due to the fact that the PJM grid has expanded greatly between 2002 and 2012. With current-day loads, the limited availability of space for solar panels and onshore wind farms means that renewables cannot cover the majority of load. Thus, pricey fossil fuel backup technologies must be used.

![Power generation, historical generation dataset](image)

*Figure 5-4: Power generation, historical generation dataset*

The average generation over the duration of the simulation is 68.3 GW, whereas the average load is 88.5 GW, which means that average generation is approximately 77% of the average load. Because the historical generation dataset covers a time period that is slightly shorter than a year, there is no cyclicality in the plot of generated power.
The graph of energy in storage is further evidence of the fact that renewable energy sources were not enough to meet demand for power. The battery seems to be discharged more often than it is charged, and the average state of charge of the battery is 21%. The use of fossil fuel backup is also visibly correlated with generation and energy in storage; when generation is high, between the 5,000\textsuperscript{th} and 6,000\textsuperscript{th} hour, there is a marked decrease in fossil fuel reliance. During other parts of the year, there seems to be a continued, intermittent use of fossil fuel energy, sometimes reaching as high as 140 GW. It is worth noting that load during the year spikes near the end of the period, during the late summer, which correlates with the point in the year at which the most fossil fuel generation was utilized.

\textbf{Figure 5-5: Energy in storage, historical generation dataset}
5.2 **FDFA ALGORITHM PERFORMANCE**

The FDFA algorithm was successful in finding the approximate least cost solution for most inputs. The results for one example run of the algorithm are shown below. Five hundred iterations of the algorithm were performed for the Budischak dataset with 2008 costs, at a fossil fuel penalty of $2500/MWh. The algorithm parameters were as follows: $\alpha = 5$, $\delta = 500$, and the initial decision variable estimate was $x_0 = \{0, \ 80000, \ 76000, \ 29000, \ 145000\}$.

After 500 iterations, the lowest cost configuration was:

$$x_{500} = \{x_{500}^{PV}, \ x_{500}^{off}, \ x_{500}^{con}, \ x_{500}^{f}, \ x_{500}^{sc}\}$$

such that:

$$x_{500}^{PV} = 0$$

$$x_{500}^{off} = 6075.95$$
\[x_{500}^{or} = 131834.17\]
\[x_{500}^{st} = 23590.99\]
\[x_{500}^{sc} = 192671.45\]

The LCOE and percent of hours covered entirely by renewables were:

\[C_{LCOE} = 0.3241\]

Percent coverage = 91.18%

Note that this solution satisfies the constraint that each dimension of the solution vector must not exceed the upper limit for that technology. To understand the path of the optimization, we plot the cost and percent coverage over the course of the 500 iterations; in addition, we plot the evolution of each dimension of the decision variable over the 500 iterations. Note that the percent coverage fluctuates wildly with decreasing amplitude for the first hundred iterations, due to the large stepsize coefficient applied at the beginning of the algorithm. As the amplitude decreases, the running average increases to a value close to its final value. After iteration 150, the algorithm makes a few small adjustments before settling down to its final value. The LCOE graph exhibits a large spike at the beginning of the algorithm (an over-correction due to large gain coefficient); then it gradually drops until it bottoms out at around iteration 350.
Figure 5-7: Percent coverage and LCOE over a single run of FDSA
Note that similarly to the path of percent coverage and cost, the decision variables fluctuate wildly for the first 100+ iterations. Then their paths start to converge more slowly over the course of the remaining 400 iterations.
power hits its optimal value, zero, at around iteration 125. Offshore wind oscillates with decreasing amplitude to gradually converge near iteration 300. Onshore wind hits its optimal value at around iteration 125 and stays there. Battery inverter size spikes up and gradually decreases over the course of the algorithm. Battery storage capacity shows the most erratic behavior; the path oscillates with decreasing amplitude, then increases unsteadily, the plateaus, then decreases again before reaching a final value. Though it seems that the battery storage capacity may not have converged, when the algorithm was run for another 500 iterations, the path did not stray from that point.

5.3 CONSTANT FOSSIL FUEL ENERGY COST

The main goal of this thesis is to understand how fossil fuel backup costs affect cost of electricity and renewable energy coverage when LCOE is optimized. Curves showing these relationships were drawn for four different scenarios. The first scenario simulates four years starting at 1999 and uses 2008 cost parameters; the second scenario simulates one year starting in 2012, with 2008 costs; the third scenario simulates four years starting in 1999, with 2030 costs; and the last scenario simulates one year starting in 2012, with 2030 costs. The results for all four scenarios are shown in the figure below.
Figure 5-9: Fossil fuel cost vs % coverage and LCOE for various scenarios

We observe in these two graphs that as the fossil fuel cost increases, it becomes more favorable to rely on renewable energy sources and batteries. It also seems that the first derivative of the LCOE function with respect to fossil fuel cost is non-increasing, meaning that the LCOE curve is concave. Notice in these graphs that optimization with 2030 costs always produces lower costs and higher rates of renewable energy coverage; this is because the cost of capital for every technology is projected to be significantly cheaper for the year 2030. However, we also notice that for fossil fuel penalties less than $500/MWh, the difference between 2008 and 2030 LCOEs is almost negligible for the historical dataset and is very small for the Budischak dataset. In addition, when the optimization is done over historical generation data, the resulting cost of electricity is two to three times higher than for the Budischak dataset. This is mostly due to the fact that load is significantly higher in the 2012 to 2013 time period, and there is a
limited supply of renewable energy sources to meet load. Another important factor is that offshore wind was not available as a source of generation for the historical dataset, thus solar power had to be used instead. However, notice that even for low fossil fuel costs ($50 - $400/MWh), for which no solar power is used in either dataset, the historical generation dataset resulted in much higher electricity costs. This is a combination of the greater load and the fact that the historical generation dataset provided lower hourly capacity factors. The result is that for the historical generation dataset, onshore wind was more readily maxed out, solar power was utilized heavily, and battery storage and power were built out to high levels.

We also note that the percent coverage curve exhibits a very strange shape for the historical generation dataset, regardless of which cost parameters are used. The curve increases rapidly at first, then hits an inflection point and levels out, then increases again until hitting a second inflection point, then increases at a much lower rate. Over the range of fossil fuel costs, the curve is concave, then convex, then concave again. Looking at the figure below showing the optimal decision variables over a range of fossil fuel costs, we notice that the first inflection point coincides with the point at which onshore wind is maxed out. The point at which the percent coverage increases again coincides with the point at which the algorithm chooses to start using solar power to supplement the onshore wind generation. As the fossil fuel continues to increase, battery capacity is added to the mix. The second inflection point occurs when photovoltaic
generation also hits its upper bound, and all gains in percent coverage must come from increased battery size.

![Graphs of fossil fuel cost vs renewable energy sources and inverter power](image)

Figure 5-10: Fossil fuel cost vs decision variables for various scenarios
It is unclear whether this strange coverage curve is due to a failure on the part of the optimization algorithm to find the true solution, or whether the deviation is due to the fact that the period of simulation is very short for the historical dataset (<1 year), which can result in over-fitting. In other words, because the time period is so short, the FDSA algorithm may be finding the optimal solution that specifically fits the vagaries and quirks of that one time period, yet its solution would fail over much longer time periods. Or perhaps the queer curves are simply the result of the way in which the upper bounds of the decision variable were set.

Below, we present the individual results for each of the two datasets and each of the two different cost parameter sets.

5.3.1 BUDISCHAK DATASET

The results for the Budischak dataset are generally very well behaved. For the results associated with the 2008 cost parameters as represented by the figures below, we notice that no solar power is employed, presumably because of high cost. In fact, only modest levels of offshore wind (about 6 GW) are utilized even when the fossil fuel penalty is as high as $2500/MWh. Presumably, there exists a fossil fuel penalty greater than $2500/MWh that is high enough to warrant a need for solar power. As with all the other scenarios being tested, the Budischak-2008 scenario uses onshore wind power even at low fossil fuel energy costs. However, when the fossil fuel cost is $50/MWh, which is the average price
of wholesale power today (in 2014), the optimal solution is to not build out any renewable energy generation capacity.

Additionally, we find that battery capacity becomes useful only when onshore wind is almost maxed out, most likely due its high cost. Like the results obtained by Budischak, we find that over-generation with some energy ‘spilled’ is cheaper than building battery storage infrastructure to store and release excess generated energy. Also, note that offshore wind does not become cost-effective until the fossil fuel penalty rises to almost $2000/MWh. This is long after the onshore wind capacity has maxed out, showing that with current prices of offshore wind, it is more effective to use storage in conjunction with onshore wind.

We also notice that there is a small peak in the battery variables at around $700/MWh. This is most likely due to the noisiness of the data or imprecise convergence of the FDSA algorithm.

Figure 5-11: Results for Budischak dataset, 2008 costs
Figure 5-12: Decision variables for Budischak dataset, 2008 costs
The results for the 2030 cost parameters show that with significantly cheaper capital costs, renewable energy becomes cost-effective at lower fossil fuel penalties. However, the main takeaways are remarkably similar to those for 2008 costs. The optimal solution at the current wholesale price of electricity is to not develop renewables at all. Onshore wind is still most cost-effective; storage is cheaper than offshore wind; and solar is only feasible when fossil fuel penalties are extremely high. Even at $2500/MWh for fossil fuel energy, we notice that it is optimal to only provide 0.6 GW of solar power.

Note that there are small bumps in the curves for each of the decision variables, which may be due to noisiness of the data or possibly imprecise convergence.

Figure 5.13: Results for Budischak dataset, 2030 costs
Figure 5-14: Decision variables for Budischak dataset, 2030 costs
5.3.2 HISTORICAL GENERATION DATASET

The results from the historical generation dataset differed greatly from those derived from the Budischak dataset. For reasons explained in the beginning of Chapter 5.3, costs were higher and percent coverage was lower for this dataset, which was derived from power output data obtained from up-and-running wind farms and solar arrays. It is notable that for this dataset, solar power seems to be more cost-effective than buying storage. This also most likely because of the much higher average load within this dataset. The percent coverage by renewables is very low for fossil fuel penalties under $1000/MWh, so any contribution from a renewable energy source is going to be cost-effective. Results from both datasets show that storage only becomes effective when there is already a large amount of generation through wind and solar, relative to load.

Another notable result is that for current whole prices of fossil fuel energy, the least cost solution, again, is to not build out renewables. Even when the price is double the current wholesale price, building out renewable generation is not the least-cost solution, when using the historical generation dataset.
Figure 5-15: Results for historical generation dataset, 2008 costs
Figure 5-16: Decision variables for historical generation dataset, 2008 costs
The results from the second dataset with 2030 costs are almost identical to those resulting from 2008 cost parameters. The main difference is that each decision variable curve is shifted left (to lower fossil fuel energy costs) by about $500, except for onshore wind, which exhibits almost the exact same curve as for 2008 costs. Again, fossil fuel energy prices at $50 - $100/MWh result in an optimal solution that relies only on fossil fuel generation.
Figure 5-18: Decision variables for historical generation dataset, 2030 costs
5.4 BID STACK

The inclusion of a bid stack cost function helps us answer the following question: what would be the optimal technology mix if all non-renewable generation were supplied through the current ‘fast’ generation stack? Recall that the current bid stack for on-demand energy is supplied through gas turbines and oil-powered internal combustion. In considering these results, we must keep in mind that we cannot prove that the objective function with bid stack integration is convex, and we know that the myopic policy for battery control is suboptimal. Thus we know that the FDSA algorithm probably does not converge to the true least cost solution. Nonetheless, the results below serve as an upper bound for the true least cost solutions. We immediately notice that the optimal percent coverage is relatively low, and the optimal LCOE is also relatively low, though it is still higher than the 5-cent per kWh current wholesale cost of electricity. Of particular note is that in this case, the difference between 2008 and 2030 results is minimal; the LCOEs differ by about a cent. However, the Budischak percent coverage jumps by five percentage points when using 2030 costs. It is interesting to note that the bid stack results for the Budischak dataset mirror the constant fossil fuel penalty results for a penalty of about $100/MWh, whereas for the historical data set, the bid stack results correlate to a fossil fuel cost of about $300 - $400/MWh. It is apparent that the increasing nature of the bid stack cost function hits the high-load historical generation dataset the hardest.
Figure 5-19: Bid stack percent coverage and LCOE

<table>
<thead>
<tr>
<th></th>
<th>Percent coverage</th>
<th>LCOE ($/kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budischak dataset /2008</td>
<td>15.30%</td>
<td>$0.10</td>
</tr>
<tr>
<td>Historical dataset /2008</td>
<td>6.15%</td>
<td>$0.25</td>
</tr>
<tr>
<td>Budischak dataset /2030</td>
<td>20.90%</td>
<td>$0.09</td>
</tr>
<tr>
<td>Historical dataset /2030</td>
<td>7.05%</td>
<td>$0.24</td>
</tr>
</tbody>
</table>

Table 5-1: Bid stack percent coverage and LCOE
The optimal decision variables are of little surprise. As with the percent coverage and LCOE results, they mirror the results derived from a constant fossil fuel cost function. Onshore wind is the most cost-effective, and for 2030 costs, the historical dataset opts to include a small amount of solar power.

6 DISCUSSION

The results of this thesis stand in stark contrast to some of the more optimistic studies done recently concerning high-penetrations renewable energy generation. In a world where energy security is of paramount importance, it is critical to understand how we might realistically move forward to create a more sustainable energy infrastructure for mankind. The question of cost is one of the most important factors when looking to understand how a shift to greener energy sources might unfold. This thesis approached the problem with the assumption that market forces drive the development of renewable energy infrastructure, thus the use of a fossil fuel penalty fee for each MWh of energy generated with non-renewable sources. This fossil fuel penalty can be thought of as representing many different things. It could represent the cost of externalities associated with burning fossil fuels; it could represent a carbon tax or cap-and-trade policy; it could represent the actual cost of fossil fuels if baseload generation were to go offline. Regardless, the inclusion of a fossil fuel energy cost in the model is a useful tool for us to understand how we as a society might respond to an increased penalty on combusting fossil fuels as an energy source.
Among the many insights that are derived from the data, a few trends stand out. We notice that solar power is the least cost-efficient technology out of the three methods of renewable generation, followed by offshore wind. Onshore wind is the most cost-efficient, and when coupled with storage, it has the ability to cover a significant percentage of load at reasonable costs for 1999 load data. Another important result is that load in 2013 has grown to be so large that solar power and onshore wind farms alone are not able to cover even half of the simulated hours. In addition, it seems that given the cost parameters, large-scale batteries are not cost-effective unless a significant percentage of hours are covered by renewables; instead, for low coverage percentages, it seems to be more cost-effective to purchase additional capacity to generate power.

Additionally, the results of this thesis shed light on the Budischak study in many ways. For example, the Budischak study uses a fossil fuel energy cost of $80/MWh for any load not met by renewables. It turns out that when fossil fuel is available at this price, the optimal strategy for minimizing costs is to buy no renewable technologies, thus allowing energy to be bought at $0.08/kWh, which is lower than any of the LCOEs that the Budischak model produces with arbitrary requirements for the percentage of hours that must be entirely met by renewables. In many ways, this thesis also corroborated the results from the Budischak study. Both of the studies found that onshore wind alone was able to provide a lower percentage of hours covered by renewables (up to 30%). At 90% coverage, both the Budischak results and this thesis’s results found that offshore wind, onshore wind, and some battery storage were necessary; however, the high
cost of solar power made it prohibitive unless a high fossil fuel penalty or arbitrary requirement for percent coverage was imposed.

The policy implications are many. We find that at today’s wholesale price of electricity, the least-cost solution is to not build out any renewable energy sources, for both 2008 and 2030 costs. Thus it is clear that structurally, we need to make changes to how fossil fuel energy costs are represented in their price. If the externalities of fossil fuel generation are included in the price of electricity itself, there is a greater incentive for the market to pursue more sustainable forms of energy. Furthermore, there is much work that needs to be done to decrease the cost of renewable energy sources. More basic research must be conducted to increase the cost-effectiveness of solar PV energy in particular, especially because solar serves as a useful complement to onshore wind due to its negative correlation with onshore wind generation and positive correlation to load. Finally, as we start to map out the next ten years of energy projects, we now know that among solar, onshore wind, offshore wind, and battery infrastructure, onshore wind provides the most cost-efficient method of increasing percent covered by renewables.

6.1 OPPORTUNITIES FOR FURTHER RESEARCH

Though this thesis is a promising start in a long journey to understanding renewable energy economics, there remain many opportunities to build upon and improve the work that was done here, both in execution and in scope.
In terms of execution, the results of this thesis could be made more meaningful through the incorporation of a larger set of historical power output data. The lack of a reliable source of data for offshore wind generation made it difficult to compare results between the Budischak dataset and the historical generation dataset. Furthermore, the short time span of the second dataset cast some doubt over the applicability of its results. One challenge in using data from a long period of time is that PJM as an organization has grown considerably in geographical scope; however, PJM records each of its constituent areas’ load data separately, so it is easy to isolate select areas from which to extract load data.

Another opportunity for further work is to increase the complexity of the model. If we want to truly understand how grid-level batteries will fare within the PJM system, a better model would need to be created that simulates how batteries age and how charge and discharge rates are affected by state of charge, and battery degradation. However, a more complex model may require either more powerful computational resources or a better optimization algorithm. Optimization is an area of exploration in and of itself. Is FDSA the best search algorithm for this problem?

Though this thesis also broaches the possibility of incorporating bid stacks into the model, the implementation was somewhat incomplete. An optimal policy for on-line battery control should be devised for when fossil fuel energy costs use a bid stack function. Instead of exploring different constant fossil fuel energy costs, one could also study optimal results simulated using shifted or
scaled bid stacks to account for the cost of fossil fuel generation. Another element that would improve the model is to consider the use of forecasts for lower coverage percentages. In general, forecasting could be used to dramatically lower costs, since cheaper existing baseload power can be used.

Finally, this thesis constrained the number of renewable energy sources that were considered, and as a result, options like hydroelectric power, geothermal power, solar thermal, and hydropower have yet to be explored. In addition, only one storage option was selected; certainly the results would differ if, for example, pumped hydro were the storage technology. Perhaps a model can be created in which multiple tiers of storage are available, which would be more realistic than just one centralized battery. Thankfully, in terms of changing cost parameters, the modularity of the computational model ensures that anybody can input an Excel sheet with alternate cost parameters to test various hypotheticals.

7 CONCLUSION

Two of the greatest challenges when integrating renewable energy sources with the grid are cost and the stochastic nature of load, wind power, and solar power. This thesis addresses both concerns by finding the least-cost configuration of renewable energy technologies for a given cost of fossil fuel generation such that all demand for power is met. First, this thesis presents a model that simulates the generation and consumption of energy on an hour-by-hour basis in the PJM grid with solar power, offshore wind, onshore wind, and
centralized battery storage as input variables. All load that is not met through renewables and storage is instead supplied through fossil fuel generation. The use of a Finite-Difference Stochastic Approximation algorithm was pioneered, calibrated, and proven to find the least-cost configuration of renewable power build-outs for any given set of cost parameters and input data detailing load and hourly capacity factors. The optimization algorithm was run for four different scenarios over a wide range of fossil fuel penalties. The resulting curves detailing the relationship between fossil fuel penalty and energy outcomes provide a roadmap for understanding how we might gradually transition from a fossil-fuel-powered energy grid to one with a high penetration of renewables.
8 REFERENCES


This thesis was printed and bound by Princeton Printer, formerly Triangle Reprocenter, est. 1939. The body of this work is typeset in Princeton Monticello, the origins of which can be traced back to Binny & Ronaldson of Philadelphia, est. 1796, the very first successful type foundry in America. The typeface was revived in the 1940s when Princeton University Press commissioned Linotype to design a historically appropriate face to be used in *The Papers of Thomas Jefferson*. In 2003, Princeton Monticello was digitized by the renowned type designer Matthew Carter, a MacArthur fellow known for designing two of the Internet’s pioneering typefaces, Verdana and Georgia.