Deterministic Scheduling for Transmission-Constrained Power Systems Amid Uncertainty

by

Joshua Daniel Lyon

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved December 2014 by the
Graduate Supervisory Committee:

Muhong Zhang, Co-Chair
Kory W. Hedman, Co-Chair
Ronald G. Askin
Pitu Mirchandani

ARIZONA STATE UNIVERSITY

May 2015

ABSTRACT

This research develops heuristics for scheduling electric power production amid uncertainty. Reliability is becoming more difficult to manage due to growing uncertainty from renewable resources. This challenge is compounded by the risk of resource outages, which can occur any time and without warning. Stochastic optimization is a promising tool but remains computationally intractable for large systems. The models used in industry instead schedule for the forecast and withhold generation reserve for scenario response, but they are blind to how this reserve may be constrained by network congestion. This dissertation investigates more effective heuristics to improve economics and reliability in power systems where congestion is a concern.

Two general approaches are developed. Both approximate the effects of recourse decisions without actually solving a stochastic model. The first approach procures more reserve whenever approximate recourse policies stress the transmission network. The second approach procures reserve at prime locations by generalizing the existing practice of reserve disqualification. The latter approach is applied for feasibility and is later extended to limit scenario costs. Testing demonstrates expected cost improvements around 0.5%–1.0% for the IEEE 73-bus test case, which can translate to millions of dollars per year even for modest systems. The heuristics developed in this dissertation perform somewhere between established deterministic and stochastic models: providing an economic benefit over current practices without substantially increasing computational times.

This one's dedicated to Becki and Belle,

who've blessed me in more ways than I can tell.

ACKNOWLEDGEMENTS

I am deeply grateful for the guidance and mentorship provided by the committee co-chairs, Muhong Zhang and Kory Hedman. From them I have learned many technical and practical aspects to research and communicating effectively. Thank you for your collaboration, securing research funding, and investing so much of your valuable time. Your concern for my well being was apparent all the way through, for which I am flattered and especially grateful.

Thank you for critical funding from the Department of Energy (DOE), Power Systems Engineering Research Center (PSERC), and the National Science Foundation (NSF). The research does not necessarily reflect the viewpoints of these funding organizations. Additional fellowships and travel support came from the department, college, and university, for which I am very grateful.

Thank you to Eugene Litvinov for inviting me to work as an intern at ISO-NE. Eugene provided encouragement and valuable insights into the electric industry. Many thanks to Tongxin Zheng, Jinye Zhao, Slava Maslennikov, and Parviz Alivand for their discussions on electric energy markets and the challenges faced by the industry. Further thanks to Yonghong Chen of MISO for her feedback on portions of this research.

I am indebted to the academic review process. Thank you to the entire committee for your time and service. Also thank you to the journal reviewers whose anonymous comments contributed significantly to improving the quality of this work.

Several student colleagues provided valuable contributions. Fengyu Wang helped finalize and test the work on locational reserve pricing and we coauthored a paper on the topic. Others who also provided written and oral feedback include Greg Thompson, Chao Li, Nikita Singhal, Yousef Al-Abdullah, and Garret LaBove. Thank you all; I will think back to working with you with fond memories.

Finally, thank you to friends and family. Thank you to my parents for instilling curiosity and a willingness to work hard; to my wife for her constant and unwavering support; and to those unnamed friends, family, and colleagues who make life what it is.

My lovely wife receives the utmost appreciation. She was at my side throughout graduate school. She provided valuable feedback for my work. She stayed home pregnant during the first internship at ISO-NE and left friends and family to travel with me during the second. Rebecca, you have brought many changes to my life and have helped make me a better person. I love you.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Overview

Power system production scheduling shares many similarities with standard supply chain problems. There are suppliers who base their activities on consumer forecasts and the product is delivered across a relatively inflexible network. However, power production scheduling has some important distinctions from other supply chain problems. The most notable distinction is that Kirchhoff's laws dictate the path power takes once it has been injected into the system. The resulting power flow equations are non-convex and, therefore, difficult to model accurately for efficient optimization algorithms. Reliability standards further increase the complexity by requiring feasibility across a large number of plausible scenarios. It is well recognized in the research community that modeling many scenarios to high precision is not yet computationally tractable for large power systems (Q. P. Zheng, Wang, & Liu, 2014; FERC, 2014a). Approximate models are, therefore, used in practice but may be unnecessarily expensive or technically infeasible if they fail to anticipate network constraints.

Commercial scheduling models have steadily improved over the years. Linear power flow approximations facilitate linear programming and mixed integer linear programming algorithms. Many scheduling decisions are made day-ahead to provide adequate lead time for generators to prepare for operation. Day-ahead scheduling models, ideally, would recognize and account for critical constraints because many decisions are fixed at this stage. Day-ahead scheduling is also coupled with electricity markets in many parts of the United States (US), and inaccurate model approximations can have the additional downside of biasing market outcomes (Pritchard, Zakeri, & Philpott, 2010). Accurate models are important to ensure that power systems run economically and reliably, and promoting such efficiency through improved computational methods is a core strategical goal of the Federal Energy Regulatory Commission (FERC, 2014b). This research supports that initiative by developing effective approaches that can be implemented in the near future.

Reliability standards regulate operations to protect against uncertainty. For example, the N-1 standard requires operators to continue serving demand following the failure of any single generator or transmission line. These resource outages are called *contingencies* in this industry. There is a 15-minute re-dispatching period allowed following a contingency for operators to rebalance supply and demand and to alleviate stress on the transmission network (NERC, 2014b). Likewise, operating conditions must be managed for other forms of uncertainty, such as intermittent generation from wind and solar plants. Dispatchable backup capacity called *reserve* is used to respond to disturbances when they occur. Reserve is considered deliverable if it can be utilized without violating transmission constraints. The focus of this dissertation is on ensuring reserve deliverability in congested networks.

Congestion is a major concern for some operators. Some areas are operating closer to their limits as transmission expansion has lagged behind new investments in wind generation

(ERCOT, 2007; Gu, McCalley, & Ni, 2012). Meanwhile, intermittent generation from renewable resources (mostly wind) is becoming more pervasive as governments enact stricter renewable portfolio standards to reduce emission of greenhouse gases. Fluctuations in available renewable energy are offset in real time by changing the production of conventional generators. This dynamic complicates operations because network flows change with the wind, which in turn makes N-1 more difficult to ensure as conditions become less predictable. There is value in operational tools that can better manage congestion because such tools will support better use of available resources.

1.2 Research Focus

Unit commitment is an optimization problem that schedules generators (i.e., units) to be available (i.e., committed) to provide power to consumers. The traditional approach is to minimize cost subject to a variety of physical and operational constraints. This dissertation focuses on managing constraints on recourse decisions (i.e., corrective actions, e.g., generator re-dispatch) made after uncertainty has been revealed. The primary goal is to ensure that operating constraints can be satisfied across a variety of plausible scenarios. This target emphasizes economics for forecasted conditions and feasibility across a range of scenarios.

The initial focus of this work is on ensuring reliability for N-1 contingencies. The costs of recourse actions are de-emphasized here because individual contingencies are unlikely to occur and, therefore, have small effects on total expected cost. However, large wind fluctuations are common in some parts of the world – common enough to significantly influence cost on a regular basis. The cost of recourse decisions is considered in the later part of this dissertation to support economical decisions in systems dealing significant wind fluctuations.

Operators prepare for uncertainty by acquiring generation and transmission flexibility ahead of time. The generation flexibility is called reserve. There are two popular approaches for ensuring reserves are deliverable. The more precise strategy is to directly model transmission constraints on recourse decisions using tools like stochastic programming or robust optimization. However, these methods have limited practical use due to their computational limitations. The approach favored by operators today is to instead model flexibility constraints that address reliability in a more approximate manner. These flexibility constraints are labeled in this dissertation as *deterministic* policies due to their indirect treatment of uncertainty. Generation flexibility is obtained in part by reserve requirements that specify how much generators and other resources must collectively be able to change production within a specified period of time. Transmission flexibility is encouraged by preemptively limiting flows below physical limits (this tends to provide more leeway for corrective actions) and by controlling the location of reserve using zones. Zonal models distribute reserve across the grid but do not distinguish between locations at a detailed level. The effect of this limitation is that solutions to the unit commitment problem may fail to satisfy reliability targets due to imprecise modeling of transmission constraints. More details on existing approaches are covered by the literature review in Chapter 3.

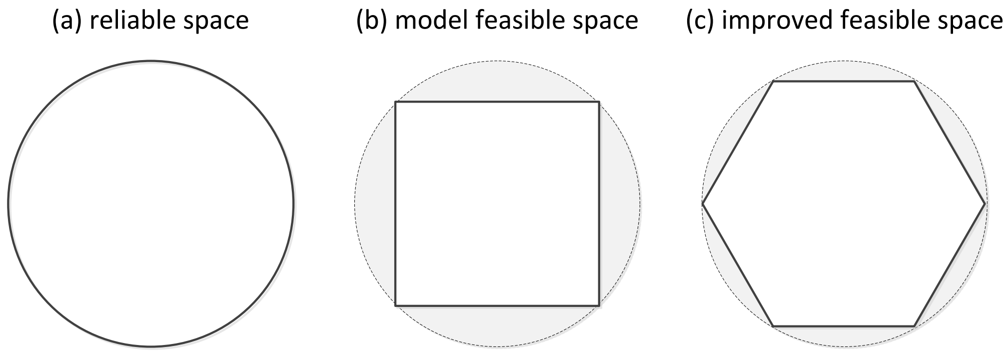


Figure 1.1: Model feasible regions compared to the N-1 reliable space.

The primary research objective is to mitigate transmission constraints on recourse decisions. Consider the simplistic representation of all N-1-reliable solutions in Figure 1.1 (a). Modeling limitations force operators to use an approximate feasible set like in Figure 1.1 (b) in their model if they want to guarantee reliability. The approximations result in missed opportunities because good solutions may fall outside the approximate feasible region. The approaches developed in this dissertation are meant to improve the accuracy of deterministic models to define a feasible set roughly akin to Figure 1.1 (c). The gains in accuracy are achieved by introducing a minimal number of additional variables and constraints, and the solution methodologies are designed to be computationally tractable. The tension between finding a good answer and finding a timely answer is managed by leveraging partial information on what are likely to be critical constraints.

Two types of techniques are proposed in this dissertation. The first approach is covered in Chapter 4. It anticipates recourse actions in a simplified way (as an affine policy of uncertainty), evaluates their influence on transmission stress, and acquires additional reserves when transmission stress is high. The additional reserve increases the range of potential corrective actions and improves the likelihood of there being a feasible re-dispatch solution. The approach is labeled *congestion-based reserve requirements* because the minimum reserve quantity is dependent on the predicted transmission stress. Congestion-based reserve requirements work best when additional reserves facilitate alternate recourse actions that can circumvent congestion. This technique is similar to other flexibility requirements but can anticipate when and where transmission bottlenecks will arise without the use of reserve zones.

The second approach is introduced in Chapter 5. It disqualifies reserves that are undeliverable so that reserve requirements must be satisfied by resources at more favorable locations. Today, reserve disqualification is already used as a corrective tool when the scheduling model returns an unreliable solution. Successful application depends heavily on operator insight and experience. This dissertation introduces a mathematical program to disqualify reserves and generalizes the structure to be scenario-specific. The generalized constraints account for different congestion patterns that may arise across periods and scenarios. Reserve disqualification is integrated into a decomposition algorithm that can be viewed as a heuristic for two-stage stochastic programming. In effect, reserve disqualifications become cuts that remove unreliable solutions from the feasible space (but may also remove optimal solutions in the process). The algorithm sacrifices the precision of stochastic models in order to improve computational tractability and it does so using a form of constraint that is already familiar to power system operators.

All methods developed in this dissertation can be used in conjunction with existing models (including many reserve requirements from the literature) to improve reserve deliverability. They are also compatible with established market settlement strategies. In Chapter 5, a new pricing scheme is developed that further rewards resources at prime locations. This represents a step forward over traditional markets that derive a single price for each zone. Locational reserve pricing is a potential benefit of stochastic programming (Francisco D. Galiana, Bouffard, Arroyo, & Restrepo, 2005), but similar developments have been elusive for deterministic models up until now.

Deterministic models largely aim to ensure sufficient reserve availability. The final part of this dissertation in Chapter 6 extends reserve disqualification to consider the cost of dispatching reserves. A new methodology is proposed that disqualifies backup capacity on the basis of it being too expensive to utilize during a coordinated scenario response. This effectively removes both unreliable and expensive solutions from the feasible space, as in Figure 1.2 (c). The approach can be particularly effective in inflexible systems when large forecast deviations are likely, such as in systems that regularly experience large deviations in available wind power. The consideration of recourse costs appears to be unexplored by deterministic reserve requirements in the literature. The new modeling approach fills this gap while also retaining some similarity to capacity requirements that are already used in practice.

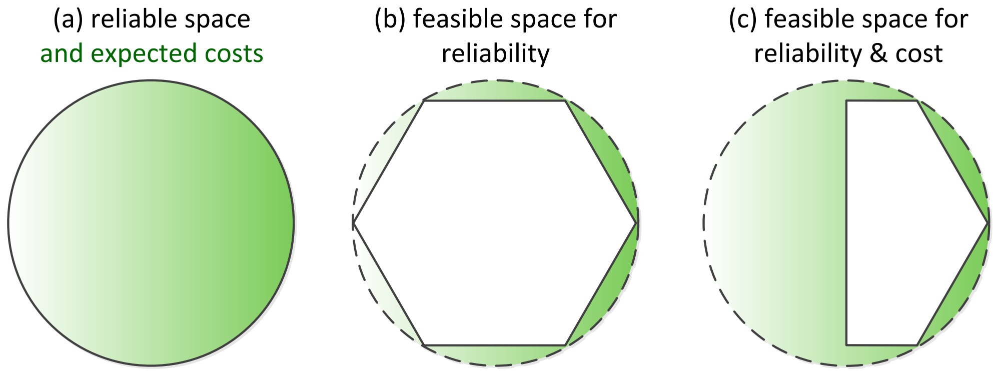


Figure 1.2: Model feasible regions that result from constraints defined for reliability (b) and for cost (c). The darker shading represents solutions with lower expected costs. The model in (c) removes solutions from the feasible set that may be expensive when evaluated over random scenarios.

1.3 Summary of Chapters

Chapter 2 overviews aspects of the electric power industry that are relevant to this dissertation. The topics include physics of power systems, defining reliability, and summarizing how reliability is typically addressed. Formulations are provided for power system scheduling models. This chapter introduces important concepts that influence the optimization models used for power production scheduling.

Chapter 3 reviews the literature on day-ahead scheduling for power systems under uncertainty. Most existing research investigates reserve quantities under the assumption that all reserves are deliverable. There is relatively little work on ensuring reserve deliverability, and this dissertation works to fill that gap. Deterministic approaches are reviewed and compared with stochastic models from the literature.

Chapter 4 investigates congestion-based reserve requirements, which help ensure reserves are deliverable within zones. The method balances generation and transmission flexibility in a rough but tractable way. Testing on the IEEE-73 bus test case demonstrates an improvement to N-1 reliability without increasing operating costs. The results indicate effectiveness for both transmission and generator contingencies. This chapter may be read out of order because the methodology is largely independent from those in later chapters.

Chapter 5 generalizes existing reserve disqualification approaches and develops a framework for their implementation. The proposed approach is scenario-specific, enabling operators to address distinct congestion patterns that can arise for different contingencies during different times of day. This contrasts the static policies that are commonly applied today to handle a wide range of conditions. The approach also allows for locational reserve pricing that is absent from traditional market structures. Testing on the IEEE-73 bus test case demonstrates its effectiveness when it is beneficial to control reserve locations *within* the reserve zones.

Chapter 6 extends reserve disqualification to mitigate wind uncertainty. This work modifies existing capacity requirements to anticipate the cost of dispatching backup capacity if and when it is needed (this cost may include penalties for violating reliability targets). The model distinguishes suitable capacity for individual scenarios by extending the concept of response sets from Chapter 5. The approach is shown to identify more economical solutions for the IEEE 73-bus test case than alternative deterministic policies in the literature.

Chapter 7 concludes and outlines opportunities for future work. There are several avenues that may be followed to improve the accuracy and scope of the proposed methodologies. Future research should investigate protecting against different forms of uncertainty and at different time scales, such as during short-term scheduling moving closer to real time. There may also be opportunities to combine the methodologies with stochastic programming in order to help those more complex procedures converge in shorter time frames.

The appendices define acronyms, nomenclature, and formulations used in this dissertation. They include details on the working example explored in Chapters 4 and 5. In addition to the appendices, all terms are identified upon first use in the dissertation body.

The work in this dissertation has lead to three journal publications thus far. The model from Chapter 4 is covered by Lyon, Hedman, and Zhang (2014). The modeling and pricing aspects from Chapter 5 are covered by Lyon, Zhang, and Hedman (2014) and Lyon, F. Wang, Hedman, and Zhang (2014). The work in Chapter 6 is under review by an academic journal.

Chapter 2

BACKGROUND

2.1 Structure of Electric Power Industry

In 2011, the United States (US) generated four billion megawatt hours (MWhs) of energy across a network comprising 200,000 miles of high-voltage transmission lines and many more low-voltage distribution lines (Energy Information Administration, 2013b, 2013a). The contiguous US and most of Canada are served by the Western, Eastern, and Texas Interconnections, each representing a large alternating current (AC) circuit with tight operating limits. These interconnections are mostly independent and are only loosely tied together by direct current (DC) transmission lines so that trading can occur between regions. Such large electrical networks are popularly referred to as the most complex machines ever constructed by mankind.

Figure 2.1 describes the electricity supply chain. The electrical network facilitates power delivery from suppliers (generators) to consumers (loads). High-voltage transmission carries power long distances and low-voltage distribution services the local residential, commercial, and industrial customers. Transmission and distribution are both understood to be natural monopolies that are highly regulated (Kirschen & Strbac, 2004; Rothwell & Gómez, 2003). Their operations are often treated independently: transmission models

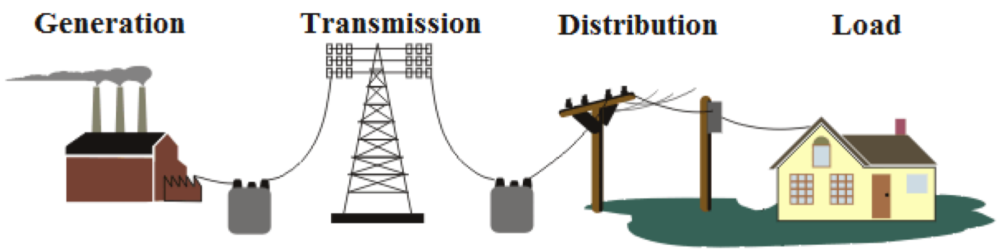


Figure 2.1: The electric supply chain (Hedman, 2009).

typically include an abstraction of the distribution grid and vice versa. The decisions considered in this dissertation are made at the level of high-voltage transmission.

Entities may be classified as suppliers, transmission, and consumers. Suppliers are controlled by central operators who coordinate production. The majority of day-to-day decisions are directly related to generator control. The remaining parts of the system are relatively inflexible. Power systems are designed in the US to make power available to consumers whenever they want it at the moment they call upon it. While consumers are mostly inflexible, some smart grid efforts aim to leverage flexible consumers to alleviate undesirable conditions. Efforts are underway to increase the flexibility of consumers and transmission, but most dispatching flexibility today originates from control of generating units.

Central operators called balancing authorities (BAs) or system operators (SOs) are responsible for maintaining operating conditions for individual control areas that typically encompass hundreds of generators and thousands of transmission lines. Some control areas are operated by vertically integrated utilities and others employ markets where participants compete with each other. In both vertically integrated and competitive markets, a centralized operator controls generator dispatch and is responsible for maintaining reliability.

Vertically integrated utilities serve about two-thirds of load in the contiguous US and are prevalent in the Northwest and Southeast regions. These utilities are granted a monopoly over generation, transmission, and distribution, and may coordinate these resources to minimize cost. However, cost minimization can be disincentivized by *cost-of-service* regulations whereby utility profits are regulated to stay near a target rate-of-return (Rothwell & Gómez, 2003). The efficiency of vertically integrated utilities is uncertain and hard to validate because operating procedures and analyses are rarely, if ever, published. Electricity markets emerged in the US over the past 20 years in hopes of improving efficiency through competition and improved operational practices.

Electricity markets are operated by non-profit independent system operators (ISOs) that are responsible for managing electricity auctions and ensuring reliability. Electricity markets are more complex than ordinary markets because they are subject to complex physical and operational constraints. ISOs operate around the same physical constraints as vertically integrated utilities but base dispatching decisions on bids from market participants: generators provide supply bids and load serving entities (LSEs) purchase energy on behalf of end-use customers [1] . Table 2.1 and Figure 2.2 summarize the span and footprint of ISOs in the US as of 2012.

All ISOs in the US are regulated by the Federal Energy Regulatory Commission (FERC). Any changes to market design require FERC approval. FERC proposed a standard market design in 2002 to encourage more uniform operations and to correct for market inefficiencies. While the push for standardization was subsequently abandoned (FERC,

[1] The LSE is often a utility that operates the distribution network in the local area.

Table 2.1: Independent system operators in the US and their approximate peak loads.

Abbreviation	Name	Peak load
CAISO	California Independent System Operator	50,000 MW
ERCOT	Electric Reliability Council of Texas	65,000 MW
MISO	Midcontinent Independent System Operator	100,000 MW*
PJM	Pennsylvania - New Jersey - Maryland	145,000 MW
NYISO	New York Independent System Operator	33,000 MW
ISO-NE	Independent System Operator - New England	28,000 MW

*The peak load for MISO will increase because it recently expanded its footprint.

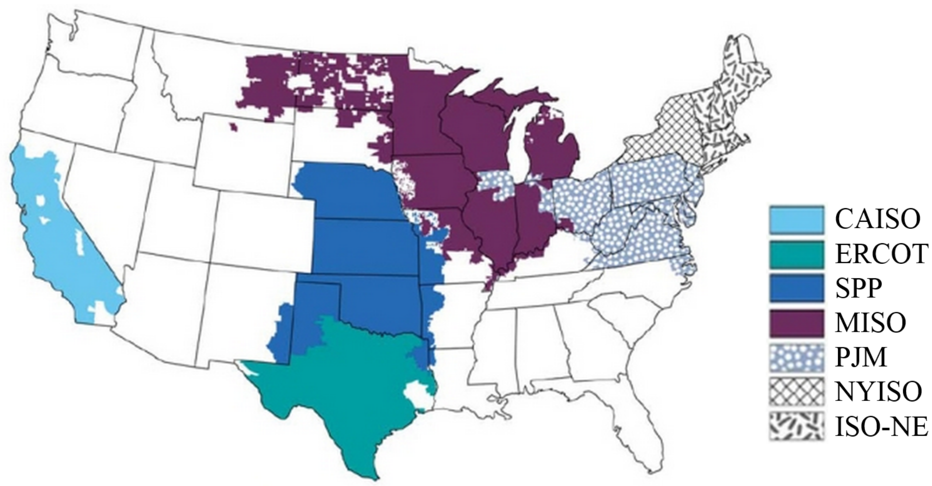


Figure 2.2: Map of independent system operators[2](U.S. Government Accountability Office, 2011).

2005), ISOs have voluntarily converged on many issues by adopting the same market design elements suggested by FERC. Nevertheless, operations still differ across ISOs in several important ways as operators experiment with different practices and market designs.

ISOs and vertically integrated utilities are subject to reliability standards that are set and enforced by the North American Electric Reliability Corporation (NERC). Therefore, operating procedures used in one setting are often applicable in another. It is common

[2]Southwest Power Pool (SPP) has markets for trading energy but is not an ISO as of 2013.

for researchers to work in the context of electricity markets because the industry practices are relatively transparent and applications developed for markets are often transferable to vertically integrated utilities.

2.2 Scheduling and Control

Power scheduling shares several commonalities with traditional supply chain problems. Operators forecast the electrical load (demand) at each location and then schedule generators (suppliers) subject to resource costs and capacities; flexibility is withheld to mitigate forecast errors; and the network models are subject to node balance and arc capacity constraints. However, there are important distinctions to be made for power systems as summarized in Table 2.2.

The first distinction for power systems is that injections and withdrawals travel quickly (at the speed of light) and are literally continuous. Adjustments are constantly needed to handle deviations in supply and demand. Conventional generators produce power

Table 2.2: Power system comparison to general supply chains.

	Supply Chain Property	Power System Analogy
✓	Limited supply	Generator constraints
✓	Demand forecast	Load forecast, etc.
✓	Backup capacity	Reserves, committed generation
~	Conflicting objectives	Competing market participants
~	Backlogging	Load shifting to different periods
×	Shipment control	Flows are subject to Kirchhoff's laws
×	Non-zero shipment times	Power is delivered at the speed of light
×	Safety stock	Power is consumed immediately (no inventory)
×	Multiple products	Only electrons are transmitted

via large magnetic turbines that are designed to spin at a nearly constant frequency. The rotational frequency only changes when mechanical torque from the fuel source does not match the magnetic torque from the load (Wood & Wollenberg, 1996). Operators always try to balance supply and demand because frequency must be kept within tight specification limits (NERC, 2014b) [3] . They maintain this balance by mobilizing backup capacity called *reserve*. At present, reserves are mainly provided by generators because consumers are inflexible and energy storage is expensive.

The second distinction from traditional supply chains is that power flows are subject to Kirchhoff's laws that describe a unique mapping from injections to power flows. These phenomena are non-convex and make power flow models significantly harder to solve. Section 2.2.1 describes a detailed power flow model and linear approximations that are commonly used in practice.

There are other power topics related to the supply chain literature not covered in this dissertation. Mitigating market power is a major concern in market design and loosely relates to non-cooperation in supply chains, e.g., game theoretic models are common in the literature. Furthermore, some demand response models defer load to later periods and are, in that way, similar to backlogging.

This dissertation improves upon scheduling algorithms applied to day-to-day operations. It is assumed that sufficient capacity is available to satisfy reliability standards. The

[3] Frequency control is a system-wide effort that benefits from coordination between BAs. The standard practice is for each area to keep net imports close to a scheduled amount, with deviations measured by the area interchange error (ACE) (NERC, 2014b).

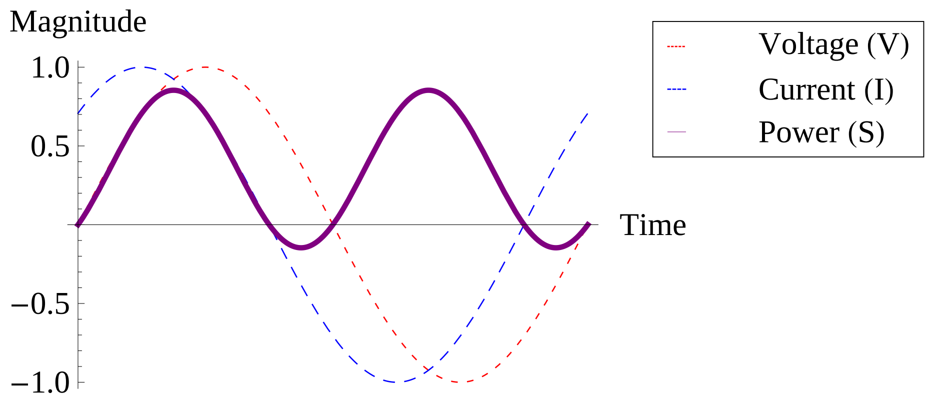


Figure 2.3: Instantaneous power as the product of voltage and current.

lack of any reliable operating solution is a medium-to-long-term planning issue that falls outside the scope of this dissertation.

2.2.1 Power Flow.

Power flows through AC circuits according to complex physical laws. The flow down individual transmission lines is driven by voltage (V) and current (I), which oscillate about 60 times per second along sinusoidal curves. These curves are rarely aligned perfectly [4] . Voltage and current become out of phase due to the electrical properties of the system, i.e., capacitors cause a lag in voltage and inductors cause a lag in current due to the delayed formation of electric and magnetic fields. The offsets between curves is crucial because it is closely related to how power travels through the system.

Total power (S) changes by time according to $S(t) = V(t) \times I(t)$. This relationship is shown in Figure 2.3, where it can be seen that the offset between voltage and current causes brief periods where the total power is negative. It is useful to decompose power into two parts: one part that is always nonnegative (real power P) and another that oscillates

[4] The offset between curves is measured by phase angles. The angle difference is zero when the curves are aligned and negative when voltage lags current.

around zero (reactive power Q) (Das, 2006). Although they are both parts of the same whole, real and reactive power are given different treatments due to their unique influences on the system. Real power performs the electrical work in consumer devices and reactive power is closely tied to voltage magnitude. Power flow models often consider only the average magnitude of real and reactive power in steady state. The power sent from node m to n across line $l = (m,n)$ is influenced by the electric properties of the line as described by Das (2006):

$$P_{mn} = \left[V_m^2 - V_m V_n \cos(\theta_m - \theta_n)\right] G_l - V_m V_n \sin(\theta_m - \theta_n) B_l \tag{2.1}$$

$$Q_{mn} = -V_m V_n \sin(\theta_m - \theta_n) G_l - \left[V_m^2 - V_m V_n \cos(\theta_m - \theta_n)\right] B_l, \tag{2.2}$$

where V_n is voltage magnitude and θ_n is the voltage angle at node n [5] . The power flows are influenced by transmission properties like conductance G_l and susceptance B_l. These electrical properties determine how easily line l admits power to flow across it – modest values impede current and encourage power to take alternative paths. The conductance G_l influences how much real power is dissipated as heat. The susceptance B_l influences how much alternating current is resisted by magnetic fields surrounding the line. These phenomena result from the resistive and inductive properties of the transmission line. Long lines also have a significant capacitive effect that is not explicitly modeled in the above formulation (Das, 2006).

[5] The voltage magnitude and phase angle define the sinusoidal curve in Figure 2.3 for a single location. Convention defines V, P, and Q as the root mean square (RMS) of the oscillating value (Glover & Sarma, 1994). Also note that voltage angles can be measured relative to an arbitrary reference point because all that matters in (2.1) and (2.2) is the angle *difference* between adjacent nodes.

AC power flow models couple equations (2.1)–(2.2) with flow balance constraints. Such models include decision variables for the voltage magnitudes and phase angles. Scheduling problems are complicated by the non-convexity of the AC power flow equations, and so the following simplifications are common to linearize the problem (Das, 2006):

- Assume that angle differences are small to substitute $\cos(\theta_m - \theta_n) \approx 1$ and $\sin(\theta_m - \theta_n) \approx \theta_m - \theta_n$. This assumption is reasonable for high-voltage transmission lines.

- Assume voltages are roughly uniform to substitute $V_m = V_n \approx 1$ per-unit. Voltage is usually controlled using ancillary services to stay between 0.95 and 1.05 per-unit.

- Assume $G \ll B$. Taken with the above assumptions, this implies that reactive power is negligible and that there are no power losses (H. Liu, Tesfatsion, & Chowdhury, 2009).

The resulting power flow equations reduce to $P_l = B_l(\theta_n - \theta_m)$ and $Q_l = 0$, which define the so-called DC power flow model. The DC power flow is a linear approximation of AC systems. Significant missing components include losses and reactive power, and so adjustments are often made to help correct for these inaccuracies (Wood & Wollenberg, 1996). This dissertation forgoes such adjustments and retains the standard DC power flow problem:

$$i_n + \sum_{l \in LO(n)} f_l - \sum_{l \in LI(n)} f_l = 0, \qquad \forall n \in N \tag{2.3}$$

$$f_l = B_l(\theta_n - \theta_m), \qquad \forall l = (m,n) \in L, \tag{2.4}$$

where i_n denotes the net injection (supply minus demand) at node n and f_l denotes the flow on line l relative to an arbitrary reference direction. Equation (2.3) is a standard flow balance constraint found in many network problems. This equation says that the injection plus the flow into a node must equal the flow out of the node. Equation (2.4) is different because it is unique to power systems. This equation says that flow depends on the voltage angle difference across the line and on how susceptible the line is to conducting alternating current, as defined by the electrical susceptance B_l. Power travels across the line toward the smaller angle because susceptance is always negative. Therefore, power can be thought of as flowing from regions with high voltage angles to regions with low voltage angles, distributed across lines according to their susceptance.

Operators generally cannot control the path power takes once it has been injected into the system. The above model solves for voltage angles because they complete the mapping from injections to flows. The existence of a unique mapping can be shown by demonstrating that there are $|N|+|L|$ decision variables (θ_n and f_l) and the same number of linearly independent equality constraints. The power flow equations (2.3)–(2.4) collectively have the correct number of constraints. However, they are not linearly independent because (2.3) implies $\sum_{n\in N}\left(i_n+\sum_{l\in LO(n)} f_l-\sum_{l\in LI(n)} f_l\right)=0$, which can be used to determine the flow for one of the lines without the need for (2.4). The required number of linearly independent constraints comes about by setting an arbitrary voltage angle equal to zero:

$$\theta_{\mathrm{REF}}=0. \tag{2.5}$$

The voltage angles are then measured with respect to the angle at the reference location $n = \text{REF}$. Equation (2.5) establishes that a unique set of voltage angles will arise from any given set of injections. However, this constraint is omitted from many studies because it does not affect the estimated power flows (which are driven by the differences between angles and not the angles themselves).

There is a simpler way to model power flows without solving for voltage angles θ. It is possible to determine parameters that directly map locational injections to power flows. In particular, *power transfer distribution factors* (PTDFs) describe how much power exchanged between two nodes will travel along a particular transmission line (M. Liu & Gross, 2004). PTDFs are called *shift factors* when defined with respect to a constant reference node: then $PTDF_{nl}$ is the flow on line l for every unit of power sent from node n to the reference. Shift factors are additive, such that sending power to node n is equivalent to first sending it to the reference node and then to n. Equation (2.6) below uses this property to characterize flow by sending all surplus generation ($i_n > 0$) to the reference node and then dispersing power to the locations with surplus demand ($i_n < 0$) [6] :

$$f_l = \sum_{n \in N} PTDF_{nl} i_n, \qquad \forall l \in L. \tag{2.6}$$

For fixed injections, this relationship provides the same solution as (2.3)–(2.4) but with fewer variables. This is a popular modeling approach because it makes it easy to ignore uncritical lines by removing the correspondence constraint from the model – which is computationally useful in large systems where many lines are rarely congested. The primary

[6] Power sent from the reference to n can be modeled as a negative transfer in the opposite direction.

limitation of using shift factors is that they must be calculated off-line and updated whenever the network topology changes. However, the benefits generally out-weigh the limitations and so shift factors are commonly used in power system scheduling (Stott, Jardim, & Alsaç, 2009).

The accuracy of DC models is reasonable when warm-start techniques are used to tune the formulation around the expected system state. Line flow errors when using such techniques are evaluated by Stott et al. (2009) for six large test cases. The errors are less than 2% on average but can be as high as 20% in the extreme [7] . The accuracy generally improves as the system state becomes more predictable moving closer to real time. However, it is a good idea to check contingencies for AC feasibility because they can cause an abrupt change to the system state. The analyses in this dissertation do not check for AC feasibility, but it is understood that infeasible solutions will be found and addressed by system operators before the system reaches real time.

2.2.2 DC Optimal Power Flow. The direct current optimal power flow (DCOPF) is a network optimization problem that uses linear approximations like those from Section 2.2.1. One potential formulation is described in this section. The objective is to dispatch generators in a way that minimizes cost [8] subject to network and generator constraints. Appendix B describes the nomenclature for the DCOPF and all other models used in this dissertation.

[7] Stott et al. (2009) only report errors for lines loaded above 50 MW and 70% of their rated capacity.

[8] The formulation presented in this chapter relates to generator decisions. Other elements can be modeled with little change to the problem structure, e.g., flexible demand are sometimes modeled as “negative” generators.

Generators typically have convex total cost curves because fuel efficiency tends to decrease with production (Wood & Wollenberg, 1996). Piecewise linear cost curves are common in practice so that the problem can be solved using linear programming (Carrión & Arroyo, 2006; Pandžić, Qiu, & Kirschen, 2013). The following objective and constraints model piecewise linear costs:

$$\text{Minimize:} \quad \sum_{g\in G}\sum_{t\in T}\sum_{i\in I(g)} c_g^i \tilde{p}_{gt}^i, \tag{2.7}$$

$$\sum_{i\in I(g)} \tilde{p}_{gt}^i = p_{gt}, \qquad \forall g \in G, t \in T \tag{2.8}$$

$$0 \le \tilde{p}_{gt}^i \le \tilde{P}_g^i, \qquad \forall i \in I, g \in G, t \in T, \tag{2.9}$$

where c_g^i is the production cost and $\tilde{p}_{gt}^i$ is the production quantity for generator g in interval i during period t. Each generator has a set of production intervals $I(g)$ with differing marginal costs. Generator g can produce up to $\tilde{P}_g^i$ in interval i and this is tied back to the total output p_{gt} by constraints (2.8)–(2.9). The cheapest intervals are filled first in all optimal solutions, which corresponds to the actual production cost when the marginal cost curve is non-decreasing, i.e., the total cost is convex.

Power flows can be modeled by solving for voltage angles (2.3)–(2.4) or using shift factors (2.6). The shift factor formulation is used here with network constraints

$$i_{nt} = \sum_{g \in G(n)} p_{gt} - D_{nt}, \qquad \forall n \in N, t \in T \tag{2.10}$$

$$f_{lt} = \sum_{n \in N} PTDF_{nl} i_{nt}, \qquad \forall l \in L, t \in T \tag{2.11}$$

$$\sum_{n \in N} i_{nt} = 0, \qquad \forall t \in T \tag{2.12}$$

$$-F_l \leq f_{lt} \leq F_l, \qquad \forall l \in L, t \in T, \tag{2.13}$$

where i_{nt} is the net injection (generation minus load D_{nt}) at node n. Constraint (2.12) specifies that total supply must equal demand and (2.13) constrains the flow on each line to not exceed its capacity F_l.

Generators are constrained with respect to capacity and ramping capability. Predetermined parameters indicate periods when the generator is on ($u = 1$), turned on ($v = 1$), and turned off ($w = 1$). A generator that is on must operate within its upper and lower limits according to

$$p_{gt} \leq P_g^{\max} u_{gt}, \qquad \forall g \in G, t \in T \tag{2.14}$$

$$p_{gt} \geq P_g^{\min} u_{gt}, \qquad \forall g \in G, t \in T \tag{2.15}$$

$$p_{gt} - p_{g,t-1} \leq R_g u_{g,t-1} + R_g^{\text{SU}} v_{gt}, \qquad \forall g \in G, t \in T \tag{2.16}$$

$$p_{gt} - p_{g,t-1} \geq -R_g u_{gt} - R_g^{\text{SD}} w_{gt}, \qquad \forall g \in G, t \in T, \tag{2.17}$$

where $P_g^{\max}$ and $P_g^{\min}$ are the maximum and minimum capacities and R_g defines how much generator g can change production (ramp) between normal periods. Three types of ramping are considered: 1) during generator start up, 2) shut down, and 3) between periods when a generator remains on. At most one term on the right hand sides of (2.16)–(2.17) is positive and the resulting ramp restrictions reflect the respective generator status. For example, a generator g cannot ramp more than R_g^{SU} when starting up or R_g^{SD} when shutting down.

2.2.3 Unit Commitment. Unit commitment (UC) is an optimization problem that schedules generators (i.e., units) to be available (i.e., committed) to provide power in an economical manner. Commercial models include physical generator constraints and a linear approximation of power flow (Pandžić et al., 2013). In this dissertation, UC is modeled as a mixed integer linear program (MILP) that contains the DCOPF with additional binary variables describing generator commitments. Fixed costs may be incurred whenever a generator is committed as well as during start up and shut down periods. These costs are termed no load (c^{NL}), start up (c^{SU}), and shut down (c^{SD}) costs, respectively. The fixed costs are combined with the marginal production costs from the DCOPF (2.7) to get the UC objective:

$$\text{Minimize:} \quad \sum_{g \in G} \sum_{t \in T} \left(\sum_{i \in I(g)} c_g^i \tilde{p}_{gt}^i + c_g^{\mathrm{NL}} u_{gt} + c_g^{\mathrm{SU}} v_{gt} + c_g^{\mathrm{SD}} w_{gt} \right). \tag{2.18}$$

The binary commitment (u), start up (v), and shut down (w) decisions are modeled by

$$v_{gt} - w_{gt} = u_{gt} - u_{g,t-1}, \qquad \forall g \in G, t \in T \tag{2.19}$$

$$0 \leq v_{gt}, w_{gt} \leq 1, \qquad \forall g \in G, t \in T \tag{2.20}$$

$$u_{gt} \in \{0,1\}, \qquad \forall g \in G, t \in T. \tag{2.21}$$

Equations (2.19)–(2.20) ensure $v_{gt} = 1$ ($w_{gt} = 1$) whenever a generator starts (shuts down). Unit commitment also includes minimum up and down time constraints that protect the physical integrity of the generators. Minimum up and down times are constrained by

$$\sum_{q=t-UT_g+1}^{t} v_{gq} \leq u_{gt}, \qquad \forall g \in G, t \in T \tag{2.22}$$

$$\sum_{q=t-DT_g+1}^{t} w_{gq} \leq 1 - u_{gt}, \qquad \forall g \in G, t \in T. \tag{2.23}$$

Equation (2.22) specifies that generator g must remain on if started in the last UT_g time periods and, likewise, (2.23) specifies g must remain off if shut down in the previous DT_g periods. Integrality constraints are optional for the start up and shut down variables because the above constraints define facets for the (v, w) projection (Rajan & Takriti, 2005). To demonstrate that this is the case, consider how the changing commitment status allows only three possible values for the right-hand side (RHS) of (2.19). When the RHS $= 1$, the generator is turned on and (2.19)–(2.20) force $v_{gt} = 1$ and $w_{gt} = 0$. When the RHS $= -1$, the generator is turned off and the same constraints force $v_{gt} = 0$ and $w_{gt} = 1$. When the RHS $= 0$, both v_{gt} and w_{gt} must be zero because they are nonnegative and (2.22)–(2.23) allow at most one to be positive.

The base unit commitment formulation consists of (2.8)–(2.23). Unit commitment and OPF problems often include additional variables or constraints aimed at improving reliability. Some of these modeling considerations are covered in Section 2.3, which introduces important concepts related to power system reliability.

2.2.4 Scheduling Epochs. Temporal constraints limit how fast generators can change production (ramp), how long generators must remain off before being restarted, and vice versa. These phenomena encourage scheduling far in advance so that production can be coordinated between periods. In practice, ISOs make most commitment decisions one day-ahead because, by this time, forecasts are relatively accurate. From there, rescheduling is a continual process that responds to updated forecasts for load, interchange with neighboring areas, and intermittent generation such as wind and solar.

Power system scheduling is an ongoing iterative process. ISOs solve unit commitment as part of a day-ahead market (DAM) based on bids from market participants. Figure 2.4 describes the day-ahead scheduling process at a high level. After the market has cleared (ex-post), operators analyze the solution and adjust the schedule to be more reliable for the forecasted conditions (Al-Abdullah, Khorsand, & Hedman, 2013). Market rules limit the changes made during this process, e.g., CAISO avoids overturning commitments made in the DAM (CAISO, 2013). Therefore, decisions made late in the sequential process are limited to locally optimal adjustments to the market solution. The adjustments are sometimes called *out-of-market corrections* (OMCs) because the market model failed to anticipate a reliability issue that needed to be corrected [9] . OMCs describe corrections made because of inaccuracies in the market model and improved policies may reduce the number of OMCs necessary to repair unreliable solutions.

[9] The term *out-of-market correction* refers to changes made outside of the market model. Terminology varies between operators. For example, these interventions are classified as exceptional dispatches in CAISO and out-of-merit energy/capacity in ERCOT CAISO, 2013; ERCOT, 2007.

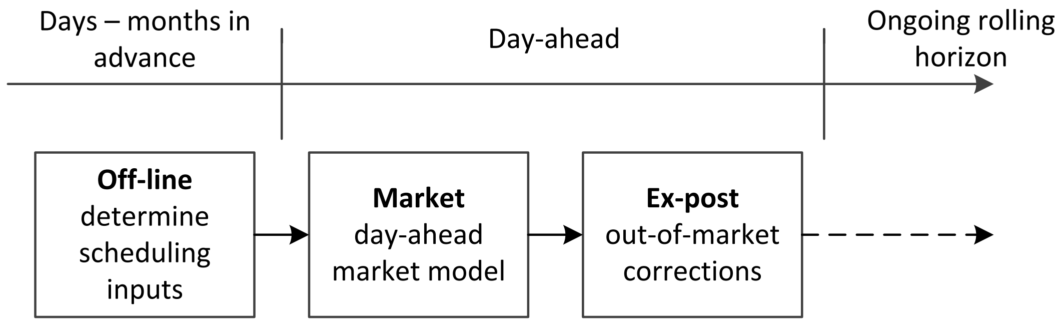


Figure 2.4: Day-ahead scheduling process in a market setting.

Table 2.3: Scheduling time horizons in CAISO.

Problem type	Horizon	Resolution	Frequency
Unit commitment	24 hours	1 hour	Daily at 10 a.m.
Unit commitment	5 hours	15 minutes	Every 15 minutes
Optimal power flow	65 minutes	5 minutes	Every 5 minutes

Generator lead times force commitment decisions to be made in advance. Fewer rescheduling options exist as operations move closer to real time. It is common for ISOs to update unit commitment over a rolling horizon as forecasts improve. For example, some important scheduling horizons in CAISO are summarized in Table 2.3 (Makarov et al., 2010). Unit commitment decisions are updated every 15 minutes and optimal power flow decisions are updated every five minutes. Moving closer to real-time (seconds to minutes), slight variations between supply and demand are corrected through separate control schemes like automatic generation control (AGC) (Wood & Wollenberg, 1996). Such frequent adjustments are necessary to respond to small forecast deviations that are constantly occurring. Generally speaking, the largest scheduling decisions are made via unit commitment and the solution is fine tuned as the system moves closer to real time.

2.3 Reliability

NERC defines operational reliability as "the ability of the electric system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system components." They outline six criteria for reliability (NERC, 2007):

1. The system is controlled to stay within acceptable limits during normal conditions.
2. The system performs acceptably after credible contingencies (sudden resource outages).
3. The system limits the impact and scope of instability and cascading outages when they occur.
4. The system's facilities are protected from unacceptable damage by operating them with facility ratings.
5. The system's integrity can be restored promptly if lost.
6. The system has the ability to supply the aggregate electric power and energy requirements of the electricity customers at all times, taking into account scheduled and reasonably expected unscheduled outages of system components.

This research aligns with the second, fourth, and sixth criteria. The second criterion relates to sudden failures of physical resources such as a generators or transmission lines (contingencies). The fourth and sixth criteria outline conditions for contingency response: NERC requires the ability to respond to the unexpected failure of a generator or transmission line without curtailing customer loads. This requirement is colloquially referred to as “N-1” to denote the loss of any single resource. The balance of supply and demand must be restored within 15 minutes of a contingency while also satisfying the emergency transmission limits (NERC, 2014b).

NERC’s definition of reliability is closely tied to system frequency. This is because many devices are designed to operate within a narrow frequency band. Frequency increases whenever supply exceeds demand and vice versa: see Figure 2.5 to visualize the impact of a generator outage. The loss of power at time τ overloads the active generators, causing them to slow down. This is indicated by the sudden drop in frequency. Generators across the network react through automatic controls to rebalance supply and demand within a few seconds. This stabilizes the frequency at its lowest point. The balancing authority overseeing the area where the contingency occurred is then responsible for dispatching enough reserve to restore frequency for the entire interconnection within 15 minutes (NERC, 2014b). Essentially, the local operator must replace the power supplied by the failed generator in order to properly contribute to frequency restoration.

Just as the frequency limits are temporarily relaxed following a contingency, thermal transmission constraints are also relaxed because lines do not overheat immediately. Figure 2.5 shows an example of how transmission limits may evolve following a contingency (Maslennikov & Litvinov, 2009). Reliability is related to how quickly operators can rebalance supply and demand while satisfying transmission constraints. A system is unreliable if it cannot progress back to normal conditions within the specified time frames. Researchers

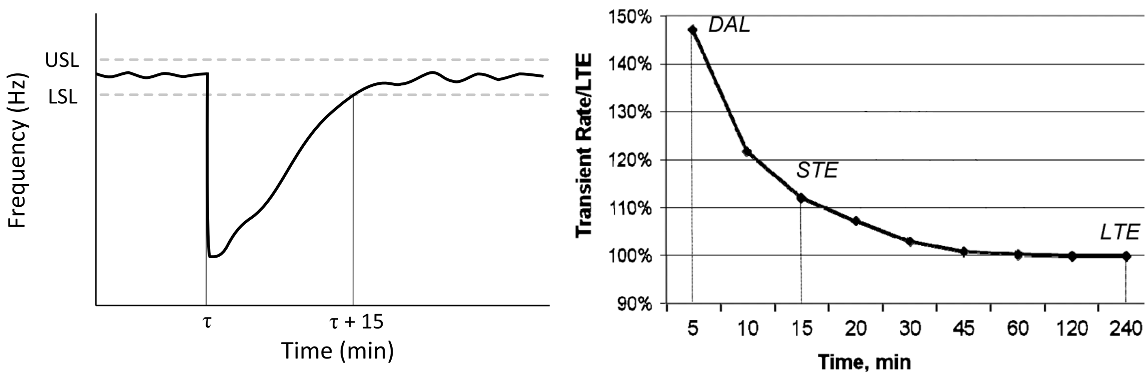


Figure 2.5: Frequency change and transmission ratings after a contingency. The frequency terms are upper and lower specification limits (USL, LSL). The transmission ratings are the drastic action limit (DAL) and short-term and long-term emergency limits (STE, LTE). The transmission sub-figure is adapted from Maslennikov and Litvinov (2009).

and practitioners tend to focus on evaluating reliability at a few fixed time points (usually 0 and 10 minutes after the contingency) [10] because simulating every moment in time is very involved and computationally intensive. This dissertation considers only the 10-minute snapshot. A mathematical formulation for contingency response is provided in Appendix C: this formulation restores the balance between supply and demand while minimizing the amount of transmission violations.

A similar process occurs on multiple time scales for other types of uncertainties. Other sources of uncertainty include load, interchange with neighboring control areas, and intermittent renewable resources like wind and solar. NERC requires frequency to remain between the upper and lower specification limits a certain percentage of the time (NERC, 2014a) and transmission constraints should also be respected in order for the system to remain reliable.

[10] Although NERC requires contingency response to restore frequency within 15 minutes, operators often target 10 minutes in their mathematical models. The five minutes of slack allows an operational buffer to solve the contingency response problem (MISO, 2012a).

The term *re-dispatch* encompasses all corrective recourse actions that fall under operator control. Recourse decisions primarily involve dispatching backup capacity in the form of generation reserve. Other forms of flexibility come from demand and transmission. Demand flexibility may involve contracting with loads to reduce consumption during emergency situations. Transmission flexibility may involve changing the network topology to relieve bottlenecks resulting from Kirchoff's laws (Hedman, 2009). This dissertation develops methods that can generally consider all of these corrective actions; however, emphasis is placed on reserves as the primary tool for corrective actions.

2.3.1 Reserves. Reserve is backup capability that allows operators to mitigate uncertainty. Generally speaking, solutions with a lot of reserve provide more recourse flexibility. There are many categories of reserve but a lack of standard terminology. Common types include regulation, contingency, operational, following, and replacement reserves, but some of these terms have slightly different meanings for different operators (Ela, Milligan, & Kirby, 2011).

Reserve is more succinctly categorized by how quickly it can be made available: 5-minute reserve can respond to frequent, small changes in load; 10-minute reserve can respond to contingencies; 30-minute reserve responds to broader uncertainties and replaces reserve expended for other purposes. Some sources of uncertainty, such as wind, can draw significantly from each of the aforementioned categories (Ela & M. O'Malley, 2012). Various types of reserve are summarized in Table 2.4 and described in more detail below. Reserve requirements are enforced for these products to ensure sufficient reserve availability.

Table 2.4: List of common reserve products. The products are ordered by the quality of service they provide (with the highest quality at the top).

Reserve product	Common name(s)
5-minute spinning	Regulation reserve
10-minute spinning	Contingency reserve, operating reserve
10-minute non-spinning	Contingency reserve, operating reserve
30-minute spinning	Supplemental reserve, operating reserve
30-minute non-spinning	Supplemental reserve, operating reserve
Total spinning	Capacity headroom

Reserve that is available in less than 5 minutes is managed by automatic generation control (AGC). A generator participating in AGC will respond to changing frequency by adjusting its output according to predetermined settings, normally according to an affine policy. AGC allows the system to respond to small uncertainties without active operator intervention. Frequency regulation provided by AGC typically corrects for small fluctuations and falls outside the scope of this dissertation.

Ten-minute reserves and above are usually procured in the up direction, e.g., the reserve requirement specifies how much generators must collectively be able to ramp up within 10 minutes. To satisfy N-1, 10-minute reserve must exceed the largest injection from a generator or importing transmission line. NERC also requires operators to maintain enough slow reserves to restore N-1 protection within 105 minutes of a contingency (NERC, 2014b). Reliability regions and system operators decide how much reserve will be required to meet NERC and their individual reliability standards.

Committed generators are said to provide spinning reserve because the turbine is already spinning and synchronized with the grid. Spinning reserve is considered high quality because it can start responding almost immediately when called upon. Other types of reserves are more likely to respond slowly or to not perform when called upon. Fast-start generators can provide non-spinning when they have not been committed. At present, the least dependable form of reserve is demand response from consumers who reduce their consumption at the operator's request (Aminifar, Fotuhi-Firuzabad, & Shahidehpour, 2009). The reliability council for each region defines how much reserve must be held of each type (NERC, 2014b). For example, the Northeast Power Coordinating Council (NPCC) obliges operators to increase 10-minute spinning reserve requirements whenever performance is poor (NPCC, 2012): the requirement might increase from 50% to 70% of the largest contingency following an event where the operator failed to recover frequency within the required time frames [11] .

The available reserves are defined in the model by

$$\dot{r}_{gt}^{\pi} \leq P_g^{\max} - p_{gt}, \qquad \forall g \in G, \pi \in \mathscr{P}(\mathrm{sp}), t \in T \tag{2.24}$$

$$\dot{r}_{gt}^{\pi} \leq \dot{R}_g^{\pi} u_{gt}, \qquad \forall g \in G, \pi \in \mathscr{P}(\mathrm{sp}), t \in T \tag{2.25}$$

$$\dot{r}_{gt}^{\pi} \leq \dot{R}_g^{\pi} \left(1 - u_{gt}\right), \qquad \forall g \in G, \pi \in \mathscr{P}(\mathrm{nsp}), t \in T, \tag{2.26}$$

where $\dot{r}_{gt}^{\pi}$ is the amount of reserve product π provided from generator g in period t. Each reserve product has a time component and a reserve type component (e.g., 10-minute spinning reserves). Spinning reserve products fall in the set $\mathscr{P}(\mathrm{sp})$ and non-spinning reserve products fall in the set $\mathscr{P}(\mathrm{nsp})$. Spinning reserves are limited by generator capacity (2.24) and ramping capability (2.25). Non-spinning reserves are subject to how quickly the off-line

[11] ISO-NE also adjusts the total 10-minute requirement for spinning and non-spinning reserves based on historical performance every three months using a *contingency reserve adjustment factor* (ISO-NE, 2011)

generators can start and ramp up their production (2.26). Other types of reserves (e.g, interruptible loads) can be also modeled using similar constructs. Reserve providers are assumed in this dissertation to follow all dispatching instructions. However, some resources may fail to perform when called upon and this can be mitigated by discounting the reserve availability from undependable resources (Weaver, 2012).

This dissertation tends to consider one reserve product (or combination of products) at a time. In most cases, the decision variable for available reserve is notated as r_{gt}. This is done to simplify the notation and to avoid restricting the formulation to too specific of applications. The context of the proposed reserve requirements will determine what type of reserve products are involved.

The availability of reserve is subject to network constraints. Reserve is said to be *deliverable* if it may be utilized without violating transmission constraints. Reserve deliverability is conditioned on the state of the system and the simultaneous re-dispatch of other units, i.e., transmission relief may require coordinated movement between multiple generators. Reserve requirements are enforced on a zonal basis to help ensure enough reserve is held within import-constrained regions (F. Wang & Hedman, 2012). Zones provide locational diversity to reduce the chance that congestion will impede deliverability. It is generally assumed that reserve is highly deliverable within a zone but may be undeliverable between zones. Figure 2.6 shows an example of reserve zones used by MISO and ISO-NE. This figure demonstrates the two types of reserve zones used in practice: mutually exclusive and nested. Nested zones usually encapsulate pockets of load in metropolitan

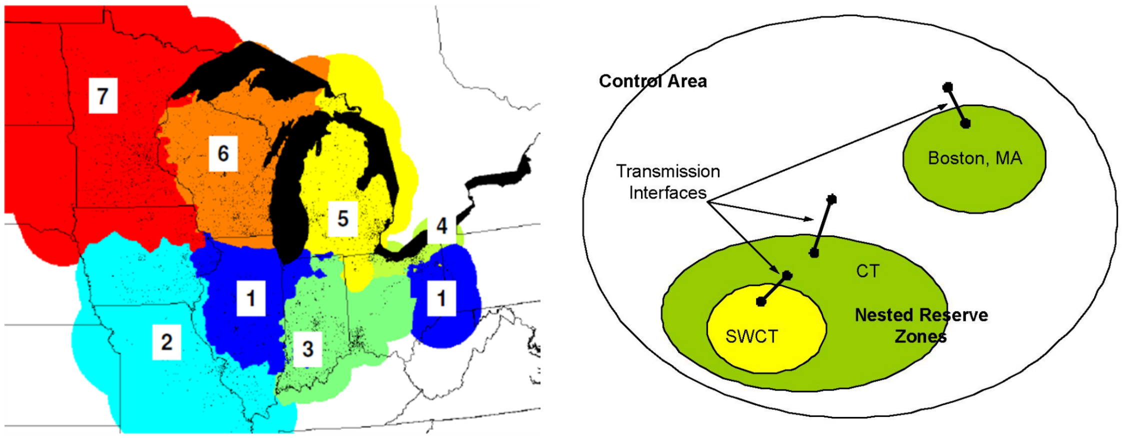


Figure 2.6: Reserve zones in MISO and ISO-NE (MISO, 2008; Cheung, Ma, & Sun, 2006). These are examples of mutually exclusive and nested zones. Note: areas “1” are controlled by PJM.

areas. These areas are typically import-constrained but can export reserve as needed. The reserve requirements for different regions are built upon these types of assumptions. The assumptions and limitations of reserve zones will be an ongoing theme of this dissertation.

2.3.2 Transmission Margins. Insufficient network transfer capabilities can hinder reserve deliverability. Verifying reserve deliverability is not trivial when there are many potential scenarios. Transfer limits are used to support deterministic scheduling models and to simplify coordination between control areas. They allow operators to schedule for the forecast and remain confident that there will be sufficient flexibility to respond to uncertainty. Transmission capability estimates formally include a transmission reliability margin (TRM) and a capacity benefit margin (CBM) (NERC, 1996). TRM helps ensure reserve deliverability within the control area and CBM improves deliverability between areas (Ou & Singh, 2002). Transmission margins are generally calculated prior to scheduling and may be overly conservative for some solutions and insufficient for others.

2.3.3 Scheduling for Reliability.

Sections 2.3.1 and 2.3.2 describe two types of flexibility, reserves and transmission margins, that help ensure reliability. Acquiring this flexibility generally comes at an opportunity cost. Operators can procure excessive reserves or transmission margins, but doing so may forgo economical solutions. Cost and reliability are often competing objectives and a solution may be overly-reliable if economics are sacrificed for the sake of meeting conservative flexibility targets. This tension emphasizes the importance of designing effective scheduling models.

Operators are responsible for ensuring N-1 reliability regardless of cost. This motivates the so-called security-constrained optimal power flow (SCOPF) and unit commitment (SCUC) problems. The simplest SCOPF and SCUC models include reserve requirements that explicitly require 10-minute reserve to exceed the largest contingency or some other, more conservative, amount. For example, PJM requires reserve to exceed about 1.5 times the largest contingency in each zone and WECC requires reserve to cover 6% of load plus 3% of exports (PJM, 2013; WECC, 2008). Reserve requirements are determined according to the balancing authority's preferences and needs.

Deterministic reserve requirements generally do not guarantee reliability because some reserve may be undeliverable. More precise models can consider how power flows change under uncertainty using stochastic programming or robust optimization. Such models can guarantee reliability (subject to the other model assumptions) and can do so economically. However, there is a computational gap between the deterministic algorithms of today and the stochastic algorithms of tomorrow. Stochastic approaches are not fully tractable because power scheduling applications are too large. Even today's deterministic models sometimes struggle to achieve small optimality gaps within the required time frame (Chen, X. Wang, & Wang, 2014). There is a practical need to develop more computationally tractable approaches that consider system conditions under uncertainty in a precise manner.

The literature review in Chapter 3 provides more details on alternative deterministic and stochastic unit commitment models. The review covers reserve requirements, reserve zones, transmission margins, stochastic programming, and robust optimization. The stochastic models frequently adopt two-stage formulations. An example of the first and second stages of a stochastic model that protects against contingencies is formulated in Appendix C.

2.3.4 Intermittent Renewable Resources. Less than 5% of nationwide production in 2011 came from renewables with intermittent properties, such as wind and solar (Energy Information Administration, 2013b). Various regions have established ambitious renewable energy portfolio standards that are spurring significant new investments. For example, the California standard requires renewables to serve 33% of load by the year 2020 (the nationwide average among all renewables was about 15% in 2011) (California Public Utilities Commission, 2013). A result of these standards is that uncertainty will increase because much of the gap will be filled by intermittent generation.

For contingencies, the cost of operating following re-dispatch is often ignored because it has negligible impacts on the total expected cost due to the low probability of an outage. More precisely, the cost of holding reserve at the ready has historically outweighed the expected difference in cost after the reserve has been dispatched. The upcoming advent of intermittent generation implies that more aggressive re-balancing will be necessary to correct random fluctuations in supply. As forecast errors increase, the deterministic scheduling algorithms used today may become less justifiable because the projected cost under forecasted conditions may poorly reflect the statistical expectation.

Furthermore, additional uncertainty from renewables will make N-1 reliability more difficult to ensure. Wind farms may be placed in isolated locations (ERCOT, 2007); therefore, fluctuations in wind power will be substituted by reserves that are injected from different locations in the network, causing congestion to deviate from the anticipated patterns. This has the potential to undermine zonal reserve models as it will be harder to predict the underlying congestion pattern.

2.3.5 Measures of Reliability. Operators have several risk measures at their disposal. Two common measures are the *loss of load probability* (LOLP) and *expected energy not served* (EENS) (Q. P. Zheng et al., 2014). LOLP represents the probability that any scenario will become so severe that load must be involuntarily curtailed in order to restore frequency and satisfy transmission constraints. EENS is the statistical expectation of load curtailment (quantity times probability summed over discrete scenarios or integrated over continuous scenarios). It is sometimes multiplied by the presumed *value of lost load* (VOLL) to estimate the expected cost of energy not served. Whenever an operator identifies a potentially unreliable scenario, it is implied that the schedule will be repaired whenever the cost of said repairs does not exceed the expected cost of energy not served.

2.4 Markets

This section provides a brief overview of important energy market considerations. This is not a comprehensive summary but an overview of concepts that are relevant to this dissertation. Market environments are more difficult to manage because most of the entities are competing with one another. ISOs are primarily responsible for maximizing social welfare while meeting established reliability standards (Kirschen & Strbac, 2004). They accept bids from producers and consumers and schedule to maximize the market surplus. These bids contain information about the cost to produce energy and the utility to consume energy.

ISOs attempt to simulate the ideal of perfect competition. With perfect competition, an "invisible hand" can be thought of influencing price, which, in turn, induces market participants to behave in a way that is globally optimal (Smith, 1869). Under the standard market design, ISOs attempt to directly maximize the market surplus using optimization and then determine market prices after the fact. Efficient prices ensure a local equilibrium where no participant wants to consume or produce differently from what was cleared (Fisher, 2006). An ideal settlement scheme will make good use of resources and provide an equilibrium where no participant has incentive to deviate from dispatch instructions.

Prices are often derived from the dual of the DCOPF [12] . For example, the dual variable of (2.10) determines the price of energy, which is called the locational marginal price (LMP). LMPs describe the sensitivity of the objective function (market surplus) to changes in supply or demand. Similarly, reserve prices are derived from the dual variables of the reserve requirements. Model imprecisions should be avoided because they can pervert market prices, e.g., by inflating prices for reserves that are incorrectly assumed to be deliverable (Lyon, F. Wang, et al., 2014). Model improvements not only stand to improve the market surplus but may also improve price signals.

The scope of market design extends far beyond what is discussed in this dissertation [13] . Changes to operating policies may result in unintended consequences, and operators often seek buy-in from market participants before implementing changes that may have noticeable impacts on market outcomes. Procedures that are expected to improve economics and reliability may be defeated if they are incongruous with the market design or unpalatable to market participants. This is one of the challenges currently faced by stochastic programming according to representatives from ISO-NE (J. Zhao, Zheng, Litvinov, & Zhao, 2013). The work in this dissertation, on the other hand, is compatible with traditional market

[12] In the case of unit commitment, prices may be derived by fixing the binary variables and solving the underlying DCOPF (O'Neill, Sotkiewicz, Hobbs, Rothkopf, & Stewart Jr., 2005). Other revisions may be made during the pricing run to manipulate market outcomes. For example, MISO recently obtained FERC approval to adopt a different pricing mechanism based on estimates of the convex hull of the feasible space (FERC, 2012b).

[13] A sample of important topics falling outside the scope of this dissertation includes: forward markets, bilateral trading, market power mitigation, non-confiscatory market policies, risk-hedging mechanisms, virtual bids, and constraint relaxation.

settlement strategies. This research develops methodologies that pose fewer challenges for practical implementation.

Chapter 3

LITERATURE REVIEW

Several approaches have been proposed to mitigate uncertainty in power systems. The earliest and most persistent concerns have been related to reliability. The consequences of curtailing customer loads are extreme, so operators take care to ensure continuous service remains available. Generally speaking, operators should be able to restore frequency (the balance between supply and demand) and satisfy transmission constraints for a prescribed set of scenarios. For example, all US control areas are required by NERC to survive the sudden outage a generator or transmission line (the N-1 requirement). The literature also proposes probabilistic criteria that emphasize protection against the most likely scenarios. Section 3.1 provides a review of traditional reserve requirements, which are primarily meant to manage reliability. Sections 3.1.3–3.2 review how reserve deliverability is encouraged in congested systems. These approaches are classified as deterministic in this dissertation because they do not model recourse decisions at a detailed level.

More recently, stochastic models have become popular in the research community. These models avoid the need to impose reserve requirements because they directly consider corrective actions. Stochastic models are more computationally expensive but can provide solutions that are more reliable and economical. Table 3.3 characterizes some of their benefits relative to deterministic models. Stochastic models can improve economics and

Table 3.1: Classification of the literature concerning reserve quantity and reserve deliverability. This dissertation* develops reserve disqualification and introduces congestion-based reserve requirements (RRs) to improve deliverability.

		Reserve quantity
		Probabilistic requirements Relax expensive requirements Iterative schemes
Deliverability	Reserve zones Transmission constraints Reserve disqualification*	Stochastic programming Robust optimization Congestion-based RRs*

reliability because they manage congestion within the scheduling algorithm itself (as opposed to deterministic models, where policies to mitigate congestion are usually established a priori). Stochastic models can also consider the cost of recourse decisions, which makes them more appealing when large disturbances are likely to occur. Section 3.3 reviews stochastic models with transmission-constrained recourse actions.

3.1 Reserve Requirements

Most of the reserve literature investigates reserve quantities while assuming transmission bottlenecks will not restrict corrective actions. NERC specifies that there must be enough 10-minute reserve to recover from the loss of any single generator or transmission line. Other requirements have been developed for alternative reliability targets. Probabilistic policies constrain risk measures based on loss of load probability (LOLP) or expected energy not served (EENS); a regular theme in the literature is the search for a risk threshold that balances cost and reliability. This review defines probabilistic models as those that *constrain* risk. Other models change reliability thresholds dynamically based on the cost of obtaining reserves, e.g., by minimizing a weighted sum of cost and reliability [1] .

Sections 3.1.1 and 3.1.2 describe probabilistic and cost-based reserve requirements while ignoring network constraints. Sections 3.1.3 and 3.1.4 locate reserves to help justify the assumption that reserve is deliverable.

3.1.1 Probabilistic Reserve Levels.

Probabilistic reserve requirements treat deviations in load and resource availability as random variables and procure reserve to satisfy a prescribed risk threshold. The general idea is to protect against only the most likely or severe scenarios. This approach adopts constraints that limit the cumulative probability of a supply shortage. However, one difficulty of this approach is that shortages are harder to anticipate when there are ramp constraints, e.g., when a generator operating 100 MW below its capacity can increase production by only 50 MW over a short time frame. Consider the example in Figure 3.1 describing how frequency changes for two contingencies 15 minutes apart. Holding enough 10-minute reserve to recover from both contingencies is unnecessary because the second contingency may use some 30-minute reserve. Designating a *single* reserve product to cover multiple uncertainties is not straightforward because disturbances may occur at different times. Nevertheless, approximate probabilistic reserve requirements abound in the literature, usually under the conservative assumption that disturbances occur simultaneously.

[1] Some authors describe models as probabilistic if they include probability terms in the objective. In this review, probabilistic only refers to models that *constrain* LOLP or EENS.

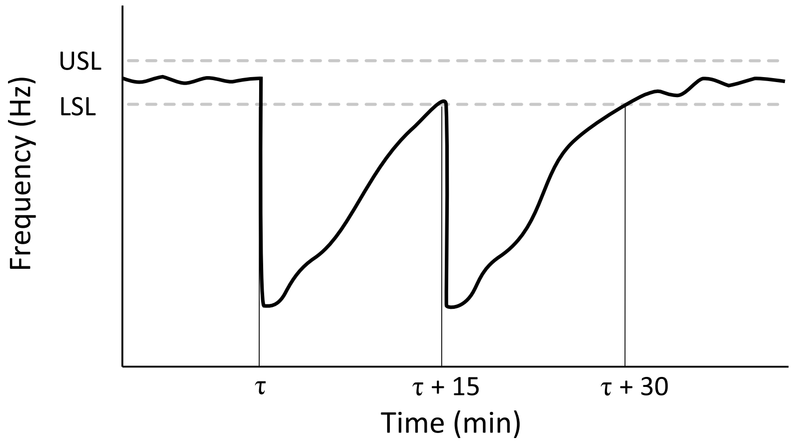


Figure 3.1: Frequency following contingencies that occur 15 minutes apart. USL and LSL are upper and lower specification limits that must be satisfied within 15 minutes of the event.

Early work by Anstine et al. (1963) requires committed capacity to exceed the forecast plus likely load deviations and generator failures. The model constrains the risk that load will exceed the committed capacity. The authors do not consider ramping restrictions; therefore, the LOLP may still be high if standby capacity cannot ramp fast enough to balance the system.

Dillon, Edwin, Kochs, and Taud (1978) suggest an iterative approach that solves unit commitment, evaluates LOLP, and then updates the unit commitment formulation based on the reliability analysis. Stopping to evaluate reliability allows the authors to calculate outage probabilities based on the most recent commitment solution. The authors perform contingency analysis using a stochastic process that models outages, the activation of "fast reserves", and replacement of fast reserves by "slow reserves". This type of contingency analysis has its limitations (e.g, it ignores network constraints), but the work does consider ramping between outages that are separated in time (attempting to address the phenomenon illustrated in Figure 3.1), an issue that is largely ignored in the probabilistic scheduling literature.

Several other iterative methods have been since proposed. Ozturk, Mazumdar, and Norman (2004) propose a way to update reserve requirements within iterative approaches when uncertainty is correlated between periods. The authors assume that load uncertainty follows a *multivariate* Gaussian distribution between periods, which cannot be inverted to determine standalone reserve requirements for each period. They address this issue by defining a lower bound for the minimum reserve quantity and updating the requirement between iterations to more closely align with the desired LOLP. The results demonstrate that considering the correlation of forecast deviations between periods may lead to better solutions (it usually the case that forecast deviations are autocorrelated).

Xia, Gooi, and Bai (2005) propose an iterative technique to drive EENS towards a target value. The authors suggest increasing reserve when EENS is high and reducing reserve when EENS is low. Apart from this paper, the convention is to bound the unreliability metric only from above [2] .

Gooi, Mendes, Bell, and Kirschen (1999) propose evaluating risk (LOLP or EENS) at each iteration of Lagrangian Relaxation and increasing the reserve requirement when risk exceeds the desired threshold. The algorithm can converge faster than other iterative techniques because it makes use of partial information before the solution converges. A

[2] In an ideal model, it is not necessary to bound risk from below because the objective function will push risk to its upper bound whenever reducing risk is costly. Targeting a certain risk could be perverse if low cost solutions are forgone in favor of solutions that are more costly *and* more risky.

backward pass is applied at the end to correct for excess reserves. H. Wu and Gooi (1999) extend the work to consider reserve ramping and hourly ramping.

It can be challenging to predict the outage probability distribution prior to knowing which generators will be committed. Bouffard and Franscisco D. Galiana (2004) handle this problem by introducing binary variables to identify combinations of failures that exceed available reserves. Doherty and M. O'Malley (2005) simplify things by estimating outage probabilities from historical commitments [3] . Chattopadhyay and Baldick (2002) dynamically estimate outage probabilities within unit commitment. The authors perform an off-line regression analysis that relates the cumulative outage distribution to a linear function of the committed generators. They then constrain the predicted LOLP against the approximate distribution. All of these approaches avoid iterative techniques by estimating LOLP within the mathematical scheduling model.

Some loads will contract with the system operator to provide contingency reserve. Aminifar et al. (2009) extend probabilistic requirements to consider the potential for non-compliance from such interruptible loads. The authors use binary variables to measure LOLP within the scheduling model, from which they derive EENS. Although the LOLP is inherently a function of binaries [4] , neither binary variables nor iterative techniques are actually necessary for estimating EENS. For example, Radi and Fox (1991) and Bai, Gooi, Xia, Strbac, and Venkatesh (2006) avoid binaries in their EENS estimations by adding

[3]Doherty and M. O'Malley (2005) lose precision by assuming outage probabilities are constant across the year. However, they wrote one of the few papers that considers generator contingencies and wind fluctuations in a unified model.

[4]For a single period, LOLP is the sum (or integral) of probabilities for scenarios with nonzero loss of load.

continuous variables anticipating the shortage for each scenario (these papers are reviewed in the Section 3.1.2).

All of the above approaches determine *how much* reserve to acquire. They do not consider the potential for transmission constraints to hinder reserve deliverability. Any can be combined with the congestion-based reserve requirements in Chapter 4 of this dissertation to support reserve deliverability.

3.1.2 Cost-Based Reserve Levels. It is common to factor cost into reserve provisions. The previous section outlined methods aimed at satisfying reliability thresholds such as a maximum LOLP or EENS. Another common approach is to allow the model to relax reserve requirements that have high shadow prices. These models dynamically change reliability targets based on the cost of reserve to avoid spending exorbitant amounts for small improvements to reliability. Generally, higher levels of reliability may be appropriate when reserve is inexpensive and vice versa.

Various ways to relax reserve requirements have been considered. The simplest approach is also the most commonly used in practice. Flynn, Sheridan, Dillan, and O'Malley (2001) propose relaxing requirements whose dual variables are large in magnitude (in a linearized version of unit commitment). This is accomplished by penalizing reserve shortages in the objective function, something that is now done by every ISO in the US under the term *constraint relaxation* (CAISO, 2009; ERCOT, 2010b; FERC, 2012a; MISO, 2012b; NYISO, 2013; SPP, 2013; T. Zheng & Litvinov, 2008). Constraint relaxation is compatible with all of the methods investigated in this dissertation. The idea is taken a step further by MISO and NYISO, who use a stepwise reserve demand curve where the presumed marginal value of reserve is inversely related to quantity (MISO, 2012b; NYISO, 2013). The theoretical justification for reserve demand curves is provided by Hogan (2009): the marginal value of reserve decreases with quantity because each additional MW protects against rarer and rarer scenarios. For example, it would be counter-productive to spend a premium protecting against N-3 when it is unlikely that three or more generators will simultaneously fail. The marginal value of reserve in each of these cases is estimated based on off-line studies.

A similar paradigm is followed by Tseng, Oren, Svoboda, and Johnson (1999), who analyze a two-player game theory model where high costs and market power are mitigated by relaxing reserve requirements. Other algorithms have been developed for solving unit commitment with fuzzy reserve requirements (Guan, Luh, & Prasannan, 1996; Chakraborty & Funabashi, 2010). The general theme in these papers is that it is difficult to pre-determine optimal reserve requirements and requirements ought to be relaxed when they become too restrictive.

Other researchers adopt a more classical multi-objective approach by minimizing a weighted sum of cost and unreliability. Radi and Fox (1991) add an estimate of EENS to the objective function. The authors model recourse decisions (without transmission constraints) in what can be considered an early application of two-stage stochastic programming for unit commitment. Bai et al. (2006) develop this approach by treating interruptible load as reserve that may fail to perform when called upon. The model considers the probability of non-performance from dispatchable loads and also proposes a settlement scheme that penalizes non-performance.

Finally, iterative schemes have also been suggested for balancing cost and reliability. O'Sullivan and M. J. O'Malley (1999) start with zero reserves and iteratively add reserves until there is no load shedding. The authors then perform a retrospective analysis on which reserve policy was the best. Ortega-Vazquez and Kirschen (2007) propose an approach that determines the reserve quantity for one period at a time. The algorithm solves much faster because it is not complicated by inter-temporal constraints. A line search technique is used when varying the reserve quantity to accelerate convergence and to avoid getting stuck at local optima. The selected hourly reserve requirements are then applied to a final run to determine the remaining unit commitment decisions.

Iterative techniques like the above have two desirable properties. First, they may be implemented concurrent to established scheduling practices that are already iterative. For example, Gooi et al. (1999) and H. Wu and Gooi (1999) update reserve requirements between iterations of Lagrangian Relaxation. Second, iterative approaches enable more in-depth reliability analyses that are able to account for more nuanced issues. For example, Ela and M. O'Malley (2012) suggest a frequency performance reliability metric and Abiri-Jahromi, Fotuhi-Firuzabad, and Abbasi (2007) propose an alternative "well-being" model – both of these metrics would be evaluated outside of the scheduling algorithm. Furthermore, a more precise contingency analysis can consider transmission constraints in order to validate the assumption that all reserves are deliverable. The work in Chapters 5 and 6 of this dissertation utilize an iterative approach to identify important transmission constraints for different scenarios.

3.1.3 Reserve Zones. Zones find various applications in power systems. They are a practical tool used to simplify decision problems. For example, energy zones have historically been used for power system scheduling. The simplified scheduling models only consider the flows on zonal interfaces. They assume that all locations within a zone have the same effects (PTDFs) on these interfaces (Alaywan, Wu, & Papalexopoulos, 2004; Daneshi & Srivastava, 2011). Zonal energy markets have now been abandoned by all ISOs in the US in favor of more precise nodal models (like the DCOPF), but zones are still ubiquitous for ancillary services such as reserves. This section reviews both energy and reserve zones because they are closely related (as reserve is nothing but power waiting to be dispatched).

Hong, Chang-Chien, Wu, and Yang (2002) and Imran and Bialek (2008) suggest that operators can determine energy zones by solving a full network model and then clustering nodes together that have similar energy prices (LMPs). The intuition is that locations with different prices tend to be separated by congestion. Furthermore, the market interpretation is reasonable because a single reserve price is typically derived for each zone. However, such approximations may be inefficient compared to a full network model. The results of Imran and Bialek (2008) suggest that zones are impractical for meshed congested systems, such as the European Interconnection. Similar concerns motivated the recent transition from zonal to nodal *energy* markets in CAISO and ERCOT (Alaywan et al., 2004; Daneshi & Srivastava, 2011). Figure 3.2a shows an example of energy zones in ERCOT. These operators would

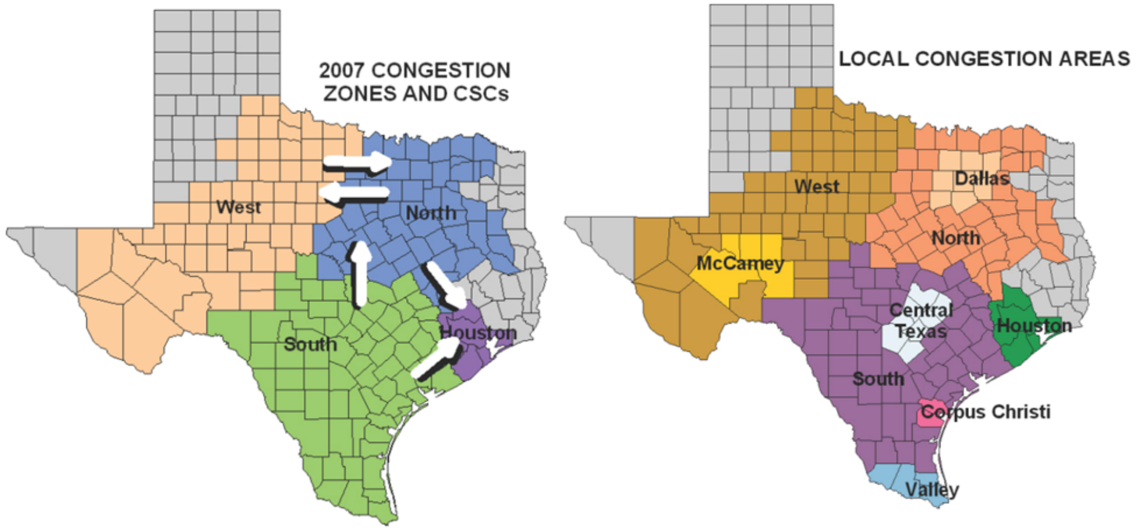


Figure 3.2: Energy and reserve zones in ERCOT, 2007 (ERCOT, 2007). The arrows indicate commercially significant constraints (CDCs), i.e., important transmission bottlenecks.

regularly observe power flow violations within zones and then repair the market solution using out-of-market corrections. CAISO and ERCOT collectively spent over $300 million on such corrections in 2007 (CAISO, 2007; ERCOT, 2007) [5] . The transition to nodal markets allowed for more precise identification of scarce resources and reduced the need for out-of-market corrections to repair infeasible or unreliable solutions.

Zones are not only used for short-term production scheduling. Blumsack, Hines, Patel, Barrows, and Sanchez (2009) propose the use of zones to simplify calculations for long-term planning studies. The authors define zones by clustering nodes based on electrical distance, which measures relative voltage and current sensitivities. The approach was developed for PJM to efficiently evaluate the congestion that could result from alternative transmission expansion plans. Cotilla-Sanchez, Hines, Barrows, Blumsack, and Patel (2013)

[5]The respective ISOs have a suite of tools available to manage congestion. These tools usually involve committing additional units or adjusting desired dispatch of different generators based on their location. A variety of techniques have been developed to make efficient interventions (Kirby & Dyke, 2002). Such interventions are classified as exceptional dispatches in CAISO and out-of-merit energy/capacity in ERCOT.

propose a slightly different electrical distance metric and partition the network based on multiple criteria using an evolutionary algorithm.

While energy zones address network constraints for the *forecasted* system state, reserve zones address network constraints under *uncertainty*. Unlike energy zones, reserve zones are ubiquitous in the US and in Europe. Zonal reserve requirements spread reserves across the system. ERCOT and MISO determine reserve zones by clustering nodes together that have similar shift factors for a prescribed set of critical paths (MISO, 2012a; ERCOT, 2010a). The main idea is to cluster nodes who may substitute for one another while preserving the flow on critical lines. This approach is formalized by F. Wang and Hedman (2012), who define a node-difference metric based on the relative shift factors (PTDFs) between nodes m and n as

$$d_{mn} = \sum_{l \in L} w_l \left| PTDF_{ml} - PTDF_{nl} \right|, \tag{3.1}$$

where w_l is a weight indicating the importance of transmission line l. A small d_{mn} indicates that nodes m and n have similar effects on highly weighted lines. Statistical clustering can be employed to group similar nodes together into zones. If large weights are assigned to heavily loaded lines, then the resulting zones tend to coincide with relatively congestion-free areas.

Reserve zones are typically updated on a seasonal basis, e.g., ERCOT updates zones every twelve months and MISO every three months (MISO, 2012a; ERCOT, 2010a). MISO has considered updating zones on a more frequent basis to adapt to changing transmission bottlenecks (Shields, Boughner, Jones, & Tackett, 2007); however, at present they will only update zones for severe conditions that cannot be resolved through other operating procedures (MISO, 2012a). F. Wang and Hedman (2014) propose a framework for updating zones on a daily or hourly basis to account for changing bottlenecks. Their approach is shown by Lyon, F. Wang, et al. (2014) to reduce costs and avoid clearing reserves that become undeliverable in real time.

Zonal reserve requirements can be expensive because they restrict the feasible dispatch solutions. Many operators recognize the economic benefit of merging reserve zones and have developed ways to account for reserve sharing when congestion allows. Reserve sharing models anticipate how much reserve can be delivered across zonal boundaries. For example, MISO determines 10-minute requirements two days ahead based on simulation of reserve deliverability between zones for different contingencies (MISO, 2012a). PJM and ISO-NE allow a portion of reserve to count towards the requirement of another zone. Their models adjust the reserve sharing estimates based on how the interface flows deviate from their anticipated values, e.g., the zonal reserve requirement may be relaxed if the import capability is estimated to be higher than expected (PJM, 2012; T. Zheng & Litvinov, 2008). Such dynamic models balance transmission and generation flexibility at the zonal level, but they may be sensitive to zone definition and they do not address intra-zonal congestion.

Reserve zones are currently an integral part of ensuring reliability in congested systems. They anticipate how congestion will limit reserve deliverability based on insights from simulation and historical experience. However, the unit commitment solution may still be unreliable if the model mischaracterizes reserve as deliverable. In such instances, operators will employ out-of-market corrections (OMCs) to repair unreliable solutions. One such OMC tool is reserve disqualification.

3.1.4 Reserve Disqualification.

Reserve disqualification is a tool used in industry to discount resources that cannot provide reserve. If some reserve is found to be undeliverable, MISO and ISO-NE will manually disqualify it so that requirements must be met by reserve at favorable locations moving forward (FERC, 2011; ISO-NE, 2012, 2013b) [6] . The benefit of these adjustments is they can be applied in a targeted manner when the reserve model assumptions are violated. The downside is they are often not co-optimized with other scheduling decisions and they are not reflected in day-ahead market prices (due to being applied after the market has cleared) (Chen, Gribik, & Gardner, 2014).

Reserve disqualification provides another means for operators to control reserve locations. The work in Chapter 5 of this dissertation introduces a mathematical framework for reserve disqualification. The new formalized procedure reduces reliance on manual interventions and supports the possible extension of reserve disqualification into day-ahead markets. When combined with reserve zones, these developments can help ensure reliability targets are met economically.

3.2 Transmission Constraints

Certain transmission constraints are regularly updated based on off-line analyses. For example, the flow across a path may be constrained below a prescribed transfer capability

[6] The exact term *reserve disqualification* is not used in industry. MISO designates undeliverable resources as *not qualified* and ISO-NE applies *reserve down flags* to these resources.

limit (NERC, 1996). These constraints protect against line violations during and after contingencies. The advantage is that off-line analyses can consider uncertainty, reactive power, stability, or other properties that are hard to include in the optimization model. The disadvantage is that transfer capability estimates may be sensitive to the system state, which is unknown at the time the estimates are calculated.

Transfer capability estimates often include margins that help ensure reserve is deliverable. For example, transmission reliability margins (TRM) and capacity benefit margins (CBM) allocate capacity for sharing reserve within and between areas (Ou & Singh, 2002). These margins are often determined empirically by simulating a large number of system states. Ou and Singh (2002) calculate TRM as a percentile of the difference between transfer capabilities *with* and *without* uncertainty. Audomvongseree and Yokoyama (2004) do similar but introduce a Markove model to simulate changing resource availabilities. bin Othman, Mohamed, and Hussain (2008) and Ramezani, Singh, and Haghifam (2009) use bootstrapping and scenario aggregation to reduce the sample size and, thereby, improve computational times for this off-line procedure. The transfer capabilities estimated using these schemes describe how flows may be restricted across a *single* path or interface to protect against uncertainty.

Papic, Vaiman, Vaiman, and Povolotskiy (2007) perform a more complex analysis that limits the relative flow on multiple lines. The operator defines a finite set of stressing patterns and evaluates how far each pattern can be extrapolated before violating transmission constraints. The feasible space for these scenarios is projected onto the flow across critical transmission paths. The multi-dimensional constraints derived from this analysis are called nomograms. A two-dimensional nomogram describes how much flow may be allowed on two transmission paths, where the maximum flow on one path depends on the flow on the other. Nomograms can be used to develop more general constraints than simple interface limits [7] . Since they are defined using simulation, nomograms may improve the model accuracy even when they have no obvious physical interpretation.

A fundamental challenge with estimating transfer capabilities is anticipating the locations of injections and withdrawals. Locations are important because more power can be transfered between points that have similar effects (shift factors) on critical lines. Purchala, Haesen, Meeus, and Belmans (2005) deal with the unknown-location issue by using a reduced network model. They collapse individual areas (or zones) into nodes before evaluating transfer capabilities. As shown in Figure 3.3, the reduced model includes one node for each area and a single line for each interface regardless of how areas are connected in the full network model. This reduced model is then used to identify a single set of shift factors shared by all locations in an area. The limitations are that these shift factors approximate inter-area flows and ignore intra-area congestion. These limitations are unavoidable when ignoring the injection locations within zones. The results of Purchala et al. (2005) demonstrate that it can be difficult to anticipate transfer capabilities without modeling the precise locations from where power will be injected. These issues speak to the difficulty of congestion management and the inherent limitations of zonal energy and reserve models.

[7] For example, nomograms can identify constraints of the form $a_1 f_1 + a_2 f_2 \leq b$, where $a_1 \neq a_2$.

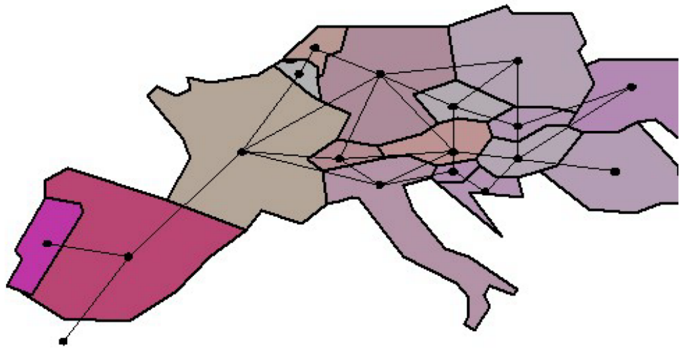


Figure 3.3: A reduced network model for the European interconnection (Purchala et al., 2005).

Transmission margins are rarely determined in conjunction with reserve requirements (e.g., they are not in the above references). However, one can envision how larger transmission margins would de-emphasize reserve location by leaving a bigger buffer on the residual network. The most economical solutions balance the cost of restricting reserve locations with the cost of leaving a transmission margin [8] . Stochastic models have shown promise, in part, because they can consider the cost of these trade-offs. The congestion-based reserve requirements in Chapter 4 of this dissertation also encourage the model to leave larger transmission margins when it is inexpensive to do so. They may, therefore, be classified as determining dynamic transmission margins as opposed to the static margins from the literature.

3.3 Stochastic Models with Transmission-Constrained Recourse

Stochastic models have been applied extensively to unit commitment. This section reviews those that include transmission constraints on recourse decisions. The key benefit

[8]The reserve sharing models of PJM (2012) and T. Zheng and Litvinov (2008) accomplish something similar on a zonal level by coupling reserve requirements with the flow across zonal interfaces.

is a locational awareness regarding the value of reserve. Historically, scheduling models became more precise when OPF replaced energy zones. In the future, a similar transition may see stochastic models replace reserve zones in order to address uncertainty more directly.

This review covers two types of models: stochastic programming and robust optimization. The stochastic programs minimize expected cost over a set of scenarios. The robust models minimize the worst cost over a convex uncertainty set. This type of robust optimization can be less computationally intensive than stochastic programming (Bertsimas, Brown, & Caramanis, 2011). It is being considered by a major industry software provider (Alstom Grid) as a tool to commit additional generators for reliability after the market has cleared (Q. Wang, X. Wang, Cheung, Guan, & Bresler, 2013), but there is no sign of it yet being proposed for day-ahead markets.

3.3.1 Stochastic Programming. Stochastic unit commitment has become more popular in the literature over the past 10 years. Francisco D. Galiana et al. (2005) propose integrating contingency recourse decisions into unit commitment. They model an explicit OPF for each scenario, causing the number of variables and constraints to increase dramatically. The authors derive some market impacts by unbundling nodal energy prices into pre- and post-contingency components. Bouffard, Galiana, and Conejo (2005) demonstrate the model's potential to identify better unit commitment solutions for various test cases.

Some related formulations ignore recourse costs and instead focused on feasibility across contingency scenarios. Hedman, Ferris, O'Neill, Fisher, and Oren (2010) show that economics for this problem can be improved by allowing changes to the network topology. All of the above authors note that computational times can be considerable for stochastic programming. J. Wang, Shahidehpour, and Li (2009) propose to make the problem more scalable via Benders' Decomposition, but the solve time for the IEEE 118-bus test case still exceeds one hour despite them considering only three contingencies (one generator and two transmission outages). Ferris, Liu, and Zhao (2014) develop ways to improve convergence that *may* make Benders' Decomposition practical in the near future; however, they approach a simpler version of the problem with only one period and no binary variables. Hejazi, Mohabati, Hosseinian, and Abedi (2011) improve computational times by using an evolutionary algorithm. Their biggest computational bottleneck is evaluating reliability at each iteration, but this problem is naturally parallelizable. Nevertheless, the practical scalability of exact solution approaches remains unproven for unit commitment problems with transmission-constrained recourse.

Stochastic programming has also been applied for continuous forms of uncertainty, such as load and wind. J. Wang, Shahidehpour, and Li (2008) apply Benders' Decomposition to a two-stage model with wind uncertainty. Morales, Conejo, and Pérez-Ruiz (2009) evaluate the sensitivity of expected operating cost to wind variability and uncertainty. One of their main conclusions is that transmission constraints can induce wind curtailments to help manage congestion. Their results support the claim that congestion will be more difficult to efficiently manage in the future as wind farms become more prevalent.

L. Wu, Shahidehpour, and Li (2007) propose minimizing expected cost across a scenario tree spanning multiple time steps. The authors use Monte Carlo simulation to generate random scenarios that diverge from a common starting point. They enforce consistent decisions in the parent nodes via coupling constraints. The authors solve the problem using dual decomposition where the coupling constraints are dualized. The different scenario paths are solved in parallel and the Lagrangian multipliers updated at each iteration. The same authors tweak the scenario tree in a later paper and generalize the model to allow for load shedding (L. Wu, Shahidehpour, & Li, 2008). A similar solution approach is used by Papavasiliou and Oren (2013) on a two-stage stochastic program. The authors emphasize the potential for parallelization and test a large number of scenarios in parallel using high performance computing (Papavasiliou & Oren, 2013; Papavasiliou, Oren, & Rountree, 2014). Even with parallelization, the simplified test case (a 225-node representation of CAISO) still takes hours to solve due to the number of iterations necessary to reach convergence. However, all of the above dual decomposition techniques reviewed in this paragraph are highly scalable with respect to the number of scenarios due to the potential for parallel computing.

3.3.2 Robust Optimization. Robust optimization has recently gained interest as an alternative to stochastic programming that can be less computationally intensive (Bertsimas et al., 2011). Robust optimization is applied extensively to the terrorist interdiction problem. In this problem, a malicious entity causes up to *k* resources to simultaneously fail and operators respond to minimize the effect of the attack. Robust optimization follows a max-min structure that identifies which interdiction causes the worst best-case response, i.e., the attack that operators struggle the most to deal with. If the inner problem is convex, the dual can be taken so that the overall problem becomes a single-level maximization problem. The resulting problem is bilinear and non-convex but can be linearized via disjunctive constraints using binary variables. This insight guides a portion of the literature interested in how well a system is prepared for malicious interdiction (Salmeron, Wood, & Baldick, 2004; Motto, Arroyo, & Galiana, 2005; Salmeron, Wood, & Baldick, 2009; L. Zhao & Zeng, 2012a).

Robust optimization has also been developed as an alternative to stochastic programming for scheduling problems with recourse. These models view nature as an adversary that tries to maximize operating costs. The models have a tri-level structure, where the inner two levels are analogous to the interdiction problem described above. The outer level determines preventative actions made prior to nature selecting a scenario from the uncertainty set. Decomposition algorithms have been developed to determine the unit commitments that optimizes the worst best-case response. For example, Q. Wang, Jean-Paul, and Guan (2013) apply Benders' Decomposition to unit commitment in order to achieve N-k reliability, i.e., to protect against any combination of k simultaneous contingencies. At each iteration, robust optimization is used to identify the worst set of contingencies and then a Benders' cut is applied to the master unit commitment problem. Due to the difficulty of the problem, the approach struggles computationally when modeling more than one contingency $(k > 1)$ because the subproblem takes a long time to solve.

Bertsimas, Litvinov, Sun, Zhao, and Zheng (2013) and Jiang, Zhang, Li, and Guan (2014) use Benders' Decomposition to minimize the worst cost over a continuous uncertainty set. The authors note that the inner problem is NP-hard and develop heuristics to quickly identify reasonable solutions. Jiang, Wang, Li, and Guan (2012) extend the model to allow surplus wind to be curtailed (this can be beneficial if excess wind is causing congestion or exacerbating other constraints). The above authors raise concerns about the tendency of robust optimization to identify solutions that have high expected cost (this is called the cost of robustness (Bertsimas & Sim, 2004)). To alleviate this concern, they advocate reducing the size of the uncertainty set to obtain solutions that are less conservative. This strategy allows operators who are more risk neutral to still utilize robust optimization, which frequently solves faster than stochastic programs.

Conventional robust optimization requires the recourse problem to be convex (in order to take the dual of the inner problem). This limits the modeling precision for systems with generators that may be turned on or off during scenario response. Such *fast* generators can greatly improve expected costs in systems with a lot of wind (J. Xiao, Hodge, Pekny, & Reklaitis, 2011). Their influence can be approximated by relaxing integrality constraints (as is done by J. Xiao et al. (2011)), but this may lead to optimistic solutions that are not technically feasible. An exact algorithm is proposed by L. Zhao and Zeng (2012b) to solve robust optimization problems using column generation. The method can model integer recourse decisions because it uses no dual information, but the authors warn that advances are needed before solving large problem instances in "reasonable" time (L. Zhao & Zeng, 2012b). The efficient consideration of binary recourse variables is one benefit provided by the deterministic policies developed in this dissertation.

3.3.3 Hybrid Deterministic Methods. Pure stochastic programming and robust optimization models do not address *all* uncertainties in a computationally tractable way. The already difficult problems in Sections 3.3.1 and 3.3.2 tend to focus on either discrete contingencies or continuous uncertainties, implying that the remaining uncertainties are negligible or addressed by other policies. This is important because deterministic policies may be applied in tandem with the state-of-the-art stochastic models from the literature.

There is also potential for hybrid methods. Ruiz, Philbrick, Zak, Cheung, and Sauer (2009) demonstrate that some benefits of stochastic programming can be retained while limiting the size of the uncertainty set by using deterministic reserve requirements to cover scenarios that are not given stochastic treatment. Figure 3.4 summarizes how improving reserve requirements can make stochastic models more scalable. There may come a time when computational advancements allow stochastic models to adequately address all uncertainty and render deterministic requirements obsolete. However, for the immediate future, reserve requirements will continue to fill the gap that cannot yet be fully addressed by stochastic programming or robust optimization.

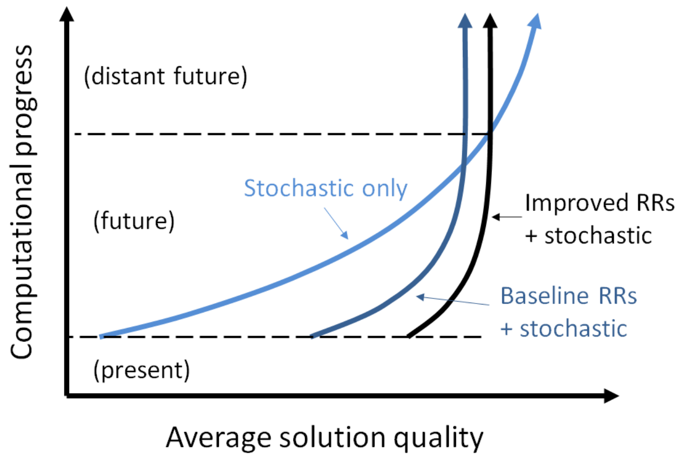


Figure 3.4: A qualitative description of how reserve requirements (RRs) support partial application of stochastic models.

3.4 Conclusion

Reliability in power systems is a matter of operating so that the system can survive disturbances. Survival requires the ability to i) balance supply and demand and ii) satisfy transmission constraints. Both tasks are critical for recourse when responding to disturbances. However, scheduling policies are often designed by considering only one task at a time: many determine the desired reserve quantity while assuming all reserves are deliverable and others address transmission constraints without considering the ability to relocate reserves. Reserve zones help control the location of reserve but are based on pre-determined (and infrequently updated) guesses as to how transmission constraints will limit recourse decisions. Less approximate models are provided by stochastic programming and robust optimization [9] but are not yet scalable enough for practical systems. There is a wide

[9] There are other types of stochastic models not covered in this review, such as chance-constrained, conditional value at risk, and minimax regret formulations (Q. P. Zheng et al., 2014). These models currently experience similar computational barriers to implementation (Q. P. Zheng et al., 2014).

computational gap between the deterministic policies being used today and the stochastic policies that are proposed for tomorrow. The research in this dissertation begins to fill that gap by developing policies that anticipate recourse decisions without solving a full stochastic model. The new policies are developed especially with transmission constraints in mind, although the work in Chapter 6 works to avoid expensive recourse decisions in general.

Chapter 4

CONGESTION-BASED RESERVE REQUIREMENTS

4.1 Introduction

Traditional reserve requirements are designed to ensure reserves are deliverable between zones. They are *not* designed to mitigate local congestion within zones. This type of congestion can be a major concern for some operators because SCUC may return an unreliable solution. The process of repairing unreliable solutions can be expensive and nontransparent to market participants (Lyon, F. Wang, et al., 2014). Effective alternatives (e.g., stochastic programming) have been delayed by computational limitations, so it can be beneficial to develop practical methodologies that improve reserve deliverability. This chapter describes a dynamic approach for addressing intra-zonal reserve deliverability for contingency response.

Current practices employ static reserve requirements to ensure N-1 reliability. These requirements are largely determined on an ad hoc basis. For example, PJM requires reserve to exceed about 1.5 times the largest contingency and WECC requires reserve to cover 6% of load plus 3% of energy exports (PJM, 2013; WECC, 2008). It is typical for reserve requirements to exceed the largest contingency by some margin: these margins tend to provide greater protection (beyond N-1) and increase the likelihood that enough reserve will be deliverable. This research addresses the issue of reserve deliverability by predicting how much reserve should be procured in order to mitigate intra-zonal congestion.

The proposed model works by linking reserve requirements to a measure of transmission stress. Figure 4.1 summarizes its relationship with other approaches from the literature. Deterministic models typically use flexibility requirements that withhold generation reserve and transmission capacity for scenario response. The proposed model integrates these types of flexibility such that less reserve is required when the network is unstressed. Conversely, the model requires more reserve as transmission stress increases so that there will be more dispatching freedom to alleviate congestion. The model performs well computationally compared to stochastic programming because it anticipates recourse decisions using a simple affine function. Although this approach sacrifices precision (compared to stochastic programs), it allows the unit commitment problem to remain scalable for large power systems because the number of additional variables and constraints are minimal.

The proposed model mitigates congestion by procuring a reserve margin. This margin can be integrated with any model from the literature [1] to help justify the assumption that all reserves are deliverable. The model is assessed in this chapter relative to cost and reliability, which are competing objectives because inexpensive solutions are often less reliable. Testing demonstrates that the proposed model provides solutions that are generally Pareto efficient and sometimes Pareto dominant over established reserve policies. Of

[1] Congestion-based reserve requirements can be used with the literature reviewed in Sections 3.1.1 and 3.1.2.

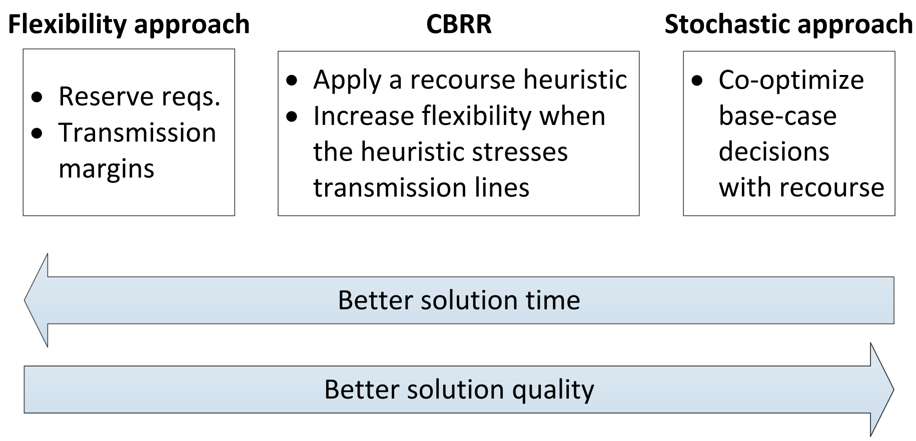


Figure 4.1: A comparison of congestion-based reserve requirements (CBRR) to traditional deterministic and stochastic models.

particular note is the model's ability to find economical solutions by increasing transmission flexibility when reserve is expensive.

The remainder of this chapter is organized as follows. Section 4.2 provides background on intra-zonal congestion. This motivates the problem by giving some indication as to the scale of intra-zonal congestion. Section 4.3 discusses the problem formulation. Section 4.4 describes how to strengthen the formulation by anticipating recourse decisions using an affine function of uncertainty. Section 4.5 suggests an interpretation and construction for the user-defined parameters. Section 4.6 illustrates the approach with a small example and Section 4.7 presents numerical results on the IEEE 73-bus test case. Finally, Section 4.8 concludes the chapter.

4.2 Extent of Intra-Zonal Congestion

To understand the prevalence of intra-zonal congestion, it helps to look at lessons learned from zonal energy markets. Until recently, both CAISO and ERCOT used zonal models to clear not only reserves but energy as well. Their scheduling models would treat resources in the same zone equivalent from a transmission perspective. Model solutions would be checked for feasibility and repaired as needed through a process they called congestion management (Kirby & Dyke, 2002; CAISO, 2007; ERCOT, 2010a). The most common congestion management tools involve turning on additional generators and changing the output of committed generators to bring power flows to within feasible limits. Another term for these adjustments is out-of-market corrections (OMCs).

CAISO and ERCOT both struggled to manage congestion in their zonal markets. Figure 4.2 shows the increase to intra-zonal congestion in CAISO from year 2005 and 2007 and Figure 4.3 shows congestion management expenses in ERCOT from 2001 to 2007. Both systems experienced growing congestion leading to more extensive out-of-market corrections to ensure feasibility and reliability. Eventually CAISO and ERCOT abandoned zonal energy markets in favor of nodal markets that use full network representations like the DCOPF discussed in Section 2.2.1 (Alaywan et al., 2004; Daneshi & Srivastava, 2011). Nodal models are more precise because they discriminate between locations in a zone.

No such transition has occurred away from zonal reserve requirements. Scheduling models still assume reserve is deliverable despite the existence of intra-zonal congestion and its potential to grow in the future as intermittent resource, such as wind, become more prevalent. Reserve deliverability is often improved in MISO by repairing the SCUC solution using out-of-market corrections [2] . MISO dislikes the need for these corrections because

[2] Reserve disqualification is the primary type of out-of-market correction MISO uses to improve reserve deliverability (FERC, 2011). This type of strategy is covered in detail in Chapter 5.

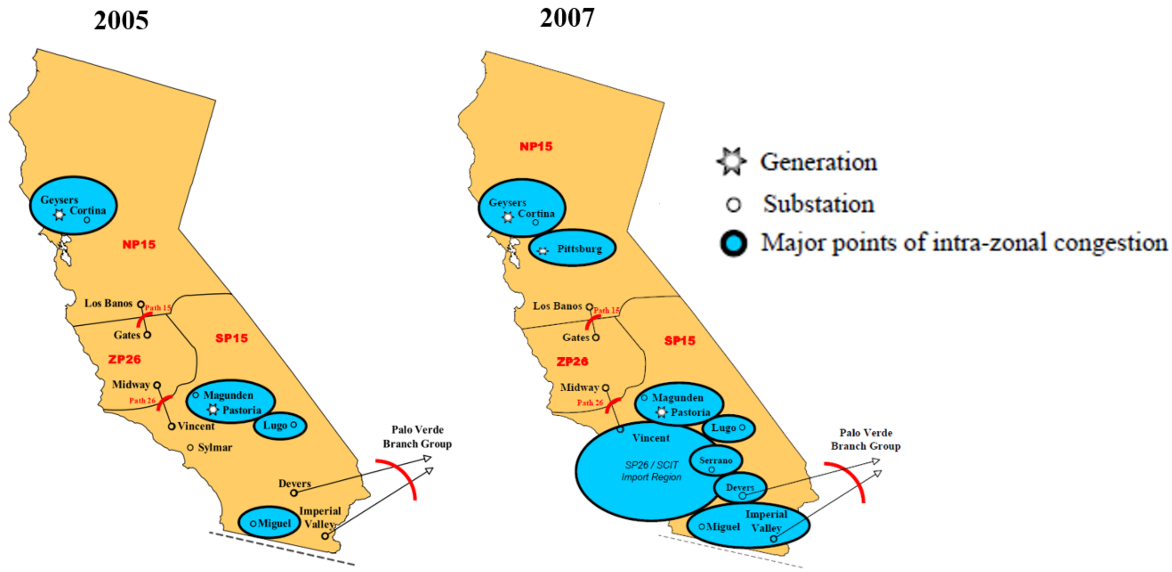


Figure 4.2: Major points of intra-zonal congestion in CAISO in 2005 and 2007 (CAISO, 2006, 2007). The black lines are zone boundaries and the blue ellipses are points of intra-zonal congestion.

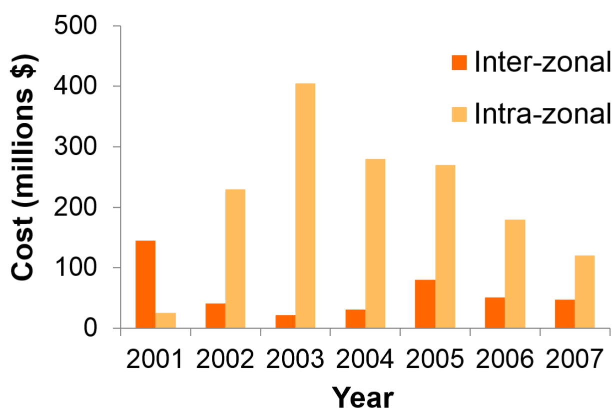


Figure 4.3: Congestion management expenses in ERCOT (ERCOT, 2007). This is the money spent repairing infeasible and unreliable market solutions. A map of the major intra-zonal congestion areas in 2007 is in Figure 3.2.

they can be expensive and they are applied after the day-ahead market has cleared (Chen, Gribik, & Gardner, 2014). The proposed model provides a dynamic approach to help ensure reserve deliverability so that SCUC can identify inexpensive solutions that require fewer out-of-market corrections.

4.3 Problem Formulation

The proposed reserve requirements build upon the DC unit commitment formulation outlined in Appendix C.1. This work develops a reserve margin η_t that can be added to any baseline requirement. For simplicity, a single-zone model is used where reserve must exceed the largest contingency plus the reserve margin η_t:

$$\sum_{i \in G \setminus g} r_{it} \geq p_{gt} + \eta_t, \qquad \forall g \in G, t \in T. \tag{4.1}$$

Prices are often derived by fixing all binary variables and solving the dual of the resulting linear program. Treating η_t as a variable imposes no change for this type of pricing structure.

4.3.1 Basic Transmission Stress Model. This section starts by measuring transmission stress as a linear function of line utilizations. The minimum reserve margin is defined by

$$\eta_t = \alpha \sum_{l \in L} \mu_{lt}, \qquad \forall t \in T \tag{4.2}$$

$$\mu_{lt} \geq \frac{f_{lt}}{F_l}, \qquad \forall l \in L, t \in T \tag{4.3}$$

$$\mu_{lt} \geq -\frac{f_{lt}}{F_l}, \qquad \forall l \in L, t \in T, \tag{4.4}$$

where f_{lt} and μ_{lt} represent the real power flow and utilization of line l during period t. The parameter α controls how much additional reserve is required as μ_{lt} changes. Larger values for α are always more conservative, i.e., put higher weight on reliability. Inequalities (4.3) and (4.4) actively measure the utilization of lines that have positive or negative flow (with respect to an arbitrary reference direction). The reserve margin η_t is unnecessarily high whenever μ_{lt} exceeds its lower bounds. Therefore, at optimality, μ_{lt} equals utilization whenever the dual variable of (4.1) in the linear relaxation is positive (otherwise the value is arbitrary, which can happen when reserve exceeds the minimum requirement).

This formulation provides an incentive to reduce power flows in order to relax reserve requirements. In particular, this occurs when the cost of increasing reserve exceeds that of adjusting the solution to reduce the required reserve margin η_t. Security-constrained unit commitment (SCUC) may have several low cost solutions with very different commitments (Sioshansi, O'Neill, & Oren, 2008) and, therefore, potentially different flow patterns. Equations (4.2)–(4.4) afford the opportunity to change the committed set of generators to significantly reduce transmission stress at relatively little cost.

The model allows substitutions between reserve and transmission stress in order to reduce *cost*. To work effectively, it is necessary for increments in reserve and decrements in transmission stress to have similar effects on *reliability*. Section 4.3.2 addresses how this relationship can be regulated more precisely.

4.3.2 General Transmission Stress Model.

The proposed approach motivates a careful look at how transmission stress affects reserve deliverability and how much of a reserve margin should be acquired in response. Consider generalizing (4.2)–(4.4) by

$$\eta_t = \alpha \sum_{l \in L} \sum_{s \in H} \varphi^s \tilde{\mu}^s_{lt}, \qquad \forall t \in T \tag{4.5}$$

$$\sum_{s \in H} \tilde{\mu}^s_{lt} \geq \frac{f_{lt} + \delta^+_{lt}}{\tilde{F}^{10}_l}, \qquad \forall l \in L, t \in T \tag{4.6}$$

$$\sum_{s \in H} \tilde{\mu}^s_{lt} \geq -\frac{f_{lt} + \delta^-_{lt}}{\tilde{F}^{10}_l}, \qquad \forall l \in L, t \in T \tag{4.7}$$

$$0 \leq \tilde{\mu}^s_{lt} \leq \omega^s, \qquad \forall s \in H, l \in L, t \in T. \tag{4.8}$$

Two new features are included in (4.5)–(4.8). First, the model anticipates how flows will change following a contingency. The term δ^+_{lt} (δ^-_{lt}) anticipates the greatest increase (decrease) of flow on line l in period t, and it is not necessary that $\left|\delta^+_{lt}\right| = \left|\delta^-_{lt}\right|$. The anticipated flows are compared to emergency ratings $\tilde{F}^{10}$, which are enforced after a contingency occurs. The second formulation change is the adoption of a convex piecewise linear transmission stress function that has slope φ^s for each segment s of width ω^s. These changes are discussed below in more detail.

The generalized model moves from measuring transmission stress in the base-case to anticipating transmission stress after a contingency. This change matters because critical lines may be relatively unutilized prior to a contingency. For example, changing flows pose enough concern for operators to simulate contingencies when estimating transfer capabilities (NERC, 1996). Likewise, δ represents the largest anticipated change in flow following contingency re-dispatch. An intelligent way to estimate δ is the subject of Section 4.4.

The convex stress function discounts lines that are unlikely to impede reserve deliverability. The term $\sum_{s \in H} \tilde{\mu}^s_{lt}$ represents the maximum predicted utilization of line l after re-dispatch for any contingency in period t. The model allows a different slope φ^s for each segment s and, at optimality, segments are used in the desired order because the reserve margin is unnecessarily high whenever segments are not filled in ascending order of slopes.

Similar transmission stress functions are used in the literature to measure how different contingencies are likely to stress the network. F. Xiao and J. D. McCalley (2009) propose a risk-based optimal power flow that replaces hard transmission limits with a convex risk function. Also related to risk, Wood and Wollenberg (1996) discuss an overload performance index (PI) as a relative measure of how severe a contingency c is likely to be:

$$PI_t^c = \sum_{l \in L} \left(\frac{f_{lt}^c}{\tilde{F}_l^{10}} \right)^{2n}, \tag{4.9}$$

where f_{lt}^c is the post-contingency line flow and n is a parameter that controls the relationship between line utilizations and PI. A graphical comparison between stress functions is shown in Figure 4.4; although they take similar forms, a distinguishing assumption in this work is that post-contingency transmission stress can be relieved by acquiring additional reserve. Section 4.5 provides a detailed interpretation of transmission stress in this context.

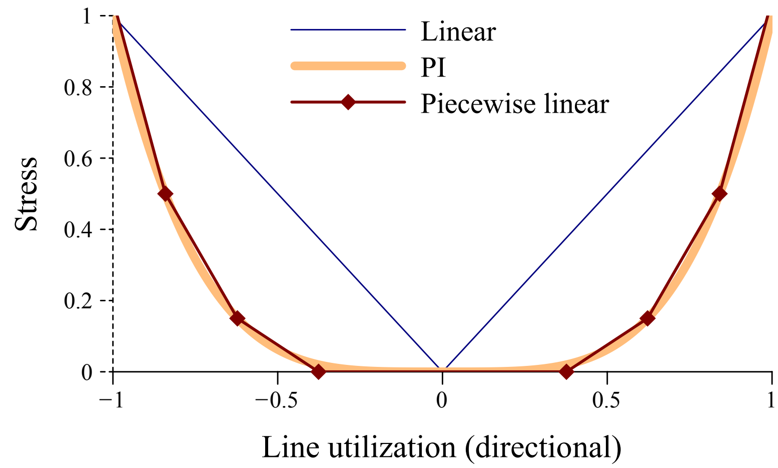


Figure 4.4: Transmission stress functions for a single line.

4.4 Anticipating Re-dispatch

The term δ (including δ_{lt}^{+} and δ_{lt}^{-}) in (4.6) and (4.7) represents the largest anticipated change in flow following contingency re-dispatch. Prudence dictates that δ should not only be practical but justifiable in order to appease market participants. This section describes an efficient approach that can base δ on operator insight or past contingency analyses, which are often readily available from past days and as the schedule is continually revised. The approach uses an affine policy based on *operational shift factors*.

Generation shift factors (GSFs) and line outage distribution factors (LODFs) are often used to calculate changes in flow following a contingency (Wood & Wollenberg, 1996). They describe how power flows change relative to the pre-contingency usage of the failed resource. This section borrows the same concept but estimates GSFs and LODFs with respect to anticipated contingency recourse actions. Estimating operational shift factors requires the specification of a response function that anticipates how generators will be re-dispatched. An affine function is used here. The response of generator i to a particular contingency in period t can be summarized by

$$\rho_{it}^{g} = \frac{\Delta p_{it}^{g}}{p_{gt}} \tag{4.10}$$

$$\rho_{it}^{k} = \frac{\Delta p_{it}^{k}}{f_{kt}}, \tag{4.11}$$

where Δp_{it}^{x} is the change in output of generator i following the failure of resource x and ρ_{it}^{x} is a normalized participation factor. Note that $\sum_{i \in G} \rho_{it}^{x} = 0$ whenever frequency (the balance between supply and demand) is restored. A precise evaluation of ρ can be calculated by stochastic programming at significant computational cost. A more computationally efficient approach is to estimate participation factors ahead of time and use them to generate operational GSFs and LODFs as follows:

$$a_{lt}^{g} = \sum_{i \in G} \rho_{it}^{g} PTDF_{il} \tag{4.12}$$

$$d_{lt}^{k} = LODF_{l}^{k} + \sum_{i \in G} \rho_{it}^{k} PTDF_{il}^{k}, \tag{4.13}$$

where $PTDF_{gl}$ (the power transfer distribution factor) is the sensitivity of flow on line l to injection by generator g, $LODF_{l}^{k}$ describes how flow changes on l immediately after line k fails, and $PTDF_{gl}^{k}$ is the power transfer distribution factor when k is not in service. These distribution factors are easy to calculate and remain the same as the operating state changes (M. Liu & Gross, 2004). They are combined above with the participation factors ρ to produce operational shift factors. The operational GSF a_{lt}^{g} is the anticipated change of flow on l per unit power produced by failed generator g, and the operational LODF d_{lt}^{k} is the anticipated change of flow on l per unit flow on failed line k. Once calculated, the shift factors can be embedded into SCUC as bounds on δ:

$$\delta_{lt}^{+} \geq p_{gt} a_{lt}^{g}, \qquad \forall g \in G, l \in L, t \in T \tag{4.14}$$

$$\delta_{lt}^{-} \leq p_{gt} a_{lt}^{g}, \qquad \forall g \in G, l \in L, t \in T \tag{4.15}$$

$$\delta_{lt}^{+} \geq f_{kt} d_{lt}^{k}, \qquad \forall k \in L, l \in L \setminus k, t \in T \tag{4.16}$$

$$\delta_{lt}^{-} \leq f_{kt} d_{lt}^{k}, \qquad \forall k \in L, l \in L \setminus k, t \in T. \tag{4.17}$$

Equations (4.14)–(4.17) contribute two constraints for each line and contingency (the number of constraints may be reduced by omitting contingency scenarios that are unlikely to threaten reliability). The use of shift factors facilitates a dynamic update of δ that incentives less utilization of a resource if its failure is likely to cause an unreliable condition. If utilization is not reduced, then more reserve will be required.

The above formulation avoids the computational intractability of more precise stochastic models by assuming that generators will be re-dispatched in proportion to the assumed participation factors, i.e., reserve will come from the same places. These factors are likely to be imprecise for at least one contingency, probably several, because reserve locations and transmission bottlenecks may change. The proposed approach accounts for this imprecision by procuring additional reserve as the predicted flows approach or exceed emergency limits in anticipation that the additional reserve will allow operators to find a re-dispatch solution that avoids transmission violations. The desired level of conservatism is achieved by tuning the user-defined parameters.

4.4.1 Alternatives for Anticipating Recourse.

4.4.1 Alternatives for Anticipating Recourse. Severe contingencies can necessitate quick response from operators to protect the system. In such situations, it can be beneficial to initiate contingency response immediately according to pre-defined rules. Such a scheme is used by MISO to immediately start deploying reserve according to participation factors while operators await an optimal dispatching solution (Chen, Gribik, & Gardner, 2014). The existence of these rules may make it easier to predict transmission stress because recourse decisions follow pre-defined policies [3] .

A more general approach to anticipate recourse decisions is explored by Warrington, Goulart, Mariéthoz, and Morari (2013). The authors develop affine policies for scenario response similar to the participation factors used above. They caution against using this approach for day-ahead scheduling because the system state is more uncertain. However, imperfect predictions may still be beneficial when used in conjunction with congestion-based reserve requirements. The anticipated recourse actions may be less accurate than stochastic programs and still do a reasonable job a) predicting where congestion is a concern and b) acquiring more reserves for additional protection in these instances. The model approaches stochastic programming when the anticipated actions align with the optimal scenario response. Until that level of accuracy is reached, the model must tuned by user-defined parameters.

[3] Independent from this dissertation, MISO published research that uses their knowledge of the participation factors to constrain post-contingency flows (Chen, Gribik, & Gardner, 2014). Their work shares the idea of using participation factors to influence security constraints, but there are also significant differences. One difference is that Chen, Gribik, and Gardner (2014) do not tie the transmission constraints to a reserve margin. Another difference is that they aggregate nodes within zones and, in doing so, lose the ability to address intra-zonal congestion. Nonetheless, they observe favorable market results when testing their approach, which underscores the potential benefit of using participation factors to anticipate recourse actions.

4.5 Interpretation of Parameters

The proposed method relies on user-defined parameters that, like other ad hoc policies used in practice, should be tuned to the particular system. Understanding the role of these parameters is relevant to build an effective model.

Let μ_{lt} be the predicted post-contingency utilization based on the assumed participation factors. An increase above $\mu_{lt} = 1$ warrants a larger reserve requirement to help operators reduce flow on the overloaded line. Properly specified models anticipate how much reserve is needed to counter transmission stress. Consider the following stress function with two parameters, β and n:

$$\eta_t = \frac{\beta}{2n} \sum_{l \in L} f(\mu_{lt}) \tag{4.18}$$

$$f(\mu_{lt}) = \begin{cases} \mu_{lt}^{2n} & \text{if } \mu_{lt} \leq 1 \\ 2n(\mu_{lt} - 1) + 1 & \text{if } \mu_{lt} \geq 1. \end{cases} \tag{4.19}$$

The contribution of an individual line has a constant slope when $\mu_{lt} \geq 1$ and is convex for $n \geq 1/2$. The slope of (4.18) controls the sensitivity of the reserve requirement to changes in line utilization:

$$\frac{\partial \eta_t}{\partial \mu_{lt}} = \begin{cases} \beta \mu_{lt}^{2n-1} & \text{if } \mu_{lt} \leq 1 \\ \beta & \text{if } \mu_{lt} \geq 1. \end{cases} \tag{4.20}$$

Equation (4.20) defines how much additional reserve the model requires to counter marginal increases in the anticipated line utilization. The parameter β controls the sensitivity

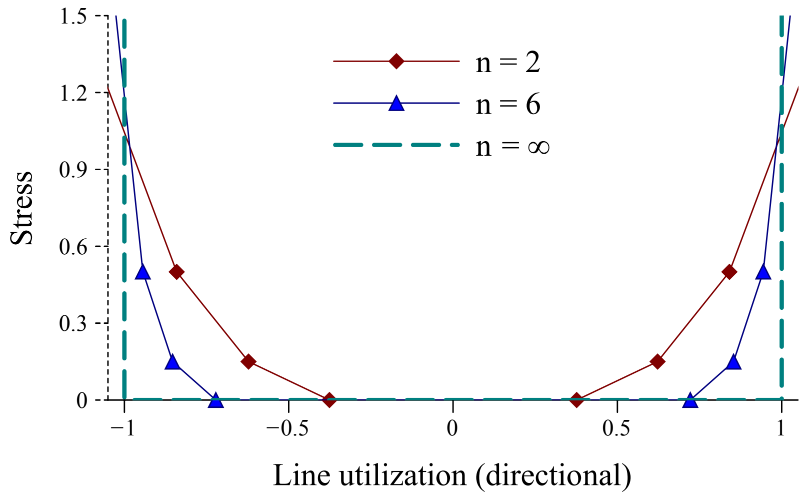


Figure 4.5: Piecewise linear approximations of the transmission stress function (4.18) with $n = 2$, 6, and ∞ and unit $\alpha = \beta/2n$.

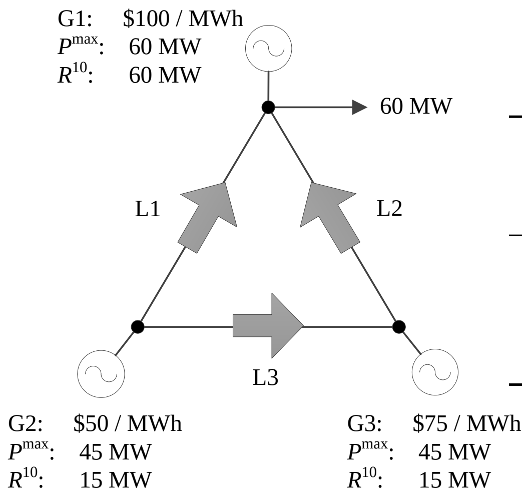
to changes in utilization when $\mu_{lt} \geq 1$. Large values of β are appropriate when large quantities of reserves are necessary to alleviate congestion. The parameter n discounts the sensitivity when $\mu_{lt} < 1$ because less additional reserve is necessary when no violations are predicted. These parameters should be tuned to the respective system in order for the model to be effective.

The general transmission stress model (4.5)–(4.8) can be made to approximate (4.18) by specifying $\alpha = \beta/2n$ and piecewise linear segments that approximate $f(\mu_{lt})$. Figure 4.5 shows the contribution of an individual line to the reserve requirement for unit α and multiple values of n. Setting $n = \infty$ is a special case that effectively fixes the reserve requirement and constrains all $\mu_{lt} < 1$. Large values of n are appropriate when the operator is confident that the participation factors accurately anticipate post-contingency decisions.

Specification of congestion-based reserve requirements should be based on off-line analysis. Such analysis could make use of simulation and statistics to compare how well different stress functions predict post-contingency flows. Furthermore, it should investigate the expected cost and reliability for different policies. Operators with intimate knowledge of the system may improve the formulation by developing distinct stress functions for different lines or by modeling dependencies between lines, such as is done today using nomograms. The value of reserve is difficult to specify a priori, but operators have expressed interest in similar solutions. For example, MISO (2012b) uses a reserve demand curve based on off-line valuation of different reserve margins. The proposed model is analogous to dynamically updating a reserve demand curve based on how transmission stress affects the anticipated value of a reserve margin.

4.6 Illustrative Example

This section demonstrates congestion-based reserve requirements with $n = \infty$ on a small test case comprising three transmission lines, three nodes, and three generators. The system attributes are shown in Figure 4.6. There is a single load of 60 MW that can be served by a generator at the same location. However, it is more economical to import power from other parts of the network. This type of arrangement is typical in power systems: the cheapest generators are usually located outside of the biggest load centers and have their power imported over high-voltage transmission lines. Consider a single-period DCOPF that includes a requirement for reserve to exceed the largest contingency. Problem details and the exact formulation are provided in Appendix D.



Line	Ratings (MW)		PTDFs		
	F	$\tilde{F}^{10}$	G1	G2	G3
L1	30	33	0	$2/3$	$1/3$
L2	30	33	0	$1/3$	$2/3$
L3	999	999	0	$1/3$	$-1/3$

Figure 4.6: The three-node test case (the arrows indicate reference directions).

Figure 4.7 shows the extreme points of the feasible space and the corresponding flows on lines L1 and L2. All optimal solutions coincide with the rightmost edge of the polyhedron. The total cost tends to be lower when the flow on L1 is large because the cheapest generator has a large effect on this line. Suppose that E_2^0 is selected as an initial solution and consider the response to the outage of generator G1. Figure 4.8 describes the only recourse action that balances supply and demand following this contingency. The initial solution is not N-1 reliable because this recourse action violates the emergency capacity of L1.

Figure 4.9 summarizes the decision space that protects against all generator contingencies [4] . This space is difficult to identify for large problem instances due to the number of extreme points. However, for small instances, such as this, it is easier to compare how

[4] Failures of G2 and G3 are ignored because they are trivially reliable in the initial solution. A description for why this is the case is provided in Appendix D

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^0	0	30	30	30	30	\$3,750
E_2^0	15	45	0	30	15	\$3,750
E_3^0	30	30	0	20	10	\$4,500
E_4^0	30	0	30	10	20	\$5,250
E_5^0	15	0	45	15	30	\$4,875

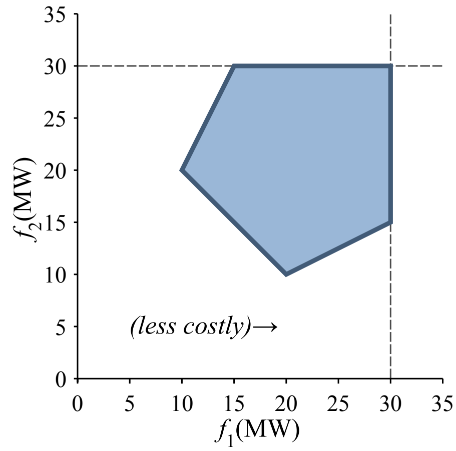


Figure 4.7: Extreme points and the feasible space projected onto f_1 and f_2.

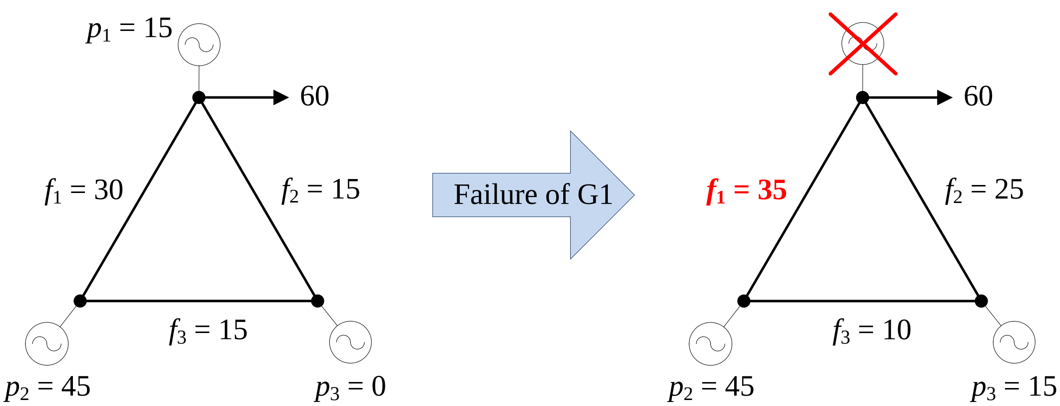


Figure 4.8: Initial solution E_2^0 and the response to the failure of G1. This is the only response that balances supply and demand, but the transmission limit on line L1 is violated.

well congestion-based reserve requirements approximate the N-1 reliable space. Ideally, the proposed approach will preserve optimal solutions and remove unreliable ones.

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^1	0	30	30	30	30	\$3,750
E_2^1	12	42	6	30	18	\$3,750
E_3^1	30	24	6	18	12	\$4,650
E_4^1	30	6	24	12	18	\$5,100
E_5^1	12	6	42	18	30	\$4,650

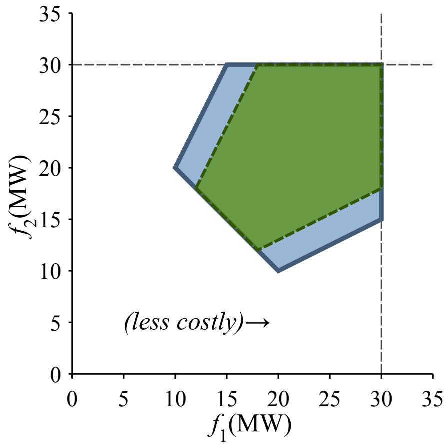


Figure 4.9: The N-1 reliable space (dashed border).

The participation factors from the response in Figure 4.8 can be used to derive operational shift factors using equation (4.12). The shift factors for lines L1 and L2 are [5]

$$\begin{aligned} a_1 &= \rho_1 \times PTDF_{11} + \rho_2 \times PTDF_{12} + \rho_3 \times PTDF_{13} \\ &= (-1 \times 0) + \left(0 \times \frac{2}{3}\right) + \left(1 \times \frac{1}{3}\right) \\ &= \frac{1}{3}, \end{aligned} \tag{4.21}$$

$$\begin{aligned} a_2 &= \rho_1 \times PTDF_{21} + \rho_2 \times PTDF_{22} + \rho_3 \times PTDF_{23} \\ &= (-1 \times 0) + \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{2}{3}\right) \\ &= \frac{2}{3}. \end{aligned} \tag{4.22}$$

The flows on L1 and L2 are anticipated to change by $(\delta_1 = a_1 \times p_1)$ and $(\delta_2 = a_2 \times p_1)$ following the loss of generator G1. Now consider congestion-based reserve requirements with $n = \infty$. Recall that this policy effectively bounds the anticipated flows

[5] Recall from (4.10) that ρ_{it}^g is the participation factor for generator i in response to the outage of generator g. The indices g and t are suppressed here from a_{lt}^g and ρ_{it}^g because this example considers only one outage (the loss of G1) during a single period.

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^2	0	30	30	30	30	\$3,750
E_2^2	9	39	12	30	21	\$3,750
E_3^2	21	39	0	26	13	\$4,050
E_4^2	30	30	0	20	10	\$4,500
E_5^2	30	21	9	17	13	\$4,725
E_6^2	$\frac{9}{2}$	21	$\frac{69}{2}$	$\frac{51}{2}$	30	\$4,088

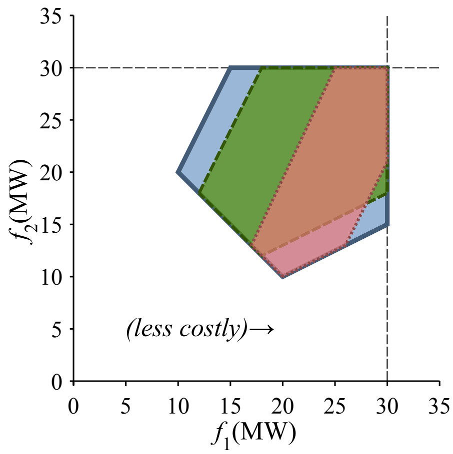


Figure 4.10: Feasible space using congestion-based reserve requirements with $n = \infty$ (dotted border).

below emergency limits. This is equivalent to adding constraints

$$-\tilde{F}_1^{10} < f_1 + \delta_1 < \tilde{F}_1^{10} \quad \longrightarrow \quad -33 < \frac{1}{3}p_1 + \frac{2}{3}p_2 + \frac{1}{3}p_3 < 33 \tag{4.23}$$

$$-\tilde{F}_2^{10} < f_2 + \delta_2 < \tilde{F}_2^{10} \quad \longrightarrow \quad -33 < \frac{1}{3}p_1 + \frac{1}{3}p_2 + \frac{2}{3}p_3 < 33. \tag{4.24}$$

Figure 4.10 shows the new feasible space considering constraints (4.23) and (4.24) as non-strict inequalities. The optimal solutions include all points on the rightmost edge of the reduced polyhedron. Unlike before, every one of these solutions is reliable to the loss of any generator. For example, Figure 4.11 demonstrates how operators may recover from the loss of G1. Reliability is established in this example by increasing the flow on L2, which is intrinsically linked with reducing the output of G1 so the contingency becomes less severe. In this small example, the congestion-based reserve requirements remove all optimal solutions that are also unreliable.

Recall that large values of n are preferred when the participation factors are predictable. This condition can be hard to achieve. The participation factors are not consistent

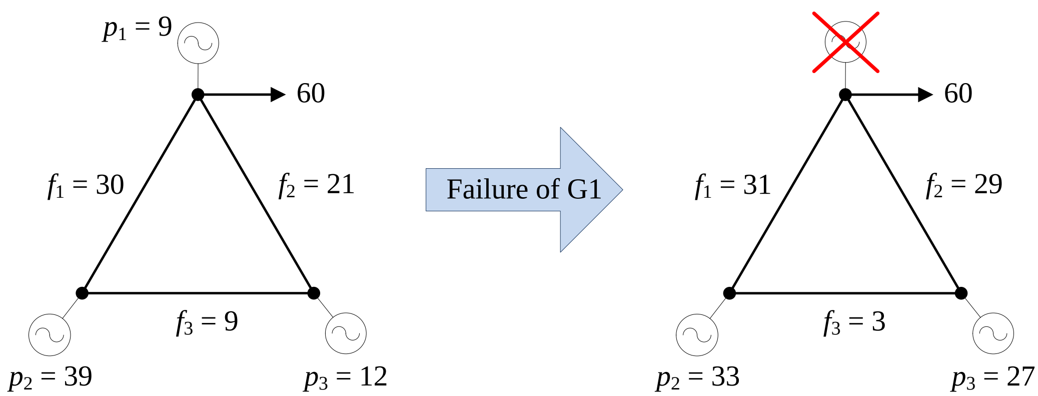


Figure 4.11: Solution E_2^2 and a feasible response to the failure of G1.

even in this example for different baseline solutions: ρ varies between $(-1,0,1)$ and $(-1,-2/3,5/3)$ in the responses shown in Figure 4.8 and Figure 4.11. It is visually apparent in Figure 4.10 that congestion-based reserve requirements provide an imperfect approximation of the N-1 reliable region. For example, the model both under and over approximates the reliable space in the neighborhood of $(f_1, f_2) = (28, 17)$. However, in this instance, the feasible region is reasonable for many *low cost* solutions. Testing on larger networks shows that smaller values of n often provide better approximations. Parameter tuning is an important step to complete because it can significantly influence the performance of the proposed algorithm.

4.7 Analysis and Results: IEEE 73-Bus (RTS 96) Test Case

Different reserve margin policies are compared using a 24-hour unit commitment model on the peak day of the RTS 96 test case. The test case is described by Grigg et al. (1999) and is publicly available from the Univ. of Washington (1999) website. The system has 73 nodes, 99 generators, and 117 transmission lines.

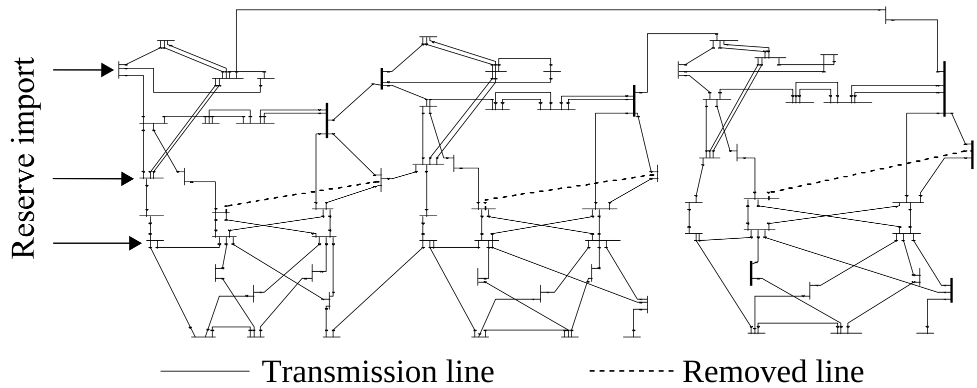


Figure 4.12: Modified IEEE 73-bus test case with up to 200 MW of reserve imports.

It is common to modify IEEE test cases. This analysis follows Hedman et al. (2010) by removing line (11–13), shifting 480 MW of load from buses 14, 15, 19, and 20 to bus 13, and decreasing the capacity of line (14–16) to 350 MW. These modifications are made to each of the three identical areas that comprise the system. The removal of lines and shifting of load was originally done by J. McCalley et al. (2004) and the changing of line capacities was originally done by Motto, Galiana, Conejo, and Arroyo (2002). A small amount of congestion is induced by tripling the capacity of inexpensive hydro power in the area consisting of buses 1–24 and removing hydro from all other areas.

The system is treated as a single zone within a larger system that can import up to 200 MW of reserve evenly across buses 3, 15, and 17; these locations mirror the connection between the left and central areas of the network as shown in Figure 4.12. All importable reserves are counted towards the respective reserve requirements. Non-spinning reserves are not modeled.

The basic and general models are tested using several values of α and transmission stress functions that approximate (4.18) with $n = 2$, 6 and ∞, as shown in Figure 4.5. The general model is first solved using the naïve assumption $\delta_{lt}^{+} = -\delta_{lt}^{-} = \tilde{F}_l^{10} - F_l$ to test performance without using shift factors to predict changing flows. Unit commitment is then solved again using shift factors derived from critical contingencies of the initial solution.

Reliability is evaluated against restoring frequency within 10 minutes while satisfying emergency transmission limits. Slack variables are added to allow transmission violations when these conditions cannot be met. Contingency analysis is performed for each hour using a DC optimal power flow that minimizes the sum of transmission violations. The contingency analysis formulation is provided in Appendix C.2. For the purpose of reporting statistical expectations, failure rates are assumed to be stationary and are inferred from the mean time to failure. Over the course of a single day, there is an 87% chance of at least one generator failure and a 10% chance of at least one transmission failure.

Testing is performed using CPLEX v12.4 on an 8-core 3.6GHz computer with 48GB of memory. SCUC is terminated after five minutes or upon reaching an optimality gap of 0.1%. All 147 runs achieve an optimality gap of 0.15% and only five do not reach 0.1% within the time limit.

4.7.1 Interpretation. Congestion-based reserve requirements are compared to two sets of baseline reserve policies: the first (B1) specifies that reserve must exceed the largest contingency as well as a percentage of load and the second (B2) specifies that reserve must exceed the largest contingency times a factor that is greater than one. These roughly reflect the reserve policies of WECC and PJM (WECC, 2008; PJM, 2013). The reserve parameters are varied to gauge performance for different levels of conservatism.

Policies are compared with respect to cost and the expected transmission violations (E[trans violation]). Transmission violations are a good indicator of reliability but do not imply that load curtailment must occur. Any violations reported by contingency analysis will be corrected by the system operator by either re-running SCUC (with additional restrictions to ensure a reliable solution) or by performing out-of-market corrections.

This analysis makes no statement as to the optimal balance between cost and reliability in the initial SCUC solution. Strong policies are distinguished as those that tend to provide more reliability at no additional cost, i.e., policies that provide Pareto efficient solutions. A solution is Pareto dominated if another solution is better with respect to both objectives.

4.7.2 Results for Normal Conditions. Figure 4.13 (a) and (b) show results from the baseline reserve and basic transmission stress policies. The reserve margins begin near zero and are progressively made more conservative. The complete range of policies and average solution times ($\overline{T}$) are shown in Table 4.1. The expected sum of violations is 0.37 MW when there is no reserve margin; this and other conservative policies are omitted from Figure 4.13 because results are similar when reserve margins are near zero. As requirements become stricter, unit commitment costs increase and reliability tends to improve. The trade-off between cost and reliability is almost indistinguishable between the baseline and basic transmission stress policies for this test case.

Figure 4.13 (c) and (d) show that the general models are more efficient for highly reliable solutions. Updating shift factors improves reliability in all cases but is accompanied

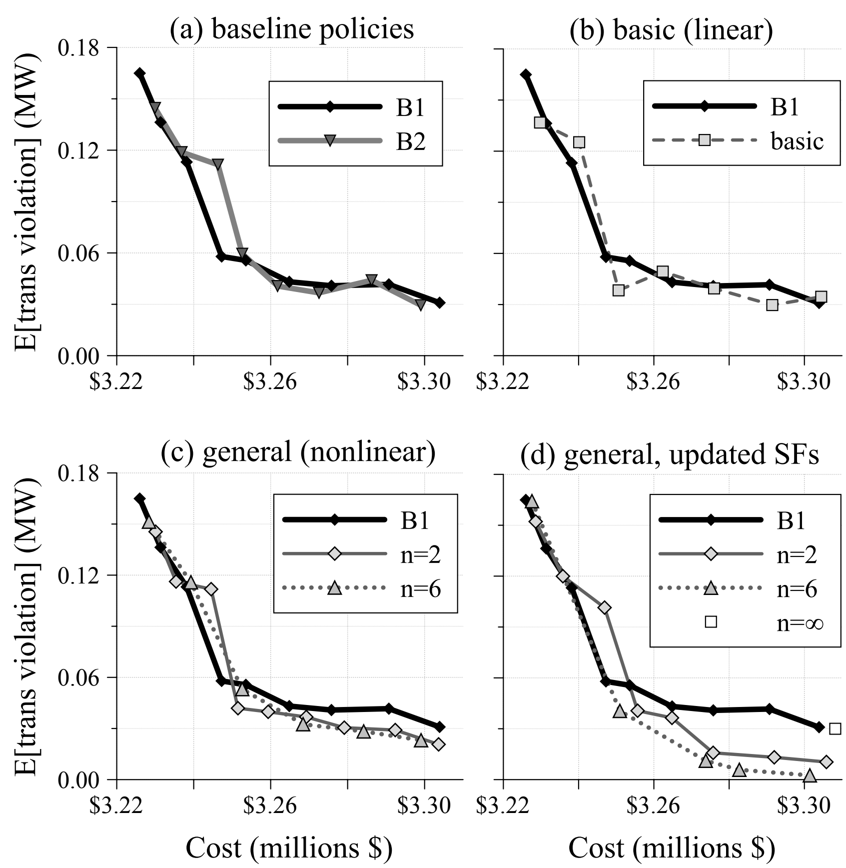


Figure 4.13: Performance of the baseline and congestion-based reserve requirements for various levels of conservatism.

by an increase in cost. Reliability improves at relatively little cost for $n = 2$ and 6 when expected violations are less than 0.05 MW, resulting in Pareto dominant solutions in this region. The general models take the longest to solve on average. However, solution times are highly variable and the fastest 66% of the general models solved quicker than the slowest 25% of either baseline policy.

Recall that $n = \infty$ is equivalent to assuming that each set of pre-defined participation factors correspond to a feasible re-dispatch with no violations. In this example, $n = \infty$ reduces the expected violations by over 90% compared to the initial run (from 0.37 MW to

Table 4.1: Test descriptions and statistics.

	Approach	Policy range (α)	$\overline{T}$ (mins)	Coeff # viol	Coeff max viol
Normal	B1 ($\times$load)	$[0.03, 0.08]$	0.59	0.48	0.62
	B2 ($\times$lrgst contg)	$[1.0, 2.1]$	0.87	0.44	0.63
	basic	$[0, 45]$	0.95	0.46	0.66
	$n = 2$	$[0, 30]$	1.32	0.47	0.68
	$n = 6$	$[0, 70]$	0.93	0.48	0.68
	$n = 2$, updated SF*	$[0, 25]$	1.17	0.47	0.72
	$n = 6$, updated SF	$[0, 50]$	0.98	0.53	0.76
	$n = \infty$, updated SF	1	2.40	–	–
Stressed	B1 ($\times$load)	$[0.03, 0.09]$	0.62	0.63	0.81
	B2 ($\times$lrgst contg)	$[1.0, 2.0]$	0.80	0.53	0.61
	basic	$[0, 50]$	0.99	0.51	0.61
	$n = 2$	$[0, 22.5]$	2.62	0.58	0.67
	$n = 6$	$[0, 80]$	0.62	0.60	0.74
	$n = 2$, updated SF	$[0, 20]$	1.45	0.62	0.71
	$n = 6$, updated SF	$[0, 60]$	0.94	0.56	0.78
	$n = \infty$, updated SF	1	5.20	–	–

*SF = shift factors

0.03 MW) without requiring any reserve margin; the gain in reliability is entirely attributed to reducing transmission stress. This approach proves costly, however, suggesting that it is more efficient to instead encourage alternative recourse decisions by procuring additional reserve. Larger values of n would become desirable if the shift factors better anticipated post-contingency decisions.

Other relevant reliability metrics include the maximum sum of violations (max viol) following a single contingency and the number of contingencies with non-zero violation (# viol). The performance of these metrics relative to expected violation is assessed using log-linear regression. The fitted regression coefficients, shown in Table 4.1, approximate the percent change in the respective metric per percent change in the expected violation (Törnqvist, Vartia, & Vartia, 1985). The coefficients measure relative sensitivities – larger values indicate that the respective metric improves quickly relative to the curve in Figure 4.13 [6] . The number of violations follows a similar relative trend for all policies, but the maximum violations improve faster using the congestion-based reserve requirements. In particular, updating shift factors appears to be especially efficient at minimizing the "worst" contingencies (max viol). This is particularly useful if the operator is interested in avoiding large amounts of load shedding during a single event.

4.7.3 Results for Stressed Conditions. It is common for resources to be unavailable due to scheduled maintenance or a forced outage. In this section, the absence of a large baseload generator (unit 90) leads to importation of cheap energy from neighboring parts of the system and the additional congestion significantly reduces reliability compared to normal operation. It may be advantageous to change the reserve zones to reflect the new congestion pattern; however, such changes are rarely implemented in practice. For example, MISO will only consider reconfiguring reserve zones when the adverse condition is projected to last multiple days and cannot be resolved through other operating procedures (MISO, 2012a).

Figure 4.14 (a) and (b) shows that the basic transmission stress model dominates only B1 when costs exceed \$3.66 million. However, Figure 4.14 (c) and (d) shows that the general models outperform both baseline approaches. Updating the shift factors usually provides efficient solutions, but the jagged results in Figure 4.14 (d) demonstrate that the

[6] Regression is used here to simplify the results and avoid replicating Figure 4.13 for every reliability metric. The fitted models all have reasonable fits, with r^2 ranging from 0.72 to 0.99.

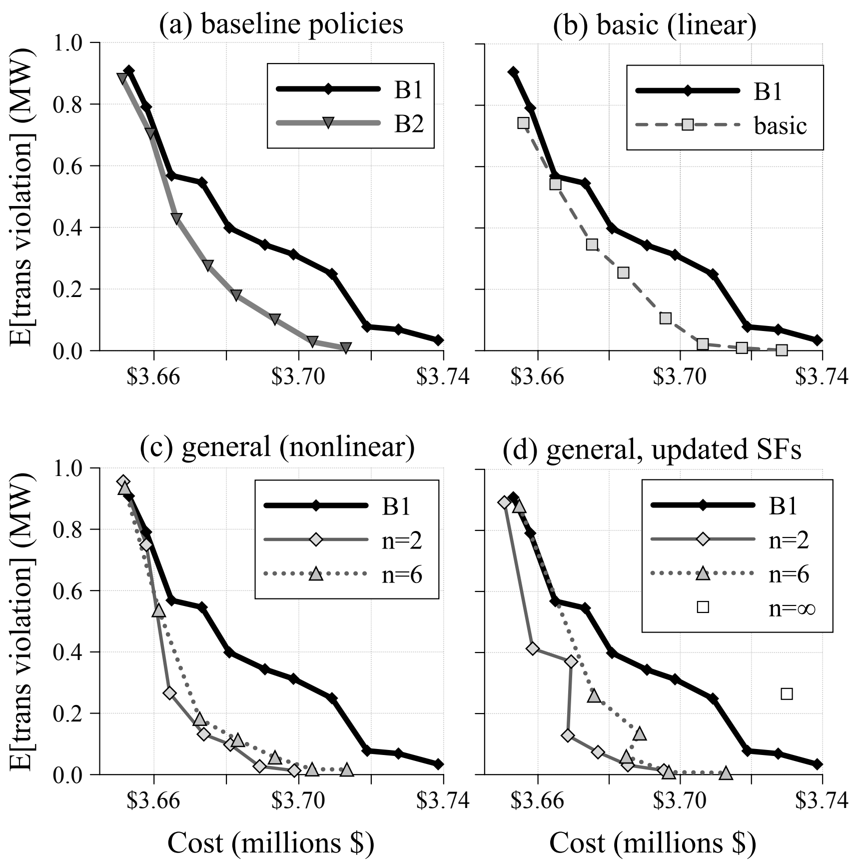


Figure 4.14: Performance of the baseline and congestion-based reserve requirements on the stressed system.

benefit can be inconsistent because generator participation in re-dispatch may differ from the assumed values.

The model with $n = \infty$ again improves reliability without increasing the minimum reserve level, but is not cost-efficient. Otherwise, all of the general policies are competitive when compared to the baseline approaches. As before, the general models are particularly efficient at improving the worst contingencies (max viol). Figure 4.15 compares the average maximum violations across hours between the baseline and general models with no shift factors. The worst contingency is often the failure of line (56–57); the congestion-based

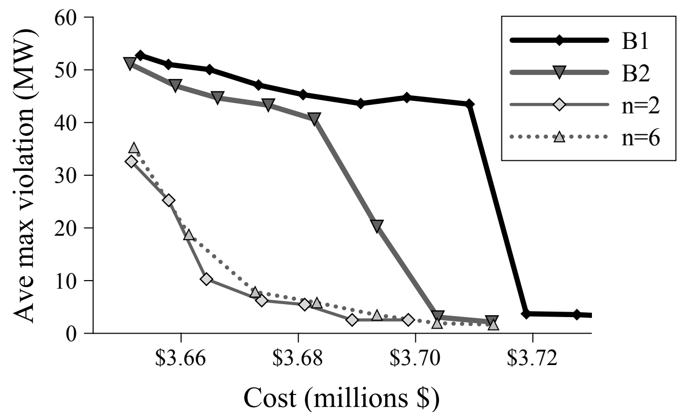


Figure 4.15: Maximum sum of violations for any contingency averaged across hours.

reserve requirements are efficient because they encourage a preemptive reduction of flow on this line to improve reliability without procuring a large expensive reserve margin.

4.7.4 Reliability by Reducing Transmission Stress. Congestion-based reserve requirements can achieve reliability more economically by reducing transmission stress when it is inexpensive to do so. This effect is demonstrated by fixing the minimum reserve quantities identified by the general model with $n = 2$ and solving SCUC with these as *explicit* requirements. Results for the stressed system, shown in Figure 4.16, show that much of the performance is due to reducing transmission stress as opposed to identifying effective reserve quantities.

It can be difficult to predict how much reducing transmission stress will improve reliability because reliability is contingent on an optimal power flow to re-dispatch generators. The third most conservative $n = 2$ policy in Figure 4.16 is dominated by the fixed reserve requirement, i.e., it is possible that reducing transmission stress will lead to a solution that is less economical *and* less reliable. Extensive off-line analysis is justified in order to ensure

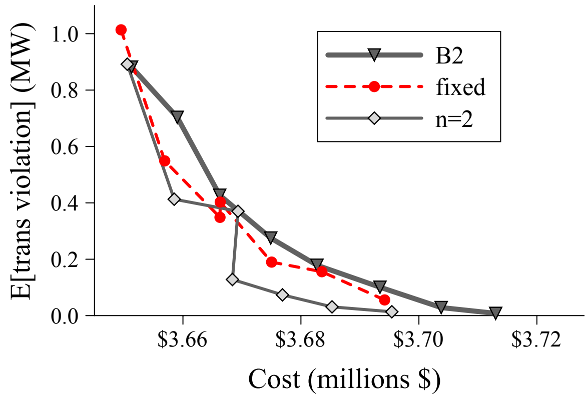


Figure 4.16: Results after fixing the reserve levels identified using $n = 2$.

consistent performance and avoid undue reductions in transmission stress when the more appropriate response is to increase reserve.

4.8 Conclusion

Most scheduling models assume reserve is highly deliverable within zones. Intra-zonal reserve deliverability remains an unresolved practical issue that operators address, in part, by procuring reserve margins that add flexibility to post-contingency dispatch decisions. These margins are often based on simple rules, such as a percentage of load or factor above the largest contingency, and may lead to overly-expensive solutions or expensive out-of-market corrections when applied to cover a broad range of uncertainty.

This chapter proposes a method for coupling reserve requirements with transmission stress. Just as important as determining a proper reserve quantity, the method may preemptively reduce flows on critical lines. The model satisfies security constraints by reducing transmission stress or acquiring a larger reserve margin, whichever is more economical. Transmission stress is based on an approximation of how power flows will change following a contingency based on pre-defined recourse participation factors. If the assumed participations would cause a violation, the model procures additional reserve to improve the likelihood that operators can find another solution that avoids all violations. Future research may improve performance by more accurately anticipating how individual reserve providers will participate in contingency response. The participation factors may also be extended to consider likely deviations in intermittent generation. The efficiency of the proposed method approaches stochastic programming when the participation factors better anticipate recourse decisions.

Results compare the congestion-based reserve requirements to common baseline policies. The benefit of the proposed approach is demonstrated under normal conditions and is pronounced when the system becomes more stressed such as during an the long-term outage of a large baseload generator. Therefore, testing suggests that the approach may help operators efficiently handle a wider range of situations without needing to update zones or perform out-of-market corrections.

Chapter 5

RESERVE DISQUALIFICATION FOR DISTINCT SCENARIOS

5.1 Introduction

Most large balancing authorities use reserve zones to ensure that enough reserve is held in import-constrained regions. However, these zones are structurally limited and are expected to address a wide range of uncertainty: requirements are inevitably overly conservative for some scenarios and insufficient for others. The limitations of zonal reserve requirements are discussed in Section 5.3. If some reserves are found to be undeliverable, operators in MISO and ISO-NE manually disqualify them so that the reserve requirements must be met by reserves at other locations (FERC, 2011; ISO-NE, 2013b, 2012). The strategy of disqualifying reserves at sub-prime locations is formalized and generalized in this chapter using the idea of response sets.

A response set defines a group of entities that are eligible to provide reserve for a particular scenario in a particular period. Scenario-specific response sets help security-constrained unit commitment (SCUC) anticipate what corrective actions will be feasible for different contingencies. This work focuses on generator contingencies, but the structure is theoretically compatible with any other type of uncertainty. Figure 5.1 shows an example of contingency response sets for a hypothetical system with three generators – only resources in the shaded regions are categorized as deliverable for the underlying contingency. The

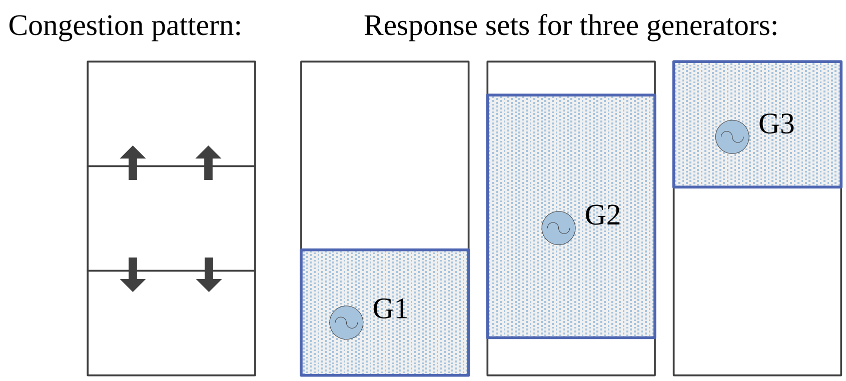


Figure 5.1: Overhead view of a hypothetical system showing the response sets for three generator contingencies. The arrows indicate the flow on congested interfaces. Each set must have enough reserve to cover the loss of the respective generator.

response sets for generators G1 and G2 are relatively small because congestion only allows reserve to be delivered from nearby areas. The benefit of response sets are that 1) different treatment can be given to entities that would traditionally fall in the same zone and 2) the response sets may change across scenario-periods to reflect the distinct congestion patterns that may arise. The proposed model integrates response sets with traditional reserve requirements in order to leverage the strengths of both approaches.

A decomposition algorithm is proposed that updates response sets while solving SCUC. The algorithm is a heuristic that mimics Benders' Decomposition for stochastic programs. The algorithm is passive in the sense that it only overrides traditional reserve requirements when reserve deliverability assumptions are violated. The approach also relates to existing reserve disqualification practices. It can be characterized as a generalized reserve disqualification procedure because removing a resource from a response set is equivalent to disqualifying its reserve *for a single contingency*.

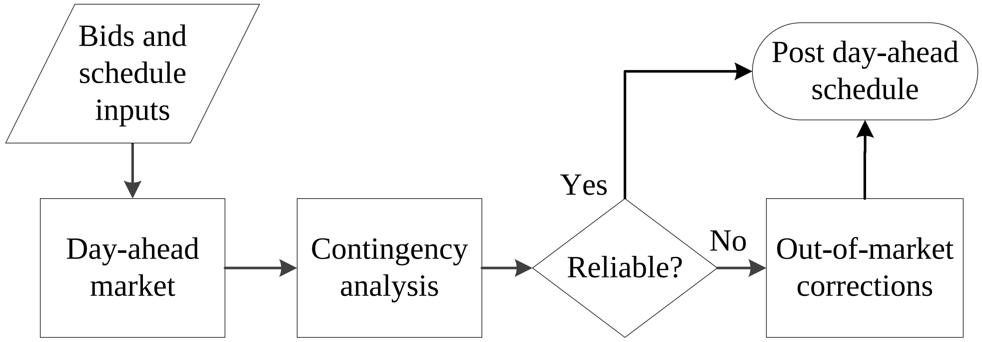


Figure 5.2: Day-ahead scheduling process.

The day-ahead scheduling process is assumed to follow the flowchart in Figure 5.2. The day-ahead market (DAM) is first cleared using SCUC to establish an initial schedule that facilitates market settlements. Operators then evaluate reliability and apply out-of-market corrections (OMCs) whenever the market solution is unreliable. More details on day-ahead scheduling is provided in Section 2.2.4 and by Al-Abdullah et al. (2013). The proposed algorithm may be integrated within the DAM or used as OMC strategy to repair unreliable market solutions in an economical manner.

The approach is assessed relative to its ability to reduce operating costs subject to satisfying N-1 reliability. The model only overrides the baseline formulation when the SCUC solution is found to be unreliable. In such cases, the algorithm is shown to improve reliability at relatively little cost. The algorithm is sufficiently fast to apply as an OMC procedure and it may also be applied within the DAM in a limited way if time allows.

The contribution goes a step further for market environments. Traditional markets compensate all members in the same zone at the same rate. This can create market inefficiencies whereby resources behind transmission bottlenecks are paid too much for reserve relative to resources at prime locations. Not all locations in a zone are necessarily able to provide equal service. This chapter proposes a reserve market settlement scheme that compensates resources based on their service to individual contingencies. Using the illustration in Figure 5.1 as an example, resources belonging to the response set for G1 may be compensated less than resources that service both G1 and G2. The proposed methodology provides a more precise payment scheme that can create different reserve prices for resources in the same zone based on congestion.

This chapter is organized as follows: Section 5.2 describes a baseline zonal approach that serves as a foundation for the proposal. Section 5.3 describes the limitations of zonal reserve requirements. Section 5.4 describes the form and implementation of the proposed approach. Section 5.5 proposes a market settlement scheme for locational reserve prices. Section 5.6 demonstrates the methodology on a small test case. Section 5.7 presents numerical results on the IEEE 73-bus test case and discusses practical considerations. Finally, Section 5.8 concludes the chapter and notes some potential future work.

5.2 Baseline Reserve Requirements

The proposed reserve requirements build upon the DC unit commitment formulation outlined in Appendix C.1. Incorporating a full network model for all contingencies (á la stochastic programming) quickly becomes intractable for large systems due to how the problem size increases. Reserve zones are used in practice to approximate transmission constraints and enable deterministic formulations. It is well-known that zonal models imperfectly characterize power flows because they estimate cross-border flows and ignore intra-zonal congestion (Purchala et al., 2005). Nonetheless, zones are ubiquitous for reserve requirements because they allow operators to model transmission constraints on recourse decisions in an intuitive and computationally efficient way.

The baseline reserve model used in this work is inspired by the market model of ISO-NE (T. Zheng & Litvinov, 2008). Two small adjustments are made to the formulation. The first (and most trivial) difference is that the model only considers 10-minute reserves procured for N-1 contingency response. The second change is a departure from nested zones, which suit the ISO-NE system but may be inappropriate for general networks. An example of the difference between nested and mutually exclusive zones is shown in Figure 5.3. The defining characteristic of nested zones is that there is assumed to be no congestion limiting reserve deliverability from child to parent zones. This assumption is based on knowledge of the underlying system and may be unreasonable for general cases. For example, CAISO, ERCOT, MISO, and PJM have found mutually exclusive zones to be preferable for their respective systems (CAISO, 2007; ERCOT, 2007; MISO, 2008; PJM, 2012) [1] . The reserve requirements in this chapter are formulated for mutually exclusive zones.

The reserve model proposed by T. Zheng and Litvinov (2008) introduces variables that describe how much reserve can be shared between zones. A minimalistic representation

[1]The reserve zones for CAISO, ERCOT and MISO are illustrated in Figures 4.2, 3.2, and 2.6. PJM has a hybrid model where just one of the zones contains a nested sub-zone (PJM, 2012).

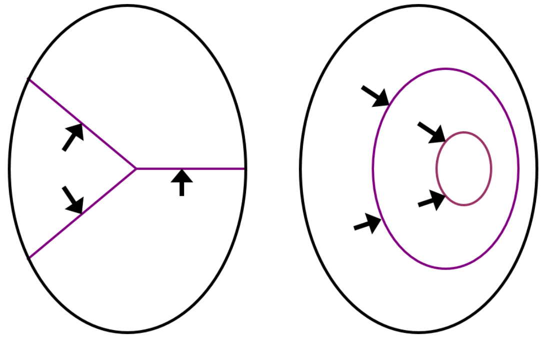


Figure 5.3: Examples of different types of reserve zones. The arrows indicate typical power flow directions across the zonal interfaces.

is

$$\sum_{k \in Z} \tilde{r}^j_{kt} \geq p_{ct} + r_{ct}, \qquad \forall j \in Z, c \in \mathscr{G}(j), t \in T \tag{5.1}$$

$$\tilde{r}^j_{kt} \leq \sum_{g \in \mathscr{G}(k)} r_{gt}, \qquad \forall j \in Z, k \in Z, t \in T \tag{5.2}$$

$$\tilde{r}^j_{kt} \leq S^j_{kt}, \qquad \forall j \in Z, k \in Z, t \in T, \tag{5.3}$$

where p_{ct} is the pre-contingency production from generator c, r_{gt} represents reserve provided by resource g, and $\tilde{r}^j_{kt}$ represents how much reserve in zone k is classified as deliverable to zone j in period t. Equation (5.1) requires there to be enough portable reserve to cover the loss of any generator, (5.2) models reserve held within the zones, and (5.3) limits how much reserve may be shared between zones. The reserve sharing bounds S may be based on off-line analyses. For example, T. Zheng and Litvinov (2008) dynamically update S^j_{kt} to be the (predefined) transfer capability estimate from zone k to j less the interface flows in the modeled system state.

The above formulation is appropriate for modeling reserve sharing between *adjacent* zones. The model of T. Zheng and Litvinov (2008) differs slightly from the above because it models sharing across layers of *nested* zones. The appropriate structure depends on the system in question. The above may be adapted to accommodate nested zones if desired [2] .

5.3 Limitations of Traditional Reserve Requirements

Traditional reserve requirements suffer from two inherent limitations. First, reserve sharing estimates may be imprecise because they are based on off-line studies performed before the system state is known. These studies must anticipate the injection and withdrawal locations. Reserve locations are important because more power can be transferred between points that have similar shift factors on critical lines.

Consider the example in Figure 5.4. Although the system states are indistinguishable at the zonal level, more reserve can be transferred to zone two in the second case because reserve is held at locations that prefer the path that has more residual capacity. Zonal models are inherently imprecise because they assume that all members of the same zone have equivalent shift factors on congested interfaces.

The reserve sharing capability is hard to determine without knowing the locations of reserves. It also depends on the location of the contingency. Ideally, reserve sharing would reflect contingency-specific transfer capabilities. MISO acknowledges this concern and uses simulations to estimate reserve sharing capabilities on a per-contingency basis (MISO, 2012a). MISO fixes the minimum reserve quantity for each zone according to the contingency with the largest loss *minus* import capability. The largest contingency is not

[2] Nested zones can be accommodated by summing the right-hand side of (5.2) over all children of zone k. This makes the same assumption as T. Zheng and Litvinov (2008) that congestion only limits reserve deliverability from parent to child zones.

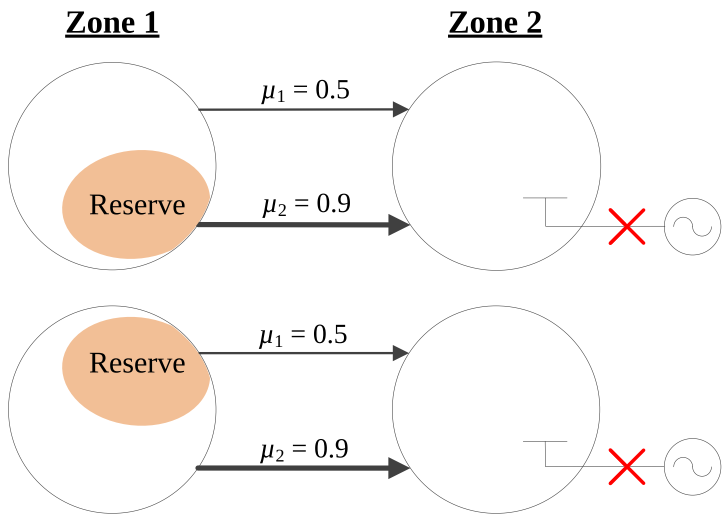


Figure 5.4: Transfer capability depends on reserve locations relative to highly utilized (μ) lines (this figure assumes that the interface lines have the same capacity).

necessarily the most severe because other contingencies may hinder reserve imports by reducing counter-flows on critical lines. Although MISO's requirements are contingency-specific, they are derived prior to SCUC and may be suboptimal whenever the schedule conflicts with the projected system state.

The second limitation is that traditional reserve requirements assume that no reserve deliverability issues will be caused by local congestion within a zone. One way to mitigate intra-zonal congestion is to procure a reserve margin for more dispatching freedom to alleviate congestion (Lyon, Hedman, & Zhang, 2014). These margins can protect against general uncertainties beyond N-1, but they are not an ideal way to improve reserve deliverability because operators lack means to control reserve locations within the zones.

5.4 Locational Disqualification

The limitations of traditional reserve requirements stem from the fact that they do not control the location of reserves within zones. To address this issue, the proposed model introduces contingency-specific response sets that suspend reserves held at unfavorable locations. The new reserve model is described by

$$\sum_{k \in Z} \tilde{r}^c_{kt} \geq p_{ct} + r_{ct}, \qquad \forall c \in G, t \in T \tag{5.4}$$

$$\tilde{r}^c_{kt} \leq \sum_{g \in \mathscr{G}(k)} \Gamma^c_{gt} r_{gt}, \qquad \forall c \in G, k \in Z, t \in T \tag{5.5}$$

$$\tilde{r}^c_{kt} \leq S^{z(c)}_{kt}, \qquad \forall c \in G, k \in Z, t \in T. \tag{5.6}$$

The zonal reserve variable $\tilde{r}^c_{kt}$ is now indexed by contingency c (instead of zone j) and represents reserve in zone k that is classified as deliverable for contingency c. The parameter $\Gamma^c_{gt} \in [0,1]$ limits the proportion of reserve that is classified as deliverable. Γ has no authority over the actual contingency response that happens in real time, but it can anticipate what reserves will be deliverable so that SCUC provides a more reliable solution.

Each response set must hold enough reserve to replace the underlying generator c, subject to the same reserve sharing constraints as the baseline model: $S^{z(c)}_{kt}$ limits sharing from zone k to the contingency zone $z(c)$. The new reserve requirements (5.4)–(5.6) are equivalent to (5.1)–(5.3) when all $\Gamma = 1$. Setting $\Gamma^c_{gt} = 0$ disqualifies the reserve provided by resource g. The proposed formulation is more general than MISO and ISO-NE practices because reserve disqualification is indexed by contingency and Γ may take any value between

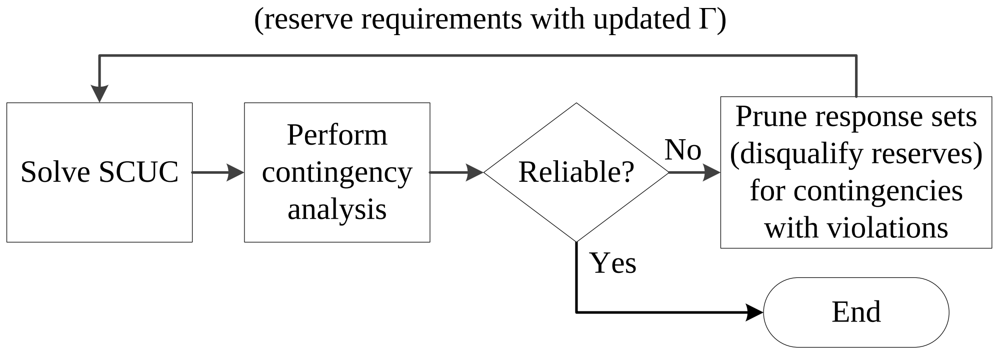


Figure 5.5: Decomposition algorithm for SCUC with reserve disqualification.

zero and one. Equations (5.4)–(5.6) enable contingency-specific management of reserve sharing and intra-zonal congestion, provided that proper values for Γ can be determined.

Figure 5.5 outlines a two-stage decomposition algorithm that iteratively updates Γ while solving SCUC. First, SCUC is solved using relaxed response sets (hereon assume that all $\Gamma_{gt}^{c} = 1$ to begin) to obtain a solution that is economical but not necessarily reliable. Contingency analysis then identifies contingencies that have insufficient deliverable reserves. The response sets are then pruned (i.e., made more restrictive by reducing Γ) for these contingencies and the restricted reserve requirements are passed back to SCUC. Each update can be viewed as a reliability cut because it restricts the amount of reserve designated as deliverable. These updates are analogous to the feasibility cuts generated by Benders' Decomposition for two-stage stochastic models. However, pruning Γ may remove larger portions of the feasible space than Benders' cuts: the proposed algorithm does not guarantee optimality but fewer iterations may be needed to converge to a reliable solution.

The proposed procedure in Figure 5.5 may be applied within the DAM or as a mechanism to determine OMCs. It is preferable to determine a reliable solution within the

DAM, but this frequently does not occur due to time limitations. When the proposed DAM solution is not reliable, market operators repair the solution using OMCs. These OMCs do not affect DAM settlements and they are often not allowed to de-commit units that were committed in the DAM framework (CAISO, 2013).

5.4.1 Pruning the Response Set for Contingency-Period (c, t). This section proposes a mathematical program to prune the response set for a single contingency c in a particular period t. The model identifies reserve disqualifications to pass back to SCUC in the process described by Figure 5.5. A practical pruning algorithm should have the following properties of speed, fairness, and effectiveness:

- To be sufficiently fast, the pruning algorithm should solve much quicker than SCUC.
- To be fair, it should not disqualify resources unduly and thereby exclude them from compensation in market environments.
- To be effective, the pruning algorithm should return a Γ that improves the SCUC solution to support reliability in an economical manner.

The empirical analysis in Section 5.7 will evaluate the speed, market outcomes, and economical effectiveness of the proposed approach. The formulation is a linear program that identifies reserve in the response set that is not deliverable. The model considers spinning reserves but can be generalized to include non-spinning reserves. For a particular contingency, let $\hat{I}$ represent the net injections (generation minus load) from the incumbent SCUC solution prior to re-dispatch, $\hat{R}$ be the available reserves, and $\hat{\Gamma}$ be the ruling disqualification factors from the previous iteration. The following equations assess the aptitude of the existing response set:

$$\sum_{n \in N} i_n = 0 \tag{5.7}$$

$$-F_l^{10} \leq \sum_{n \in N} PTDF_{nl} i_n \leq F_l^{10}, \qquad \forall l \in L \tag{5.8}$$

$$i_n = \hat{I}_n + \sum_{g \in G(n)} e_g, \qquad \forall n \in N \tag{5.9}$$

$$0 \leq e_g \leq \hat{R}_g \hat{\Gamma}_g, \qquad \forall g \in G, \tag{5.10}$$

where e_g is the amount of reserve dispatched from generator g. The response set is sufficient if there is a solution to these linear power flow equations. Equations (5.7) and (5.8) are flow balance and transmission constraints, (5.9) models locational injections, and (5.10) constrains reserve availability. The constraints may not be satisfied if congestion prevents reserve from being delivered. The pruning model (P_t^c) defined below relaxes the power flow and is guaranteed to be feasible whenever the reserve quantity exceeds the size of the contingency:

$$(P_t^c): \quad \text{Minimize: } \sum_{g \in G} R_g \gamma_g \tag{5.11}$$

Subject to:

$$\sum_{n \in N} i_n = 0 \tag{5.12}$$

$$-F_l^{10} \leq \sum_{n \in N} PTDF_{nl} i_n \leq F_l^{10}, \qquad \forall l \in L \tag{5.13}$$

$$i_n = \hat{I}_n + \sum_{g \in G(n)} e_g, \qquad \forall n \in N \setminus n(c) \tag{5.14}$$

$$i_n = \hat{I}_n + \sum_{g \in G(n)} e_g + \sum_{g \in G} m_g, \qquad n = n(c) \tag{5.15}$$

$$0 \leq e_g \leq \hat{R}_g \hat{\Gamma}_g (1 - \gamma_g), \qquad \forall g \in G \tag{5.16}$$

$$0 \leq m_g \leq \hat{R}_g \hat{\Gamma}_g \gamma_g, \qquad \forall g \in G \tag{5.17}$$

$$0 \leq \gamma_g \leq 1, \qquad \forall g \in G. \tag{5.18}$$

Constraints (5.14)–(5.18) separate corrective actions into two components: reserve dispatched through e is injected as normal and reserve dispatched through m is moved directly to the location of the contingency $n(c)$ without using any transmission capacity. A trivially feasible solution is to dispatch all reserve using m so that the nodal injections are the same as in the pre-contingency state. The variable γ identifies resources with non-deliverable reserve and the objective (5.11) minimizes the amount of reserve marked as non-deliverable. The solution to this problem is used to update the response sets using the relation

$$\Gamma_{gt}^c = \hat{\Gamma}_g (1 - \gamma_g^*), \tag{5.19}$$

where γ_g^* is from the optimal solution to problem (P_t^c). All resources with $\gamma_g^* = 1$ are completely removed from the response set because their reserve was not deliverable during simulation of a coordinated re-dispatch. The disqualified reserves may not count towards the reserve requirement (5.5) in SCUC during the next iteration. The critical assumption is that the percentage of reserve that is undeliverable will remain the same moving forward. This assumption is obviously uncertain because several aspects of the SCUC solution may change, e.g., reserve quantities can change, congestion can change, and near optimal solutions can generally have vastly different sets of committed generators (Sioshansi et al., 2008). The objective of (P_t^c) minimizes the reserve disqualified at each iteration in order to avoid hasty decisions. Hasty reserve disqualifications could impair performance and be unfair to market participants whose reserve is actually deliverable in the final solution (because participants are not paid for disqualified reserves). The overall performance depends on how well the updated Γ helps SCUC anticipate reserve deliverability based on previous solutions.

There is an optional step to force Γ_{gt}^c to be the same for all generators at the same node. This step would avoid market solutions where, by chance, generators at the same location have different percentages of qualified reserve. Market operators may take this precaution in order to give uniform treatment to resources at the same location. This optional step is not used or elaborated on here because it is not difficult nor perceived as likely to affect the algorithm's performance.

Computationally, the pruning algorithm is comparable to contingency analysis that operators already solve frequently to verify that the schedule or the current operating state is reliable. The problem should remain tractable for large-scale instances because pruning can be performed in parallel for different contingencies. The main concern is not the time to solve problem (P_t^c) but the number of iterations because re-solving SCUC at each iteration can be time consuming.

The pruning algorithm need not be convex. For example, problem (P_t^c) can be updated to include non-spinning reserves or AC power flow equations. Substituting an AC power flow would provide greater accuracy and allow for reserve to be disqualified based on voltage limitations. There has also been recent interest in using topology control to alleviate congestion (Korad & Hedman, 2013), and problem (P_t^c) may be amended to consider such corrective actions. The proposed algorithm can be extended to generally consider a wider scope of decisions than Benders' Decomposition because the recourse problem does not need to be convex (Geoffrion, 1972).

Response sets may be used for other types of scenarios besides the generator contingencies considered in this chapter. If a different reserve model is used, then the bounds on e (5.16) and m (5.17) should be adjusted to reflect the reserve up and reserve down requirements that are enforced in SCUC. The relaxation of (5.7)–(5.10) should also be amended so that m corresponds to the location of the disturbance being modeled.

5.5 Market Settlement

In US electricity markets, SCUC is used to clear energy and reserve bids. The market prices are based on dual variables from the solution to a linearized version of the problem (e.g., one linearizing technique fixes all binary variables after acceptable values have been determined (O'Neill et al., 2005)). Service providers are compensated based on these dual variables. Locational marginal prices (LMPs) are used to settle energy and reserve marginal prices (RMPs) are used to settle reserves. RMPs increase as reserve bids increase and when locational reserves are scarce.

Reserve sharing complicates the valuation of reserves at different locations. ISO-NE uses a settlement policy that rewards resources at key locations. Their nested zone model assumes that transmission only limits reserve sharing into child zones, such as load pockets (T. Zheng & Litvinov, 2008). Resources are not compensated for ancillary service to child zones because additional reserve is assumed to be undeliverable. This type of policy is appropriate when operators can predict which transmission limits will be binding and which ones will not. Otherwise, a more general mechanism is necessary that acknowledges service provided to neighboring zones only when marginal reserves are deliverable. Such a payment scheme is provided below:

$$\phi_{gt} = \sum_{c \in C} \Gamma^c_{gt} r_{gt} \left| \lambda^c_{z(g),t} \right|, \tag{5.20}$$

where $\Gamma^c_{gt} r_{gt}$ is the amount of qualified cleared reserve and λ^c_{kt} is the dual variable for 5.5 (note that $z(g) = k$ is the reserve zone containing generator g). Economic theory specifies that λ^c_{kt} is a shadow price that reflects the marginal value of reserve in zone k providing ancillary service for contingency c (O'Neill et al., 2005)). The payments defined by 5.20 compensate reserve providers based on service for individual contingencies. Service may be valuable to even a small contingency if other resources have been disqualified for that particular contingency.

The reserve price for resource g in period t is effectively $\sum_{c \in C} \Gamma_{gt}^{c} \left| \lambda_{z(g),t}^{c} \right|$. Reserve disqualification changes Γ and can thereby lead to different prices for resources in the same zone. Resources receive higher prices if their reserve is qualified for contingencies with substantial shadow prices. This is in contrast to traditional policies that derive a single reserve price for each zone.

Recall that the proposed model is equivalent to the baseline when all $\Gamma = 1$. The settlement scheme reverts back to zonal pricing in this condition. By initializing the algorithm with all $\Gamma = 1$, the proposed method *only* overrides the baseline requirements when the initial solution is not reliable. Proper compensation can be given to resources with deliverable reserve and payment is withheld from resources with undeliverable reserve. The quality of the prices (and the solution in general) depends on how Γ is determined. As a heuristic, the proposed algorithm may disqualify reserves that end up being deliverable in the final solution. However, the process is less heavy handed than existing forms of reserve disqualification because it still partially compensates resources for reserves that are only partially deliverable or only deliverable for a subset of contingencies.

5.6 Illustrative Example

This section evaluates three approaches that can be used to ensure reserve deliverability:

1. Reserve zones with reserve sharing (this is the traditional approach).

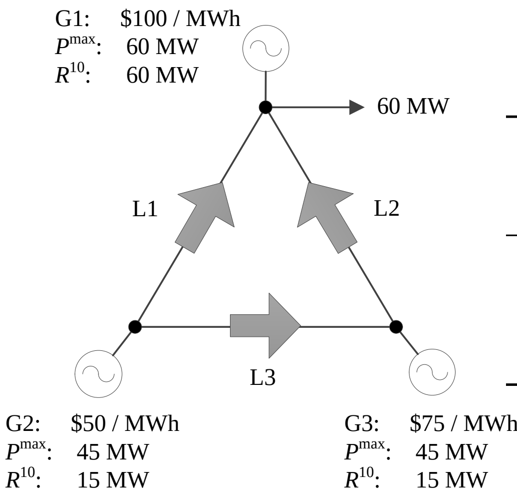
2. Reserve disqualification using response sets.

3. Benders' Decomposition.

The alternatives are compared on a small test case. An optimal solution can be found for this case using any of the above methods. However, the first approach requires users to define limits on the amount of reserve that can be shared between zones. The latter two approaches can obtain optimal solutions without any parameter tuning. Each alternative is compared in this section via visual analyses that show how well the feasible regions approximate the set of reliable solutions.

The test case has three nodes, three lines, and three generators. It was first introduced in Section 4.6 to analyze congestion-based reserve requirements. For convenience, the system attributes are again shown in Figure 5.6. There is a single load of 60 MW that can be served by a generator at the same location. However, it is more economical to import power from other parts of the network. This type of arrangement is typical in power systems: the cheapest generators are usually located outside of the biggest load centers and have their power imported over high-voltage transmission lines. Consider a single-period DCOPF that includes a requirement for reserve to exceed the largest contingency. Problem details and the exact formulation are provided in Appendix D.

Figure 5.7 describes the feasible space when using a single reserve zone. Also included is a graphical projection showing the subset of solutions that are reliable to all generator contingencies. The optimal solutions correspond to the rightmost edge of the



Line	Ratings (MW)		PTDFs		
	F	$\tilde{F}^{10}$	G1	G2	G3
L1	30	33	0	$2/3$	$1/3$
L2	30	33	0	$1/3$	$2/3$
L3	999	999	0	$1/3$	$-1/3$

Figure 5.6: Three-node test case (the arrows indicate reference directions).

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^0	0	30	30	30	30	\$3,750
E_2^0	15	45	0	30	15	\$3,750
E_3^0	30	30	0	20	10	\$4,500
E_4^0	30	0	30	10	20	\$5,250
E_5^0	15	0	45	15	30	\$4,875

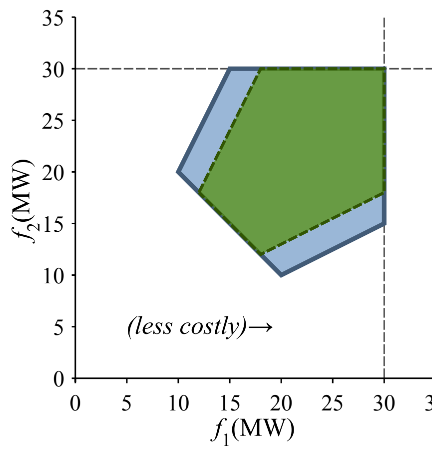


Figure 5.7: Extreme points of the feasible space when using a single reserve zone. The graph projects the feasible space (solid border) and reliable region (dashed border) onto the flow across lines L1 and L2.

feasible polyhedron. Some of these solutions are outside of the reliable region, indicating that not enough reserve is deliverable. The methods listed above are analyzed for their ability to improve reserve deliverability.

5.6.1 Two-Zone Model. The traditional approach is to improve reserve deliverability by introducing zones. Appendix D shows that the only generator contingency threatening reliability is the failure of G1. Therefore, let zone A correspond to the node containing G1 and let zone B correspond to the rest of the network. This zone definition is justified because zone A corresponds to the only location where reserve may be undeliverable. The baseline reserve sharing model defined by (5.1)–(5.3) is applied below to cover the loss of G1:

$$\tilde{r}_A^A + \tilde{r}_B^A \geq p_1 + r_1, \qquad \text{(reserve must cover the loss of G1)} \tag{5.21}$$

$$\tilde{r}_A^A \leq r_1, \qquad \text{(reserve in zone } A\text{)} \tag{5.22}$$

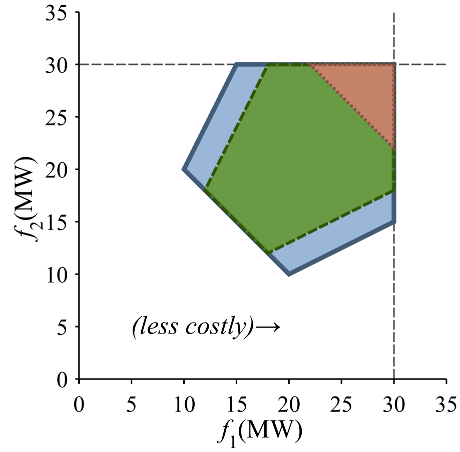
$$\tilde{r}_B^A \leq r_2 + r_3, \qquad \text{(reserve in zone } B\text{)} \tag{5.23}$$

$$\tilde{r}_B^A \leq S_B^A, \qquad \text{(reserve sharing limit from zone } B \text{ to } A\text{).} \tag{5.24}$$

The sharing parameter S_B^A limits the amount of reserve designated as deliverable from zone B to zone A. Large values of S_B^A make (5.21)–(5.24) equivalent to the single-zone model, and smaller values of S_B^A restrict production from G1 because less reserve can contribute to the left-hand-side of (5.21). Figure 5.8 describes the feasible space for several reserve sharing limits. These limits need to remove large portions of the feasible space before they ensure a reliable solution.

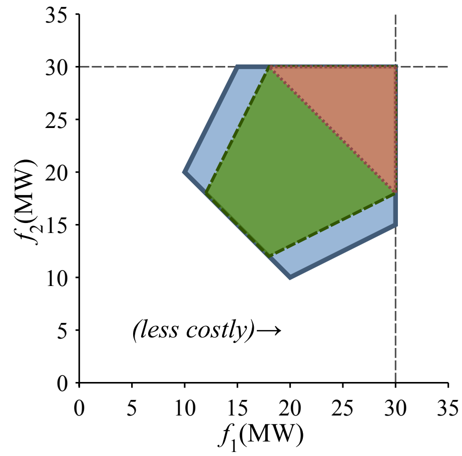
The model guarantees reliability only after removing much of the feasible space. This is undesirable because a change to the objective function (perhaps due to a different set of committed generators) could render the best solution infeasible. It would be preferable for the formulation to guarantee reliability *and* retain portions of the reliable region that are optimal or nearly optimal.

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^3	0	30	30	30	30	\$3,750
E_2^3	8	38	14	30	22	\$3,750
E_3^3	8	14	38	22	30	\$4,350



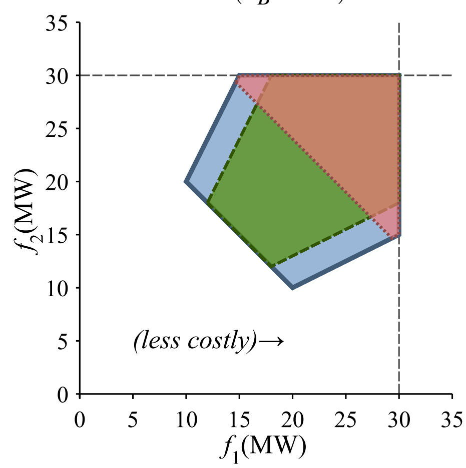
(a) Restrictive reserve sharing ($S_B^A = 8$).

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^4	0	30	30	30	30	\$3,750
E_2^4	12	42	6	30	18	\$3,750
E_3^4	12	6	42	18	30	\$4,650



(b) Least-restrictive sharing that ensures all optimal solutions are also reliable ($S_B^A = 12$).

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^5	0	30	30	30	30	\$3,750
E_2^5	15	45	0	30	15	\$3,750
E_3^5	16	44	0	$\frac{88}{3}$	$\frac{44}{3}$	\$3,800
E_4^5	16	0	44	$\frac{44}{3}$	$\frac{88}{3}$	\$4,900
E_5^5	15	0	45	15	30	\$4,875



(c) Lax reserve sharing ($S_B^A = 16$).

Figure 5.8: Feasible space using different zonal reserve sharing limits (dotted border). The graphs also show the reliable region (dashed border) and the feasible space of the single-zone model (solid border).

5.6.2 Reserve Disqualification. This section demonstrates reserve disqualification when using a single zone. Suppose that the initial solution is selected to be E_2^0 from Figure 5.7 (which falls outside of the reliable region). The only response that balances supply and demand after G1 fails is shown in Figure 5.9: G3 is the only generator holding reserve, and its reserve cannot be dispatched without violating the emergency limit of line L1. Therefore, it is reasonable to disqualify some of the reserve from G3 in hopes that it will be replaced in subsequent solutions by reserve that is deliverable. The pruning problem (P^1) defined by (5.11–(5.18) is applied below for this scenario:

$$\text{Minimize: } 0\gamma_1 + 0\gamma_2 + 15\gamma_3 \tag{5.25}$$

Subject to:

$$i_1 + i_2 + i_3 = 0 \tag{5.26}$$

$$-33 \leq \frac{2}{3}i_2 + \frac{1}{3}i_3 \leq 33 \tag{5.27}$$

$$-33 \leq \frac{1}{3}i_2 + \frac{2}{3}i_3 \leq 33 \tag{5.28}$$

$$-100 \leq \frac{1}{3}i_2 - \frac{1}{3}i_3 \leq 100 \tag{5.29}$$

$$i_1 = -60 + e_1 + m_2 + m_3 \tag{5.30}$$

$$i_2 = 45 + e_2 \tag{5.31}$$

$$i_3 = e_3 \tag{5.32}$$

$$0 \leq e_1 \leq 0\gamma_1 \tag{5.33}$$

$$0 \leq e_2 \leq 0\gamma_2 \tag{5.34}$$

$$0 \leq e_3 \leq 15\gamma_3 \tag{5.35}$$

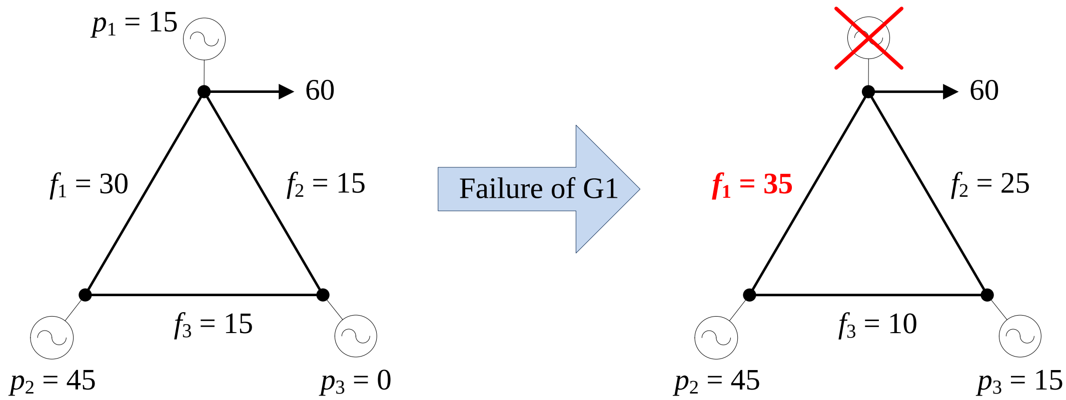


Figure 5.9: Initial solution E_2^0 and the response to the failure of G1. This is the only response that balances supply and demand, but the transmission limit on line L1 is violated.

$$0 \leq m_1 \leq 0(1-\gamma_1) \tag{5.36}$$

$$0 \leq m_2 \leq 0(1-\gamma_2) \tag{5.37}$$

$$0 \leq m_3 \leq 15(1-\gamma_3). \tag{5.38}$$

The optimal solution to the pruning problem is $\gamma_3^* = 2/5$ (γ_1^* and γ_2^* are ignored because neither G1 nor G2 hold any reserve). In other words, there would be no transmission violation if $2/5 \times 15 = 6$ MW of reserve could be moved from G3 to the location of G1 without using any transmission capacity. The response set is updated by $\Gamma_3^1 = (1-2/5) = 3/5$ according to equation (5.19) and the new single-zone reserve requirement is

$$r_1 + r_2 + \frac{3}{5} r_3 \geq p_1 + r_1. \tag{5.39}$$

Figure 5.10 describes the feasible space after applying constraint (5.39). All optimal solutions are reliable and the feasible space aligns with the reliable region for all low-cost solutions. Only a small portion of the feasible space is unreliable in a region that happens to comprise expensive solutions. This example shows that reserve disqualification can be

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^6	0	30	30	30	30	\$3,750
E_2^6	12	42	6	30	18	\$3,750
E_3^6	24	30	6	22	14	\$4,350
E_4^6	24	6	30	14	22	\$4,950
E_5^6	15	0	45	15	30	\$4,875

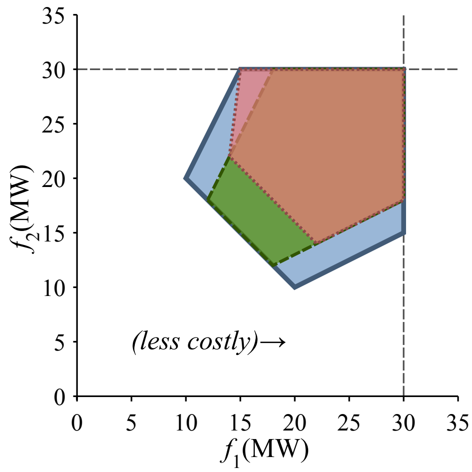


Figure 5.10: Extreme points and feasible space after one iteration of reserve disqualification (dotted border). The graph also shows the initial feasible space and the reliable region (dashed border).

a more precise than traditional reserve requirements. Reserve disqualification is effective in this example because generator injections have a very different effects (PTDFs) on the critical line L1 and the pruning algorithm can differentiate between these resources, even those that fall in the same zone.

The market settlement scheme in (5.20) states that the reserve price for generator G3 will be $3/5$ the price of G2 due to the disqualification. This is an example of the settlement scheme *not* working as intended. If anything, generator G3 should be paid more because it is at a better location. It is at a better location than G2 because it has a smaller PTDF on the critical line L1 (this line is always heavily utilized in optimal solutions). However, no reserve could be disqualified from G2 under the proposed algorithm because the generator provided no reserve in the initial solution. This points to one of the limitations of the proposed algorithm: as a heuristic, it may mischaracterize the deliverability of some reserves. The hope is that the settlements turn out better on the whole.

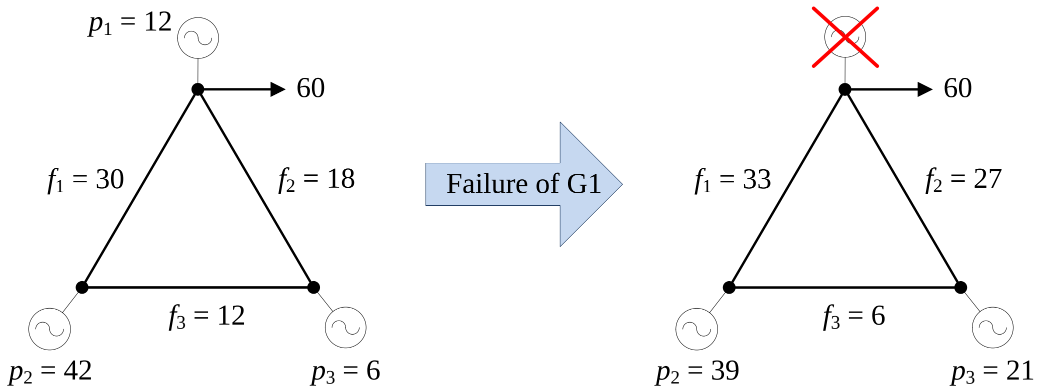


Figure 5.11: Solution E_6^2 and the only feasible response to the failure of G1.

Suppose that the solution is selected to be E_2^6 from Figure 5.10. The only feasible response after G1 fails is shown in Figure 5.11: G3 is the only generator with deliverable reserve. Generator G2 actually has to be ramped down to alleviate congestion on line L1. In practical settings, operators would recognize that reserve from G2 is undeliverable and manually disqualify it. The fact that the disqualification algorithm did not automatically identify this issue suggests that the role of operators is still important to ensure favorable market outcomes, just as it is today.

None of the considered approaches must be applied in isolation. For example, the proposed model combines reserve disqualification with zonal reserve sharing in (5.4)–(5.6) in order to leverage the strengths of both approaches. Both methods may also be combined with Benders' Decomposition when the recourse decisions are convex. The next section looks at Benders' Decomposition applied by itself to a single-zone model.

5.6.3 Benders' Decomposition. Benders' Decomposition may be applied to improve reliability in an iterative manner. This technique allows part of the problem to be modeled indirectly. It is sometimes advantageous for two-stage and multi-stage problems because the formulation can focus on the first stage decisions without directly modeling recourse for different scenarios.

In the case of N-1 reliability, Benders' Decomposition just needs to ensure a feasible response exists for every contingency scenario. A master problem is first solved considering only the first-stage decisions. The incumbent solution from the master is unreliable if the dual of the recourse problem is unbounded. Benders' Decomposition looks for extreme rays that point in unbounded directions and then constrains the master problem so that the identified extreme rays remain bounded. The process continues until the dual recourse problem is bounded, which certifies that the primal problem is feasible due to the strong duality of linear programs (Nemhauser & Wolsey, 1999).

The recourse problem for responding to the outage of G1 is defined in Appendix D. The dual of the recourse problem is

$$\text{Maximize: } 60\pi_1 + 45(\pi_2 + \pi_3) + (p_2 + 15)\pi_4 + (p_3 + 15)\pi_5 + 33(\pi_6 + \pi_7) \tag{5.40}$$

Subject to:

$$\pi_1 + \pi_2 + \pi_4 + \frac{2}{3}\pi_6 + \frac{1}{3}\pi_7 \leq 0 \tag{5.41}$$

$$\pi_1 + \pi_3 + \pi_5 + \frac{1}{3}\pi_6 + \frac{2}{3}\pi_7 \leq 0 \tag{5.42}$$

$$\pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7 \leq 0. \tag{5.43}$$

The first-stage solution from Figure 5.9 (with $p_2 = 45$ and $p_3 = 0$) is not reliable, so it follows that the dual recourse problem must be unbounded. The problem is indeed unbounded along the extreme ray $\pi^* = \left\{1, 0, 0, 0, -\frac{1}{2}, -\frac{3}{2}, 0\right\}$. Benders' Decomposition ensures that the dual of the recourse problem remains bounded along this direction in future iterations by adding the following constraint the master dispatching problem:

$$60\pi_1^* + 45(\pi_2^* + \pi_3^*) + (p_2 + 15)\pi_4^* + (p_3 + 15)\pi_5^* + 33(\pi_6^* + \pi_7^*) \leq 0, \tag{5.44}$$

which simplifies to $p_3 \geq 6$. The new feasible space with this constraint is shown in Figure 5.12. It is visually apparent that the cut removes all optimal solutions that are also unreliable. Benders' Decomposition performs as well as reserve disqualification for this test case because it ensures reliability after just one iteration. Benders' Decomposition is actually more precise because it always applies valid inequalities, which never remove any portion of the reliable region (Nemhauser & Wolsey, 1999). However, this apparent strength can become a weakness for larger problems if it takes many iterations to ensure a reliable solution. Benders' Decomposition is a precise approach that may be encumbered by computational challenges for larger problem instances. The literature shows that solution times can be excessive for unit commitment with transmission-constrained corrective actions [3] . Ferris et al. (2014) show the state of the art *may* allow application in the near future for simpler problems (ones involving a single period and no binary variables) but only after making significant enhancements over vanilla Benders' Decomposition.

The proposed algorithm may find a reliable solution in fewer iterations because it is not limited to valid inequalities. It can instead make more liberal cuts like those shown in Figure 5.13, thereby removing larger portions of the feasible space at each iteration and potentially reducing the number of iterations needed to find a reliable solution. The

[3] See the literature review in Section 3.3.

Extreme Point	p_1	p_2	p_3	f_1	f_2	Objective
E_1^7	0	30	30	30	30	\$3,750
E_2^7	12	42	6	30	18	\$3,750
E_3^7	30	24	6	18	12	\$4,650
E_4^7	30	0	30	10	20	\$5,250
E_5^7	15	0	45	15	30	\$4,875

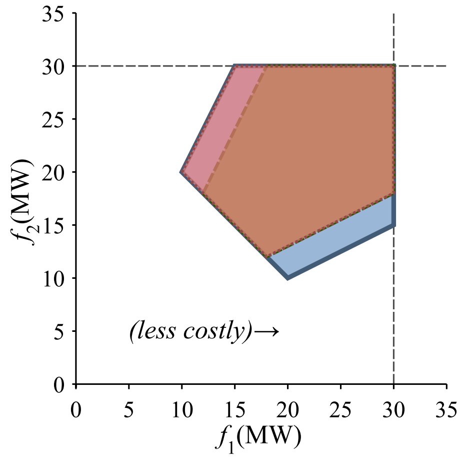


Figure 5.12: Extreme points and feasible space after one Benders' feasibility cut (dotted border). The graph also shows the initial feasible space and the reliable region (dashed border).

reduction in iterations is essential because re-solving unit commitment at each iteration is a time consuming process. The compromise is that response sets do not guarantee optimality.

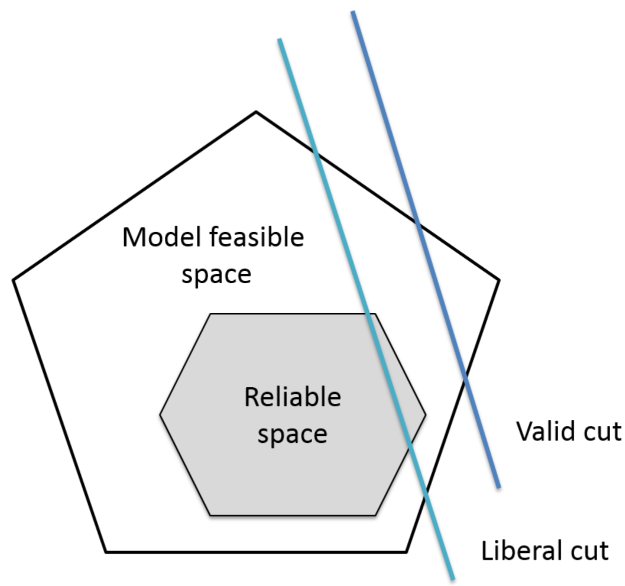


Figure 5.13: Comparison of liberal cuts and valid inequalities. In this example, further iterations are necessary when applying an algorithm that only identifies valid inequalities because all optimal solutions are outside of the reliable space.

5.7 Analysis and Results: IEEE 73-Bus (RTS 96) Test Case

The proposed reserve disqualification procedure is evaluated on seven days across the peak week of the IEEE 73-bus (RTS 96) test case (Grigg et al., 1999). Weekend testing captures the impact of the proposed method on off-peak days. Modifications are made to the test case as described in Section 4.7. This analysis uses 10-minute spinning reserves to respond generator contingencies.

Reserve zones are identified using the partitioning method of F. Wang and Hedman (2012), which uses statistical clustering so that reserve can compensate for disturbances in the same zone with little change to flow on critical lines. Critical lines are identified by solving SCUC for the peak day using a single zone and then performing contingency analysis (the contingency analysis formulation is provided in Appendix C.2). Slack variables

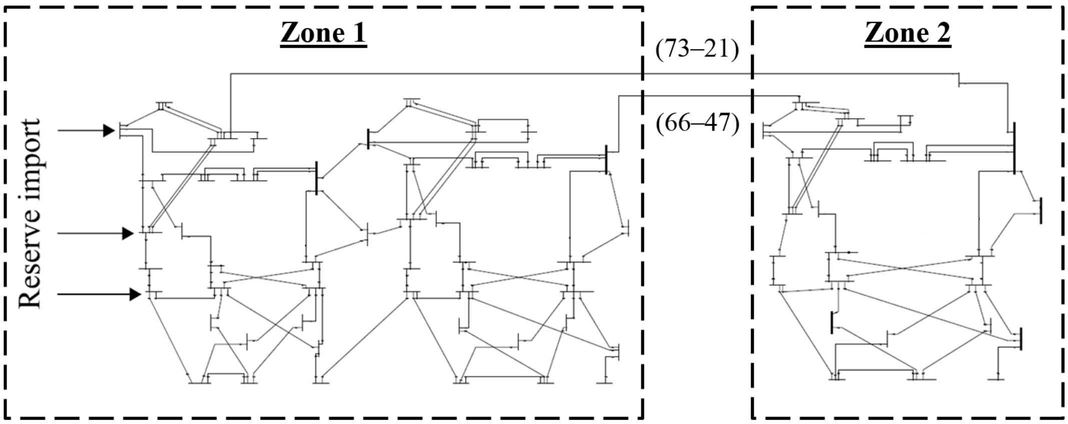


Figure 5.14: Reserve zones for the modified IEEE 73-bus test case with up to 200 MW of reserve imports.

allow transmission violations for when insufficient reserves are deliverable. Line (73–21) is classified as the only critical line because no other line limit is ever relaxed to improve access to reserves. Figure 5.14 shows the resulting zone partition. Zone one comprises nodes (1)–(48) and zone two comprises nodes (49)–(73). Intra-zonal congestion is rarely seen in either zone but interface line (73–21) often limits reserve imports into zone two. This analysis will demonstrate that reserve disqualification can still be effective when only one line is congested.

Testing is performed using CPLEX v12.5 on an 8-core 3.6GHz computer with 48 GB of memory. Unless otherwise stated, SCUC is terminated after five minutes or upon reaching an optimality gap of 0.1%. The model includes 99 generators and 24 hours, which translates to 2376 response sets. The reserve bids are assumed to all be zero because bid data is not published for the RTS 96 test case. This assumption is aligned with some markets (e.g., the real-time market in ISO-NE (Entriken & DePillis, 2009)) but different from others that allow positive reserve bids [4] .

5.7.1 Baseline Reserve Requirements.

The baseline reserve requirements (5.1)–(5.3) oblige the user to specify reserve sharing limits. This analysis tests several different policies and compares them post hoc. Although operators do not have time to compare many alternatives in the midst of making day-to-day decisions, this strategy allows the proposed approach to be evaluated alongside both conservative and liberal reserve sharing policies.

Constraints (5.45)–(5.47) are introduced to limit reserve sharing between zones. The sharing limit between each zone and itself is an arbitrarily large value M, which should exceed reserves to avoid imposing an artificial limit on reserve availability within zones. The emergency capacities for the zonal interface, which comprises lines (73–21) and (66–47), sum to 1250 MW. The sharing limit between the adjacent zones is defined in relation to the import headroom on these lines. The parameter $\alpha \in [0,1]$ de-rates the sharing capability. This is appropriate because it may be impossible to utilize every component of the interface to its full extent.

$$S^1_{1t} = S^2_{2t} = M, \qquad \forall t \in T \tag{5.45}$$

$$S^2_{1t} = \alpha 1250 - f_{(73\text{–}21),t} - f_{(66\text{–}47),t}, \qquad \forall t \in T \tag{5.46}$$

$$S^1_{2t} = \alpha 1250 + f_{(73\text{–}21),t} + f_{(66\text{–}47),t}, \qquad \forall t \in T \tag{5.47}$$

[4] Many ISOs allow reserve bids in the DAM. While there is no appreciable physical cost for holding reserves, positive bids are allowed to help generators regain their lost opportunity costs from not selling their capacity as energy. The motivation is to seek an equilibrium where the market prices naturally incentivize participants to follow dispatch instructions from the ISO.

Table 5.1: Baseline results averaged over the 7 test days.

Policy (α)	Cost	Time (mins)	Optimality gap	E[viol] (MW)	# viol	max viol (MW)
45%	$ 2,393,971	4.8	0.17%	0.004	0.9	1.9
50%	$ 2,382,857	4.8	0.16%	0.032	2.9	5.8
55%	$ 2,374,080	3.8	0.13%	0.046	5.1	10.2
60%	$ 2,370,421	3.9	0.12%	0.208	12.9	28.2
65%	$ 2,369,969	3.1	0.11%	0.385	17.3	42.8
stochastic	$ 2,374,328	60.2	0.17%	0	0	0

Five reserve sharing policies are tested corresponding to $\alpha = \{0.45, 0.50, 0.55, 0.60, 0.65\}$. Table 5.1 summarizes the average results over the seven test days. Conservative reserve sharing policies (small α) result in higher operating costs but tend to be more reliable. Three statistics measure reliability: "E[viol]" is the expected sum of transmission violations from contingency analysis, where contingency probabilities are inferred from the mean times to failure; "# viol" is the number of contingencies with some violation; and "max viol" is the maximum sum of transmission violations for any contingency. Sections 5.7.2 and 5.7.3 evaluate the use of reserve disqualification to eliminate these violations when the initial solution is not reliable.

Table 5.1 also summarizes results from the deterministic equivalent to a stochastic model that ensures feasibility across contingency scenarios. This problem was solved to obtain proper optimality gaps for the proposed algorithms. This extensive form SCUC explicitly models corrective actions for a subset of contingencies [5] . A single reserve zone requires reserve to exceed the largest contingency and corrective actions are modeled to

[5] The formulation is a combination of the unit commitment and contingency analysis models in Appendix C with a very high penalty placed on contingency violations.

ensure reserve deliverability. The stochastic model identifies a solution that is about 0.8% cheaper than the most reliable baseline policy.

Note that solution times for stochastic SCUC are prohibitively long and the extensive form model does not scale well for large systems. The reported solution time of 60 minutes in Table 5.1 was only obtained by limiting contingencies to those that failed contingency analysis at some point during the proposed decomposition algorithm. Without such prior information, the solution times are generally much longer.

5.7.2 Disqualification as an Out-of-Market Correction. Reserve disqualification is evaluated as an ex-post mechanism to improve the reliability of the DAM solution. Such procedures have been classified in this chapter as out-of-market corrections (OMCs). To be consistent with market rules, all generators that are committed by the DAM model must stay committed. Therefore, this particular application of the algorithm from Figure 5.5 will not de-commit any units that were committed by the market model.

Two different reserve disqualification strategies are considered. The first approach applies the proposed algorithm and is referred to as generalized reserve disqualification (GRD). The second approach disqualifies reserves uniformly across contingencies and is referred to as traditional reserve disqualification (TRD). The pruning algorithm from Section 5.4.1 is adjusted to accommodate TRD. The first adjustment discharges the idea of partial disqualification by adding an integrality constraint on γ. The response sets are then pruned one contingency at a time and global disqualification is enforced by specifying that resources disqualified from one contingency may not offer reserves for any other contingency during the same period. These amendments provide an automated approach that mimics the manual reserve disqualifications used by operators today.

Figure 5.15 illustrates the progress of the reserve disqualification procedures on the highest and lowest load days (Tuesday and Sunday) of the peak week. SCUC determines the operating costs (x-axis) and contingency analysis measures the reliability (y-axis). The red diamonds represent DAM solutions with varying reserve sharing policies (α): small values of α tend to be more reliable but are also more expensive because they require additional reserve within local areas. The GRD and TRD algorithms start with these initial solutions, disqualify reserves, and solve SCUC again with reserves procured from different locations. The dotted lines represent the progress at each iteration. Both procedures converge to N-1 reliable solutions that are economical compared to the most reliable DAM solution. This suggests that reserve disqualification adds robustness to the baseline policy because operators can recover from unreliable solutions and still maintain reasonable operating costs. Figure 5.15 also demonstrates that GRD consistently outperforms TRD, suggesting that it is beneficial to define different response sets for distinct contingencies.

At first glance, it may seem counterintuitive that an off-peak day would require a more conservative reserve sharing policy to achieve the same level of reliability. The reason for this phenomenon is that generation serving the base load creates relatively little flow on interface line (66–47). The import capability estimate into zone two is deceptively large because reserve imports are limited by the bottleneck on line (73–21) while line (66–47) remains relatively underutilized. The underlying issue is the same as in Figure 5.4. The

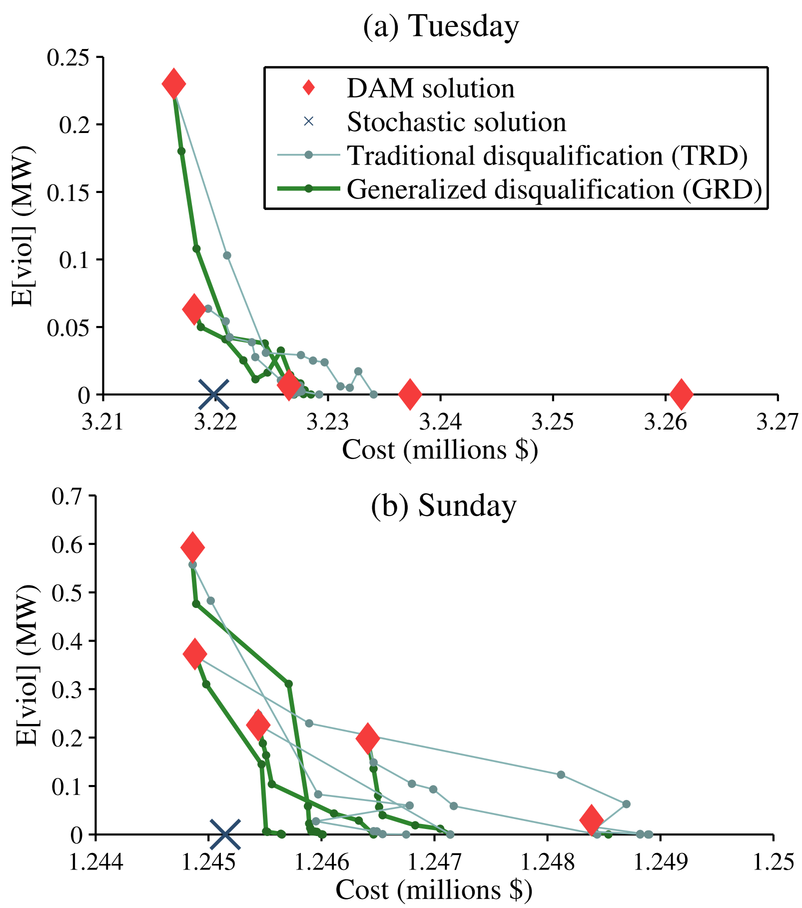


Figure 5.15: Progress of iterative reserve disqualification when applied to unreliable DAM solutions.

accuracy of reserve sharing is not driven by the amount of congestion but how well the reserve sharing model anticipates the influence of congestion on corrective actions. In this case, reserve disqualification is effective on low load days because it encourages reserve to be held at locations that favor underutilized paths. Figure 5.16 shows that reserve proximal to the heavily utilized line (73–21) is disqualified from the response set for a large contingency in zone two.

Figure 5.17a and Figure 5.17b describe the final costs (after OMCs have been applied) relative to the stochastic solution. The solutions derived using GRD are more

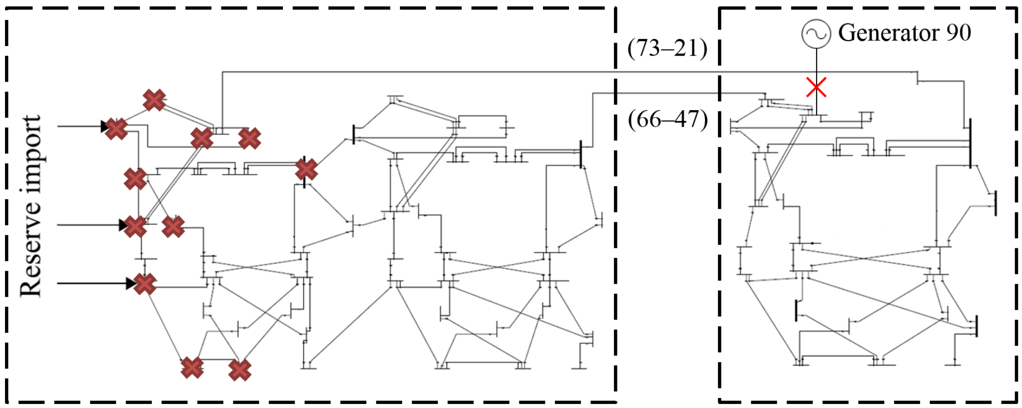


Figure 5.16: Locations with reserves disqualified from the response set for generator 90 on hour four of Sunday, day 357. Line (73–21) is the most heavily utilized.

consistently close to the lower bound obtained from the stochastic model. Conservative reserve sharing policies ($\alpha = 0.45$ and 0.50) often underperform because SCUC obtains an overly conservative solution at the expense of economics. These results suggest that it may be more economical to acquire a risky initial solution (large α) and then improve reliability through OMCs instead of starting off with an overly conservative reserve sharing policy.

5.7.3 Disqualification for the Day-Ahead Market. In Section 5.7.2, reserve disqualification was applied after the market model terminated. In this section, the proposed algorithm is tested as a means to clear the DAM itself. Figure 5.17c shows that the final costs are mostly lower when reserve disqualification is applied within the DAM, which can be attributed to there being no restrictions on generator commitments across iterations. It is generally preferable to apply this algorithm within the DAM when time allows.

Table 5.2 summarizes the total time and the number of iterations to converge to a reliable solution. The majority of time is spent solving SCUC and relatively little effort is spent on contingency analysis and pruning response sets. The proposed algorithm can

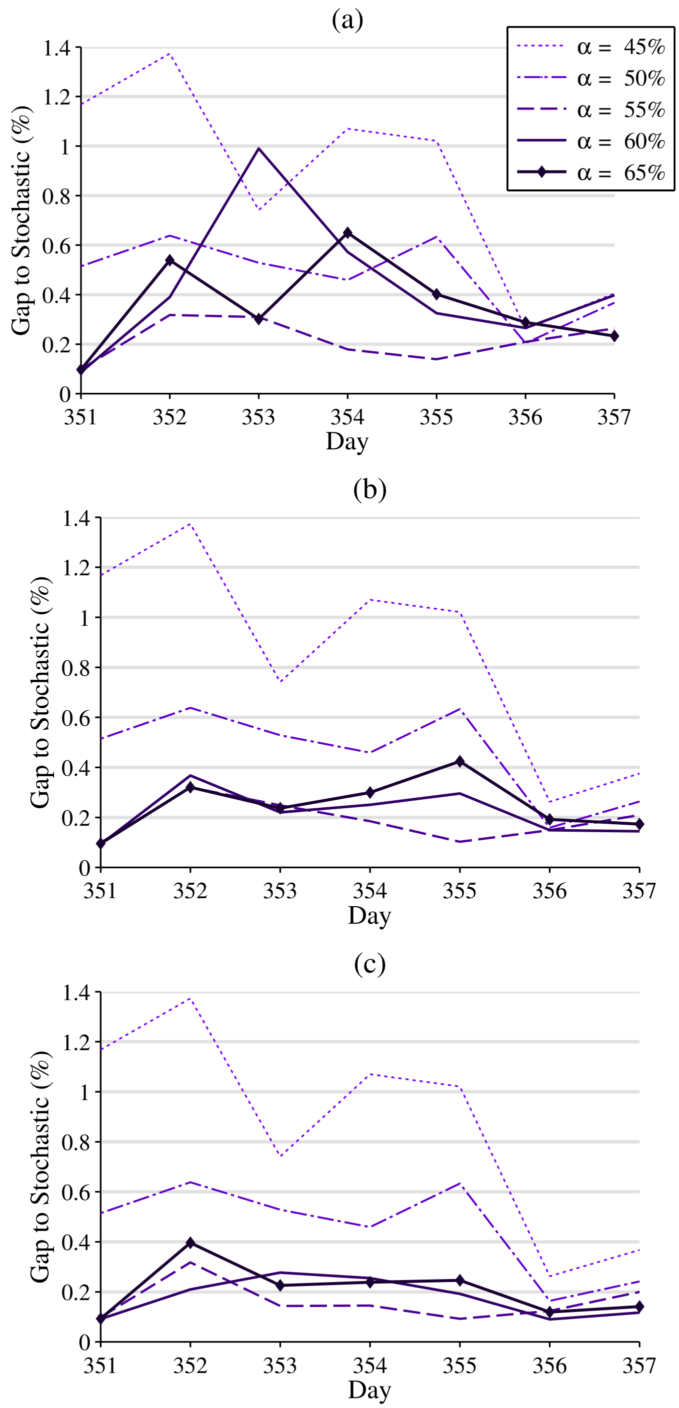


Figure 5.17: Final costs for N-1 reliable solutions relative to the stochastic model. (a) OMCs using traditional reserve disqualification. (b) OMCs using generalized reserve disqualification. (c) DAM using generalized reserve disqualification.

Table 5.2: Computing statistics for the 7 test days.

Policy (α)	# iterations			Time (min)		
	Min	Median	Max	Min	Median	Max
45%	1	1	8	2.6	5.2	33.9
50%	1	1	10	5.1	5.2	37.0
55%	1	7	20	1.1	32.5	104.7
60%	6	11	18	21.0	41.5	71.2
65%	5	11	18	15.1	33.7	97.2

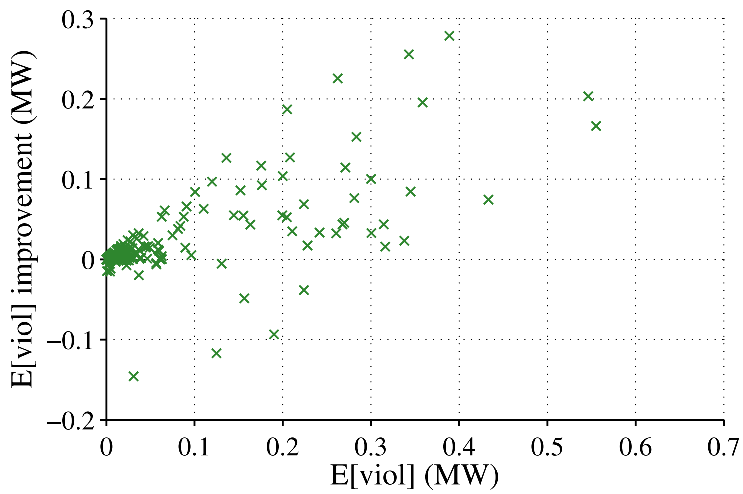


Figure 5.18: Reliability vs. reliability improvement at each iteration.

take 20 iterations to eradicate all violations. Such worst-case convergence times may be unacceptable for some operators. Fortunately, the first iterations tend to be the most influential because pruning is more aggressive when violations are abundant. Figure 5.18 plots E[viol] against the improvement to reliability at each iteration. The biggest reliability gains are obtained in early iterations, followed by a long tail of small changes as E[viol] approaches zero. Figure 5.18 shows that the majority of iterations occur after the incumbent solution is already nearly reliable.

Table 5.3: Would-be average DAM payments to reserve providers before and after reserve disqualification.

Policy (α)	Initial solution	Final solution
45%	\$ 17,591	\$ 16,958
50%	\$ 14,898	\$ 14,611
55%	\$ 15,481	\$ 14,241
60%	\$ 20,716	\$ 16,036
65%	\$ 17,164	\$ 21,808
Overall	\$ 17,170	\$ 16,731

The aggregate markets settlements are summarized in Table 5.3 relative to the first and last iterations of the decomposition algorithm. The market initiates with zonal prices and evolves as reserves are disqualified. The average sum of payments, $\sum_{g\in G}\sum_{t\in T}\phi_{gt}$, does not change greatly as reserves are disqualified. This is important because some market participants may question how aggregate payments will change. Will payments go up since resources are compensated for many individual contingencies? The empirical evidence suggests that the answer is "not significantly." Although the reserve prices are derived from multiple contingencies, most response sets have a surplus of reserve: no more than 27 of the 2376 reserve requirements in (5.4) have positive dual variables during any single day. There are relatively few contingencies that dominate the settlements and the overall payments do not necessarily increase relative to zonal markets.

The primary motivation is not to influence the aggregate payments, but to develop prices that are better at reflecting the quality of service provided by individual suppliers. An example of the locational prices is as follows for hour four of day 357. The considered reserve sharing level is $\alpha = 0.65$, which requires many reserve disqualifications to ensure reliability (see Figure 5.16 for the locations of disqualified reserve for one particular contingency during this period). The market results are elaborated below because this period has more reserve disqualifications than any other. Figure 5.19(a) shows the shadow prices for transmission capacity in the final market solution. Three lines are congested before any contingency occurs: lines (16–17), (21–22), and (73–21). Reserve deliverability is generally harder to ensure in the proximity of these lines.

Figure 5.19(b) shows the reserve locations and prices in the market solution. Reserve in the rightmost zone is valued high because it can serve more contingencies without overloading transmission (reserve that flows from right to left across line (73–21) counters congestion). Reserve in the leftmost zone has three unique prices. The prices are generally lower near congested lines. The settlement scheme still provides some payment to resources near these lines because the model recognizes the reserve is deliverable for a subset of contingencies. This example demonstrates what would be expected of locational pricing: the procedure rewards market participants that are at key locations relative to congested lines. Traditional reserve zones do not provide this level of precision. Another example of locational prices for a different test case is shown by Lyon, F. Wang, et al. (2014).

5.7.4 Practical considerations. Many ISOs already use iterative approaches to clear the DAM, and it may be practical to incorporate response sets with minimal changes to existing procedures. The number of iterations is a key factor because time is typically limited when clearing the DAM. The results in Table 5.2 show that a large number of iterations may be necessary before a reliable solution is reached.

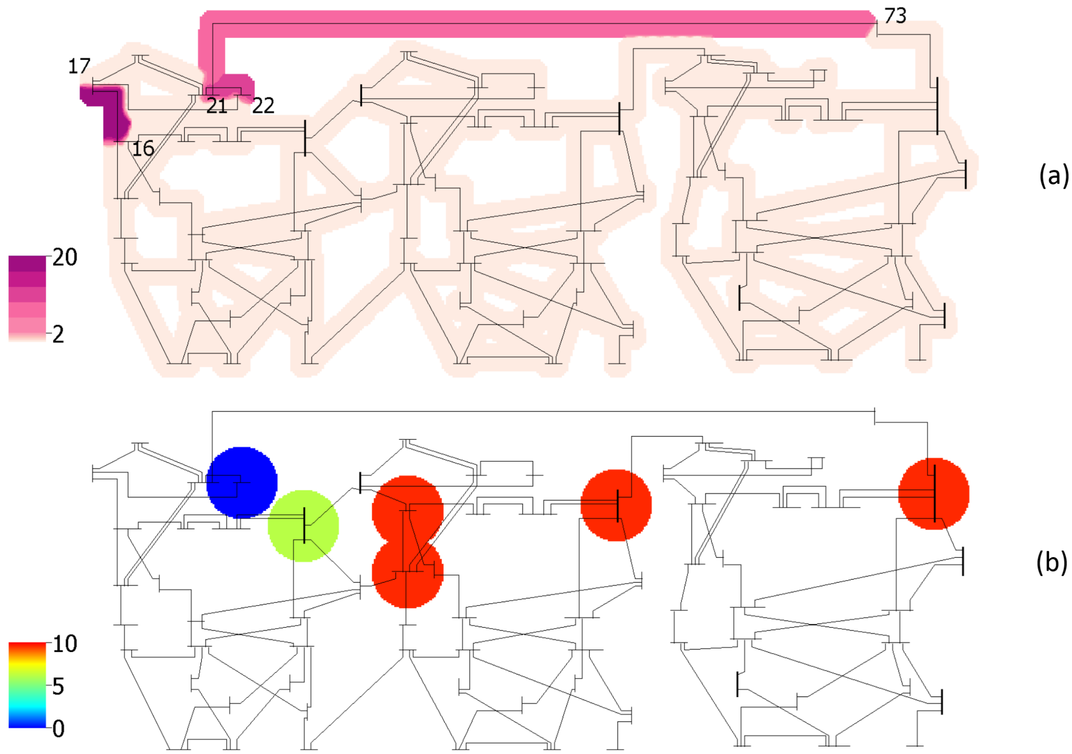


Figure 5.19: Locational market results during hour four of Sunday, day 357. (a) Transmission capacity shadow prices. (b) Reserve locations and prices (excluding reserve imports).

Several options are available when time is scarce. First, operators may adopt conservative baseline reserve requirements to encourage the initial solution to be more reliable. Second, operators may use off-line analysis to identify initial response sets that require fewer updates. Finally, OMCs may still be applied when a reliable solution is not found within the DAM time limit. The proper balance between these strategies depends on the time restrictions, the average performance of the baseline model, and the operator's tolerance for relying on OMCs.

The full market implications of the proposed algorithm are not yet clear. Any market change of this magnitude should be thoroughly vetted to determine the impact on bidding behavior and market outcomes. The good news is that the approach only overrides the baseline zonal model when the solution is unreliable. The approach can, therefore, be given a gradually more prominent role by slowly relaxing the conservative reserve requirements used today. This would allow for a trial period to evaluate the effects of the proposed approach before making large changes to market operations.

5.8 Conclusion

Transmission constraints can make reserves unavailable in ways that are hard to anticipate before the SCUC solution is available. As a result, operators disqualify reserves that end up falling behind unanticipated transmission bottlenecks. Such out-of-market changes can be expensive because they are manually implemented and not co-optimized with the DAM decisions.

The proposed reserve disqualification procedure can be used as a heuristic for clearing the DAM or as an improved OMC mechanism. Reserves are disqualified on a per-contingency basis to address the distinct congestion patterns that may arise following different contingencies in different periods. The decomposition algorithm solves SCUC and bases reserve disqualification on system conditions from the incumbent solution. The model is able to better characterize scarce resources and thereby improve economics and promote better prices. This algorithmic approach also improves upon traditional OMCs where operators manually disqualify reserves. Testing on the RTS 96 test case demonstrates reduced costs to protect against generator contingencies, particularly when the reserve locations within the same zone have dissimilar effects on critical lines.

With the upcoming advent of intermittent renewable resources, transmission bottlenecks will become harder to predict and traditional models will be less justifiable because the projected costs under forecasted conditions may poorly reflect the realized costs. New optimization tools and algorithms are appropriate to determine solutions that are not only reliable but cost effective. The work in the next chapter extends the concept of response sets to help mitigate uncertainty from intermittent resources.

Chapter 6

CAPACITY DISQUALIFICATION FOR WIND UNCERTAINTY

6.1 Introduction

The focus of this dissertation, thus far, has been on ensuring reliability across contingency scenarios. The cost of contingency response has been ignored because severe contingencies happen infrequently (the hourly probabilities for generator and transmission contingencies are around 10^{-3} and 10^{-4} (Grigg et al., 1999)). Ignoring scenario costs has been a reasonable limitation for deterministic models because critical contingencies are rare enough to hardly impact expected costs. However, scenario costs will become more important in the future as growing wind energy increases the likelihood of large forecast deviations.

Renewable portfolio standards in parts of the US are requiring up to 33% of energy to come from renewable resources, such as wind, by the year 2020 (California Public Utilities Commission, 2013). Large forecast deviations are inevitable because wind availability remains within 20% of the day-ahead forecast only 80% of the time (Hodge, Florita, Orwig, Lew, & Milligan, 2012). Deeper wind penetrations will require more aggressive rebalancing and, consequently, operating costs are likely to become more volatile. This increases the motivation for capturing scenario costs within SCUC to identify solutions that are economical across a range of scenarios. This chapter extends the idea of response sets from Chapter 5 in order to limit operating costs across scenarios instead of solely targeting feasibility.

Stochastic models have been proposed for unit commitment to minimize cost across scenarios. Two-stage stochastic programming is one such tool for dealing with wind uncertainty (L. Wu et al., 2007, 2008; Ruiz et al., 2009; Papavasiliou, Oren, & O'Neill, 2011; Morales et al., 2009; Papavasiliou & Oren, 2013). These models target economical commitment decisions. They make commitment decisions in the first stage and assume all other decisions can be deferred until after the wind scenario is revealed. Two-stage robust optimization has also been applied in a similar way (Jiang et al., 2012, 2014; Bertsimas et al., 2013). It is currently being investigated for practical application by a major software (Alstom Grid) vendor to change commitments during the operating day (Q. Wang, X. Wang, et al., 2013), but there is no evidence of day-ahead market applications at this time. Two-stage stochastic programming and robust optimization require computational advancements before becoming tractable for day-ahead markets, where there are many binary commitment decisions making the problem harder to solve.

Energy markets today use deterministic models and protect against uncertainty using reserve and capacity requirements. Sioshansi and Short (2009) and Papavasiliou and Oren (2013) require reserve to exceed a prescribed percentage of the wind forecast. Doherty and M. O'Malley (2005) and Vos and Driesen (2014) require reserve to cover likely wind fluctuations [1] . Practitioners like CAISO and ISO-NE also commit additional capacity

[1] Incidentally, the reserve requirements derived by Vos and Driesen (2014) are nearly equivalent to the alternative approach of simply requiring reserve to cover a percentage of the wind forecast (see Figures 6 and 9 in the reference).

within the market model for reliability (CAISO, 2010; Parent, 2013a). These procedures do not guarantee enough reserve will be deliverable, so more capacity may be later procured manually to repair unreliable market solutions. Formalizing this repair process is one of the potential applications being considered for robust optimization. However, the most popular robust optimization models do not consider the capability to change generator commitments moving closer to real time [2]. Having the option to change the commitment status of certain *fast* generators can improve expected costs by 0.1%–2% according to some stochastic wind-integration studies (Tuohy, Meibom, Denny, & O'Malley, 2009; J. Xiao et al., 2011). The approach proposed in this chapter considers the ability to commit fast generators when determining capacity requirements.

The proposed method augments capacity constraints for day-ahead SCUC. The problem is challenging because capacity may be:

- expensive to dispatch,
- unable to ramp fast enough,
- or undeliverable due to transmission constraints.

A preferred approach would acquire capacity that is sufficiently inexpensive, fast, and deliverable. The proposed approach follows these principles while avoiding the computational limitations of stochastic programming and robust optimization. The approach

[2] Most solution methods being investigated for robust optimization require the recourse problem to be convex because they use dual information (See Section 3.3.2). There is an exception in recent work by L. Zhao and Zeng (2012b), who use a primal decomposition approach. The method can model integer recourse decisions exactly but still needs further development to improve scalability (L. Zhao & Zeng, 2012b).

anticipates what capacity can contribute to inexpensive scenario response. The scenario response may be *expensive* if it relies on inefficient generators or if it cannot satisfy demand without violating some other constraint (demand imbalances are penalized in the objective function at a high cost because they can lead to curtailing customer loads). This is an extension over existing deterministic reserve requirements that only target satisfying demand across scenarios. The ability to consider production cost for different scenarios is a contribution that takes a step closer to stochastic models.

The proposed algorithm solves SCUC using a decomposition structure similar to the one used in Chapter 5 [3] . The first stage solves a nominal SCUC model with capacity requirements for individual wind scenarios – each requirement must be satisfied by capacity from the response set for the respective scenario. The second stage identifies capacity that should not have been treated as available because it is too expensive, slow, or undeliverable. The algorithm disqualifies this capacity so that more favorable resources will be committed in subsequent iterations.

The results show that capacity response sets can improve expected costs relative to existing reserve policies from the literature. The cost improvement is about 0.5% for the IEEE 73-bus test case. The computational performance is also promising relative to results reported for state-of-the-art stochastic models. Therefore, the proposed approach appears to provide a trade-off between the precision of stochastic models and the computational tractability of existing deterministic models. Most of the performance comes from reducing

[3] The approaches from Chapters 5 and 6 may even be applied simultaneously if desired, one addressing contingencies and the other addressing wind uncertainty.

load imbalances through anticipating recourse constraints (like transmission and ramp limits). This underscores the relevance of alternative approaches that seek only reliability. The premium for reliability notwithstanding, the more general framework of incorporating production costs makes the work consistent with stochastic models and allows the potential to exploit production cost improvements when such opportunities exist.

Another observation is that upfront commitment decisions are less important when fast generators can change their commitment status. Fast generators can significantly reduce the value of perfect information, to the point that advanced models may offer little improvement to expected operating costs. This observation corroborates the findings of Tuohy et al. (2009) and J. Xiao et al. (2011). The proposed approach can consider fast generators for the sake of generality and to provide operational continuity in changing systems (e.g., those moving towards a greater number of fast generators).

The chapter outline is as follows. Section 6.2 formulates the stochastic SCUC problem. Section 6.3 describes the form of the proposed capacity requirements, which replace the direct modeling of scenarios in the stochastic formulation. Section 6.4 presents the decomposition algorithm. Section 6.5 describes the methodology used to generate wind scenarios for testing. Section 6.6 reports numerical results for the IEEE 73-bus test case and describes strategies to improve performance. Finally, Section 6.7 concludes and summarizes the potential for future work.

6.2 Stochastic Unit Commitment

The stochastic SCUC problem (SUC) commits generators while considering several different scenarios that may occur. As such, many variables and parameters are now indexed by scenario s. This is a two-stage problem that makes base commitment decisions upfront (designated here by $s = 0$) and evaluates the potential response for different scenarios. The objective is to minimize a function of cost $f(x)$ across scenarios. The problem definition is

$$(SUC): \text{ Minimize } f(x) \tag{6.1}$$

Subject to:

Wind availability:

$$k^s_{nt} \leq K^s_{nt} \qquad \forall n \in N, t \in T, s \in S \tag{6.2}$$

Network constraints:

$$\sum_{n \in N} i^s_{nt} = 0, \qquad \forall t \in T, s \in S \tag{6.3}$$

$$i^s_{nt} = \sum_{g \in G(n)} p^s_{gt} + k^s_{nt} + b^s_{nt} - D^s_{nt}, \qquad \forall n \in N, t \in T, s \in S \tag{6.4}$$

$$-F_l \leq \sum_{n \in N} PTDF_{nl} i^s_{nt} \leq F_l, \qquad \forall l \in L, t \in T, s \in S \tag{6.5}$$

Generator capacity and ramping constraints:

$$P^{\min}_g u^s_{gt} \leq p^s_{gt} \leq P^{\max}_g u^s_{gt}, \qquad \forall g \in G, t \in T, s \in S \tag{6.6}$$

$$-R_g \leq p^s_{gt} - p^s_{g,t-1} \leq R_g, \qquad \forall g \in G, t \in T, s \in S \tag{6.7}$$

Minimum up and down times:

$$v^s_{gt} - w^s_{gt} = u^s_{gt} - u^s_{g,t-1}, \qquad \forall g \in G, t \in T, s \in S \tag{6.8}$$

$$\sum_{q=t-UT_g+1}^{t} v^s_{gq} \le u^s_{gt}, \qquad \forall g \in G, t \in T, s \in S \tag{6.9}$$

$$\sum_{q=t-DT_g+1}^{t} w^s_{gq} \le 1 - u^s_{gt}, \qquad \forall g \in G, t \in T, s \in S \tag{6.10}$$

Nonanticipivity constraints:

$$u^s_{gt} = u^0_{gt}, \qquad \forall g \notin G^{\text{fast}}, t \in T, s \in S \tag{6.11}$$

Reserve requirements (not shown), e.g., contingency reserves

Variable domains:

$$u^s_{gt} \in \{0,1\}, 0 \le v^s_{gt}, w^s_{gt} \le 1, \qquad \forall g \in G, t \in T, s \in S \tag{6.12}$$

$$k^s_{nt}, b^s_{nt} \ge 0, \qquad \forall n \in N, t \in T, s \in S. \tag{6.13}$$

Commitment decisions are coupled with an optimal power flow that anticipates how generators will be dispatched for each scenario. The objective (6.1) minimizes a given cost function $f(x)$, where x is the vector of decision variables. Alternative objective functions are described in Section 6.2.1. The only difference between scenarios is a limit on the available wind production in (6.2), where k^s_{nt} is wind production and K^s_{nt} is the available wind power at node n during period t.

Network constraints model power balance (6.3), locational injections (6.4), and transmission capacity limits (6.5). The power flows are modeled using power transfer distribution factors, where $PTDF_{nl}$ is the sensitivity of flow across line l to injections at node n. This type of linear approximation is common in power system scheduling (Stott et al., 2009). If network constraints are proving difficult to satisfy, then the flow at node n can be balanced via the dummy variable b^s_{nt} at a penalty to the objective function. Different penalty costs for load imbalance are discussed in Section 6.2.1.

Conventional generator constraints model capacity (6.6), ramping capability (6.7), and minimum up and down times (6.8)–(6.10). A description of these constraints can be found in Section 2.2.3.

This model is formulated as a two-stage decision problem. The first stage positions generators for base conditions (designated as $s = 0$) before the scenario becomes known. These decisions influence the second stage through the nonanticipivity constraints (6.11), which specify that only fast generators (in G^{fast}) may change commitment status for scenario response. Two-stage models, such as this, assume uncertainty is revealed at a single point in time. More precise models in the literature can include three or more stages (L. Wu et al., 2007, 2008); however, multi-stage models are harder to solve and detailed predictions can be difficult to make outside of simulating operations on a fine time scale over a rolling horizon (Ela & M. O'Malley, 2012). The intent of stochastic models is to anticipate operating cost well enough to make informed first-stage decisions.

The first stage determines what capacity to commit from slow generators. All other decisions can be deferred until after the scenario is revealed. The same two-stage structure is used by Ruiz et al. (2009), Papavasiliou et al. (2011), Morales et al. (2009), Papavasiliou and Oren (2013), Bertsimas et al. (2013), and Jiang et al. (2014), although some of these references do not model fast generators (effectively assuming $G^{\text{fast}} = \varnothing$). The potential benefit of stochastic programming depends on how restrictive the first-stage decisions are, e.g., the upfront commitment decisions are more important when fewer fast generators are available to change their commitment status in response to uncertainty.

Problem (SUC) is deterministic when $S = \{0\}$, which effectively removes all second-stage scenarios and makes the problem much smaller and easier to solve. The downside is that optimal solutions to this simplified problem may perform poorly for one or more scenarios. As described in the introduction, deterministic formulations will often procure reserves or backup capacity to compensate for this loss of precision. One such policy is the basis of this research.

SCUC may have more features than detailed by (6.1)–(6.13), e.g., storage, flexible demand, and interface limits. Reserve requirements may still be used to cover forms of uncertainty not included in the scenarios. Different objective functions are discussed in Section 6.2.1.

6.2.1 Objective Function. Production costs for committed generators are frequently modeled using convex functions because fuel efficiency tends to decrease with production (Wood & Wollenberg, 1996). For simplicity, linear marginal costs are used here. Parameter c_g is the marginal cost per unit of energy produced by generator g. Fixed costs c_g^{NL} and c_g^{SU} (no-load and start-up) are incurred whenever the generator is committed or started. Load imbalance is also allowed at a high penalty ν. Taking everything together, the total cost for a single scenario s is

$$c^s(x) = \sum_{g \in G} \sum_{t \in T} \left(c_g p_{gt}^s + c_g^{\mathrm{NL}} u_{gt}^s + c_g^{\mathrm{SU}} v_{gt}^s \right) + \nu \sum_{n \in N} \sum_{t \in T} b_{nt}^s. \tag{6.14}$$

The parameter ν is a linearized penalty for violating node balance constraints. This would translate into physical load shedding if the imbalance were to propagate into real-time. For this reason, proper values for ν are bounded by the value of lost load ($\nu \leq$ VOLL). The true cost of curtailing load depends on the type of customer, time of day, and the duration of the outage. Reasonable estimates can put VOLL anywhere between \$2000/MWh and \$25,000/MWh for most customers during outages that last one hour (Sullivan, Mercurio, Schellenberg, & Freeman, Sullivan & Co., 2009). Traditional load shedding is not the only interpretation. With the evolution of the smart grid, more customers may voluntarily curtail themselves in order to avoid high prices or to receive compensation from the system operator. Such demand response actions can be modeled mathematically as load imbalances relative to the forecast.

Load imbalances do not have to represent physical load reductions. Operators may be able to reduce the expected load shedding by adjusting the unit commitment solution after the day-ahead market has cleared. The adjustment period has the advantage of modeling intermediate commitments of semi-fast generators, i.e., generators that are not classified as fast but may still be "fast enough" to commit after the day-ahead market has closed. Such commitments occur through the *look-ahead unit commitment* in MISO and the *reserve adequacy assessment* in ISO-NE (MISO, 2012a; ISO-NE, 2014c). These commitment decisions are difficult to model in two-stage formulations that categorize everything as either day-ahead or real-time. Bertsimas et al. (2013) allow v to be the anticipated cost of repairing unreliable solutions after the fact (they choose a value of \$5000/MWh). Given the many interpretations of v, its use in this chapter is only intended as a rough penalty for not satisfying all constraints in the modeled scenarios.

Costs can be combined in various ways to create different (SUC) objectives, such as

$$f(x)^{\text{stoch}} = \sum_{s \in S} \pi^s c^s(x) \tag{6.15}$$

$$f(x)^{\text{robust}} = \max_{s \in S} \{c^s(x)\} \tag{6.16}$$

$$f(x)^{\text{base}} = c^0(x). \tag{6.17}$$

Equation (6.15) is a stochastic programming objective that weighs the cost of individual scenarios by probability π^s. Equation (6.16) is a robust objective that measures the most expensive scenario in the uncertainty set. Equation (6.17) is a deterministic objective that only measures cost of first-stage decisions (the base case). Deterministic models neglect second-stage costs and may perform poorly across scenarios. However, their performance may be improved by committing backup capacity for scenario response. The proposed model uses objective (6.17) because only the base scenario $s = 0$ is modeled directly in SCUC. The costs across other scenarios are addressed by proxy through capacity constraints that are derived on an iterative basis.

6.3 Capacity Constraints

Operators do not apply (SUC) to day-ahead markets today because it is too hard to solve in the alloted time. Instead, they protect against uncertainty using deterministic policies like reserve and capacity constraints. Similar constraints have been used since the 1960s, when PJM adopted probabilistic capacity requirements for unit commitment (Anstine et al., 1963). CAISO and ISO-NE now include capacity constraints in their day-ahead markets (CAISO, 2010; Parent, 2013b). Without such policies, the system may be exposed to shortages that drive up costs and threaten reliability. Their inclusion in day-ahead markets can lead to more efficient decisions. There is also a market design benefit because capacity requirements can improve convergence between day-ahead and real-time prices by addressing reliability needs in the market model and limiting the number of out-of-market adjustments necessary to repair unreliable solutions (CAISO, 2010; Parent, 2013a).

The capacity requirement used by CAISO (2010) is equivalent to

$$\sum_{g \in \tilde{G}(j)} \mathscr{C}_{gt} u_{gt}^{0} \geq Q_{tj}, \qquad \forall t \in T, j \in J, \tag{6.18}$$

where $\mathscr{C}_{gt}$ is qualified capacity from generator g and Q_{tj} is the minimum required capacity in area j during period t. These parameters are tuned by the system operator to provide the desired amount of protection. Capacity requirements may be applied system wide or to local areas $j \in J$, where $\tilde{G}(j)$ comprises generators in the respective area (CAISO, 2010). The proposed approach adopts a similar requirement to procure capacity that is useful for scenario response.

The augmented requirement used in this chapter is

$$\sum_{g \in G} \mathscr{C}_{gt}^{s} u_{gt}^{0} \geq Q_{t}^{s}, \qquad \forall s \in S, t \in T, \tag{6.19}$$

where

$$Q_{t}^{s} = \sum_{n \in N} (D_{nt}^{s} - K_{nt}^{s}), \qquad \forall s \in S, t \in T. \tag{6.20}$$

The primary difference from (6.18) is the new constraint is indexed by scenario s. Constraint (6.19) says committed capacity must cover demand minus wind production. This is a minimal requirement. Acquiring an additional margin by making $\mathscr{C}_{gt}^{s}$ smaller is usually prudent to provide more protection and because some capacity may be constrained down due to transmission or ramping limits. Some generators should have their capacity discounted $\left(\mathscr{C}_{gt}^{s} < P_{g}^{\max}\right)$ if they are expensive or otherwise made unavailable by dispatching constraints. This formulation introduces wind response sets that are similar in principle to those in Chapter 5 and Lyon, Zhang, and Hedman (2014): $\mathscr{C}_{gt}^{s}$ is the amount of qualified capacity from committed generator g belonging to the response set for scenario s in period t.

Figure 6.1 illustrates what one response set might look like in a hypothetical system. The map's lightly shaded regions contain capacity that is unavailable due to congestion. The bar chart identifies some specific reasons capacity has been disqualified. The ramping and transmission criteria come from physical dispatching constraints for the given scenario-period. The expense criterion reflects the cost of utilizing the capacity during a coordinated dispatch: even generators with low productions costs may be classified as expensive if

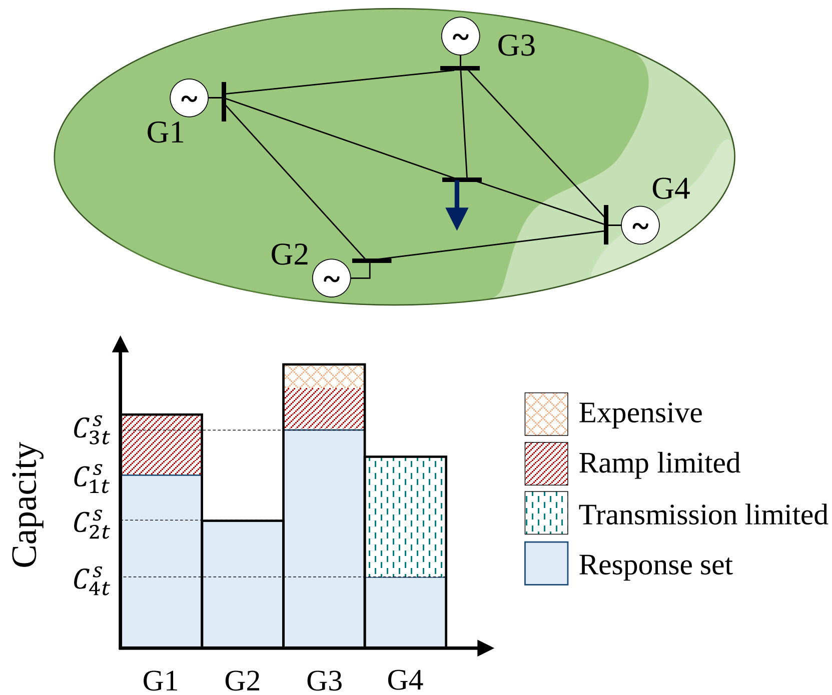


Figure 6.1: Stylized response set for a particular scenario-period (s,t).

they restrict production from other units [4]. A mathematical model for identifying and disqualifying unavailable or expensive capacity is proposed in Section 6.4.3.

Constraint (6.21) below describes a more general form of capacity requirement that accounts for uncommitted fast generators which may be turned on during scenario response:

$$\sum_{g \in G} \mathscr{C}^s_{gt} u^0_{gt} + \sum_{g \in G^{\text{fast}}} \mathscr{U}^s_{gt} \left(1 - u^0_{gt}\right) \geq Q^s_t, \qquad \forall s \in S, t \in T, \tag{6.21}$$

where the parameter $\mathscr{U}^s_{gt}$ is the capacity from uncommitted generator g that is qualified for scenario s. The proposed algorithm determines both $\mathscr{C}$ and $\mathscr{U}$ to help ensure that efficient

[4] For example, injections are generally less valuable at nodes that have small locational marginal prices (LMPs) due to congestion. Generation is less valuable where LMPs are small because the LMP reflects the marginal value of supply at the given location. Generators with high productions costs may still be valuable if the LMP exceeds their marginal cost. The proposed algorithm in Section 6.4 classifies expensive capacity based on the overall cost of dispatching the entire system and not just the marginal cost of the respective generator.

capacity is committed for scenario response. The resulting formulation assumes (6.21) will ensure the existence of a satisfactory response for each scenario, i.e., a response that satisfies all second-stage constraints in an economical manner. The proposed algorithm works by projecting information from the second stage back onto SCUC via revised response set definitions ($\mathscr{C}$ and $\mathscr{U}$). The heuristical nature of this approach arises from the entire recourse problem being reduced to a single constraint for each scenario-period.

6.4 Decomposition Algorithm

The proposed algorithm uses a two-stage structure to determine response sets. The first stage solves the deterministic SCUC with capacity requirements and the second stage identifies capacity that should not have been treated as available because it is too expensive or held back by dispatching constraints. This approach aims to reduce scenario costs without solving the more difficult (SUC) problem.

Figure 6.2 provides a flowchart representation of the proposed algorithm. The deterministic SCUC is initially solved with large response sets and the resulting solution is evaluated. If the scenario costs are too high, the response sets are pruned ($\mathscr{C}$ and $\mathscr{U}$ are made smaller) using the pruning algorithm in Section 6.4.3. The new response sets are then used to update the capacity requirement (6.21) to encourage SCUC to commit preferable sets of slow generators in future iterations.

Individual steps from the process in Figure 6.2 are now described in detail.

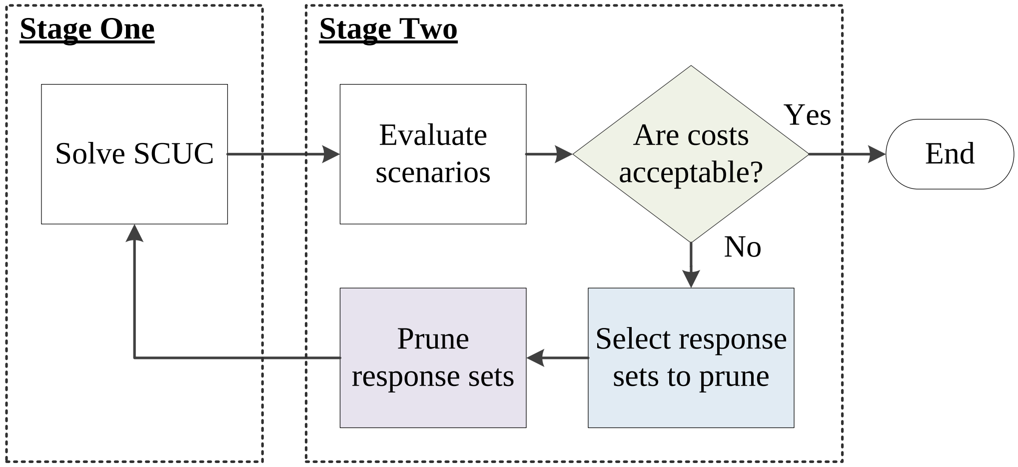


Figure 6.2: Decomposition algorithm for unit commitment.

6.4.1 Evaluating Scenarios. The evaluation subroutine assesses performance of the first-stage solution for different scenarios. This evaluation fixes the base commitments u^0 and solves the second stage of problem (SUC) (solving the second stage alone is considerably faster than the entire problem because scenarios are independent). The scenarios may be evaluated in parallel to reduce computational times because u^0 is fixed in the nonanticipivity constraints.

6.4.2 Stopping Criterion. The decomposition algorithm terminates when costs from the evaluation subroutine are deemed acceptable. The stopping criterion may be based on the stochastic or robust objectives (6.15) and (6.16) or other criteria based on operator preference. This research adopts a cost threshold for each scenario. Scenarios that do not satisfy the cost threshold are placed in set $\tilde{S}$ to be pruned (the set $\tilde{S} \subseteq S$ contains only scenarios with unsatisfactory costs in the current iteration).

The cost threshold is defined here relative to regret, i.e., the scenario cost above what could be obtained if a perfect forecast were available. The tolerable percent regret $\bar{R}$ is relative to a lower bound $\underline{C}$ on the perfect-forecast solution:

$$c^s(x) \leq (1+\bar{R})\underline{C}^s, \qquad \forall s \in S. \tag{6.22}$$

Constraining regret is a heuristic for reducing expected regret, which is equivalent to reducing expected cost. This approach is generally more risk averse than stochastic programming and more risk neutral than robust optimization (Jiang, Wang, Zhang, & Guan, 2013). Regret is estimated here relative to a lower bound on the perfect dispatch solution as this bound can be calculated efficiently.

Applying such cost constraints is not guaranteed to minimize expected costs. However, there is also a desire to protect against extreme scenarios. Operators already apply reserve requirements for wind uncertainty that have been described as conservative (Holttinen et al., 2012). Risk discussion in power systems frequently centers on reliability, but there is also risk of expensive solutions. Methods from the literature that are risk-averse include robust optimization, conditional value at risk (Huang, Zheng, & Wang, 2014), and minimax regret (Jiang et al., 2013). The ISOs are chartered to maximize social welfare but they are not required to be risk neutral (as long as reliability standards are met). Risk-averse approaches like minimizing regret have long been recognized as a rational (Savage, 1951), and the allowance for risk aversion within the mathematical model may help to further clarify organizational goals and improve transparency for market participants.

6.4.3 Pruning Response Sets. At this point, the decomposition algorithm has identified expensive scenarios and placed them in set $\tilde{S}$ to be pruned. The qualified capacity from committed generators and uncommitted fast generators is represented as

$$\hat{P}^s_{gt} = \begin{cases} \mathscr{C}^s_{gt}, & \text{if } u^0_{gt} = 1 \\ \mathscr{U}^s_{gt}, & \text{if } u^0_{gt} = 0 \text{ and } g \in G^{\text{fast}} \\ 0, & \text{if } u^0_{gt} = 0 \text{ and } g \notin G^{\text{fast}}. \end{cases} \tag{6.23}$$

The pruning subroutine now identifies the portion of this capacity $\hat{P}^s_{gt}$ that cannot be dispatched at low cost given fixed base commitments u^0. The mathematical model for identification of costly capacity (ICC) is defined below:

$$(ICC): \text{ minimize} \sum_{g\in G}\sum_{t\in T}\sum_{s\in\tilde{S}} \hat{P}^s_{gt}\gamma^s_{gt} \tag{6.24}$$

subject to:

$$c^s(x) \leq (1+\bar{R})\underline{C}^s, \qquad \forall s\in\tilde{S}. \tag{6.25}$$

$$i^s_{nt} = \sum_{g\in G(n)} p^s_{gt} + \sum_{g\in G} m^s_{gnt} + k^s_{nt} + b^s_{nt} - D^s_{nt}$$

$$\forall n\in N, t\in T, s\in\tilde{S} \tag{6.26}$$

$$\sum_{n\in N} m^s_{gnt} \leq \hat{P}^s_{gt} - p^s_{gt}, \qquad \forall g\in G, t\in T, s\in\tilde{S} \tag{6.27}$$

$$\sum_{n\in N} m^s_{gnt} \leq \hat{P}^s_{gt}\gamma^s_{gt}, \qquad \forall g\in G, t\in T, s\in\tilde{S} \tag{6.28}$$

$$\sum_{g\in G} m^s_{gnt} \leq \sum_{g\in G(n)} \mathscr{C}^s_{gt}\left(1-u^0_{gt}\right), \qquad \forall n\in N, t\in T, s\in\tilde{S} \tag{6.29}$$

$$m^s_{gnt} \geq 0, \qquad \forall g\in G, n\in N, t\in T, s\in\tilde{S} \tag{6.30}$$

Wind availability (6.2)

Network constraints (6.3), (6.5)

Generator capacity and ramping constraints (6.6), (6.7)

Minimum up and down times (6.8)–(6.10)

Nonanticipivity constraints (6.11)

Reserve requirements (e.g., contingency requirements)

Variable domains (6.12).

Problem (ICC) coordinates dispatch across the entire horizon (e.g., 24 hours). It resembles the second stage of problem (SUC) with three differences. First, the scenario costs are constrained to be below the maximum regret (6.25). Second, generator capacity is limited to what currently belongs in the response set (6.27). Third, costly capacity may be replaced by dummy injections if the dispatching constraints cannot be satisfied. The dummy injection m^s_{gnt} is made at node n to replace capacity from generator g. The idea is that this injection represents capacity from slow generators that could be committed in future iterations when a stricter capacity requirement will be used. As such, (6.29) limits dummy injections to locations that have uncommitted capacity from slow generators.

The proportion of capacity replaced by dummy injections is tracked by γ through (6.28). The objective function minimizes this value so that dummy injections are only used to replace capacity that cannot be dispatched without violating a constraint, including the constraint on production costs (6.25). Let γ^* be the optimal solution of problem (ICC). The response sets are pruned based on γ^* according to

$$\mathscr{C}^{s(\text{new})}_{gt} = \begin{cases} \mathscr{C}^s_{gt}\left(1-\gamma^{s^*}_{gt}\right), & \text{if } u^0_{gt}=1 \\ \mathscr{C}^s_{gt}, & \text{if } u^0_{gt}=0, \end{cases} \tag{6.31}$$

$$\mathscr{U}^{s(\text{new})}_{gt} = \begin{cases} \mathscr{U}^s_{gt}\left(1-\gamma^{s^*}_{gt}\right), & \text{if } u^0_{gt}=0 \\ \min\left\{\mathscr{U}^s_{gt}, \mathscr{C}^{s(\text{new})}_{gt}\right\}, & \text{if } u^0_{gt}=1. \end{cases} \tag{6.32}$$

The capacity of each response set member is discounted in (6.31) and (6.32) by the proportion replaced by dummy injections in problem (ICC). A slightly modified rule is used in (6.32) for uncommitted fast generators to enforce $\mathscr{U} \leq \mathscr{C}$. This relation may help to speed up convergence and is justified because fast generators that are too expensive when committed are unlikely to be used when they have not been committed due to the fixed start-up costs.

Problem (ICC) can be easily parallelized because nothing links scenarios. The model for each scenario is comparable to the evaluation sub-routine from Section 6.4.1 except for the inclusion of dummy injections m. Most of these additional variables must equal zero because (6.29) forbids dummy injections where qualified capacity does not exist (the number of nodes greatly exceed the number of generators in power systems). Therefore, problem (ICC) has similar computational requirements as the evaluation sub-routine. Both problems are relatively easy to solve quickly due to restrictions on binary commitments in the second-stage.

6.5 Wind Scenario Generation

To evaluate the proposed response sets, it is necessary to generate wind scenarios that reflect the day-ahead uncertainty faced by SCUC. Wind forecasting and simulation methods typically begin by forecasting wind speed and then estimating the corresponding electric power generated from that wind. This approach is common because advanced physical weather models are available, making wind speeds *relatively* easy to forecast. The speed is converted to power via a transformation like that in Figure 6.3, which is based on the physical characteristics of the generators. The relationship between wind speed and power is

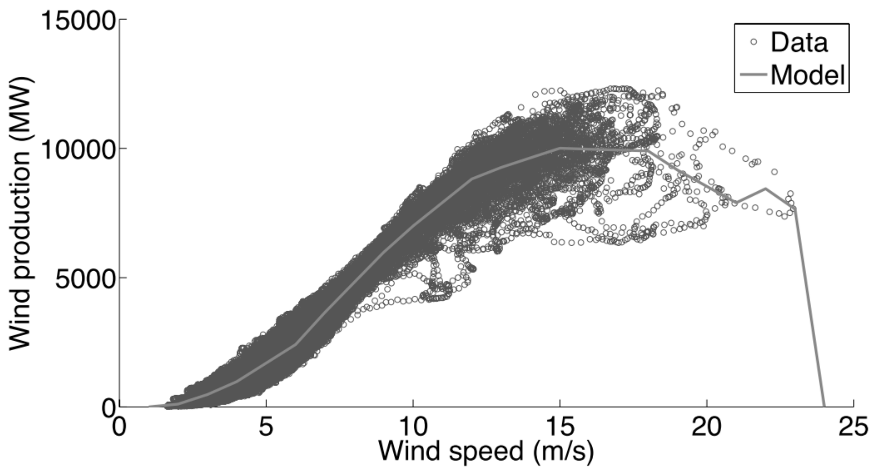


Figure 6.3: Power curve for a wind farm (Papavasiliou et al., 2011).

nearly linear but tapers off near maximum capacity. Some precision is lost because a single wind speed is used to cover multiple nearby turbines (Wan, Ela, & Orwig, 2010). Locational approximations notwithstanding, most studies concerned with wind power generation start by characterizing a single wind speed for the localized area.

The most common wind speed distributions in the literature are Weibull and Releigh (which is a special case of the Weibull distribution). A review of the literature by Dorvlo (2002) highlights that both distributions have been used extensively and the best choice depends on the dataset. However, operations are also influenced by temporal and spatial correlations. The correlation between periods is generally strong, so strong that the persistence model (i.e., a moving average with unit window length) can be reasonable for forecasting a few hours ahead (Landberg, 2001). A visual indication persistence model accuracy is compared to a state-of-the-art forecasting method is shown in Figure 6.4. The second type of correlation is spatial. Adding wind across the system tends to reduce the overall volatility (due to the law of large numbers), but this effect is subdued in regions where wind speeds are

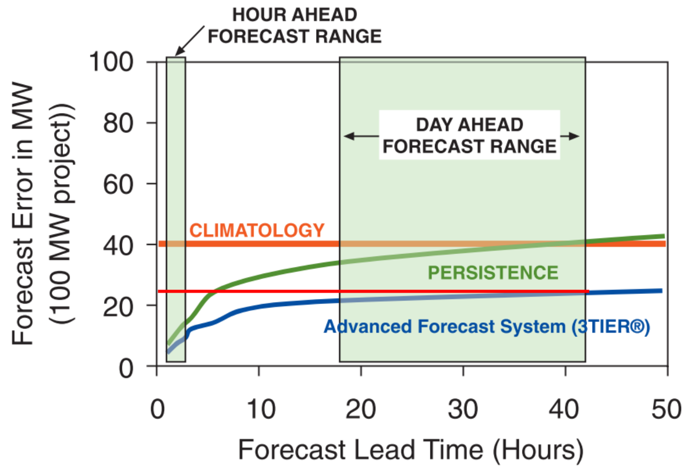


Figure 6.4: Typical forecast errors using climatology (moving average with a very wide window), persistence, and an advanced forecasting model (Lerner, Grundmeyer, & Garvert, 2009).

correlated. The spatial wind variation is also relevant because different locations uniquely influence power flows and congestion. Models that ignore spatial temporal correlations may overlook important phenomena with large impacts on system operations.

The wind forecasting literature is reviewed by Lei, Shiyan, Chuanwen, Hongling, and Yan (2009). Forecasting models can be categorized into three categories: physical, statistical, and artificial intelligence. Physical models simulate the Earth's atmosphere while taking into account factors like terrain, temperature, pressure, and humidity. They are generally appropriate for long-term wind forecasting but are also used to generate inputs for short-term models. Artificial intelligence is extensively used for short-term forecasting (Lei et al., 2009). Popular methods include artificial neural networks, fuzzy logic, and support vector machines. Statistical models frequently involve autoregressive moving average (ARMA) techniques, which are well suited for characterizing temporal correlation (Lei et al., 2009). Statistical models are also convenient for generating random scenarios because they often include a random component that lends itself well to Monte Carlo simulation.

Statistical models are widely used for simulating wind speed. ARMA models incorporate autocorrelation by relating each data point to information from previous periods (Montgomery, Jennings, & Kulahci, 2008). The structure of an ARMA model is

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{6.33}$$

where y_t is the value of the time series in period t, ϕ_j and θ_j are autoregressive (AR) and moving average (MA) parameters for time lag j, and ε is a random error term normally distributed with $\mathrm{E}[\varepsilon] = 0$. The error terms can be interpreted as random shocks that determine how the time series progresses between periods (Montgomery et al., 2008). In ARMA models, the data at every time point is a function of the previous data and the random shocks to the system. The model includes parameters looking back p periods for the AR terms and q periods for the MA terms.

ARMA models can be modified to account for spatial correlation (cross-correlation). For example, multivariate ARMA (VARMA) models incorporate additional variables that link information from between locations. These additional variables tend to make VARMA models difficult to fit for large problem instances (Montgomery et al., 2008). However, such models have been used for applications with fewer than five cross-correlated wind series (S´ooder, 2004; Lojowska, 2009; Ewing, Kruse, Schroeder, & Smith, 2007). More tractable approaches for larger problems are proposed by Xie and Billinton (2009), Morales, Mínguez, and Conejo (2010). These authors reduce the number of parameters by capturing the cross-correlation through the error terms ε. Their strategy is to generate random errors (shocks) that are normally distributed but also correlated across locations. Both references fit independent models for each wind location, characterize the error correlations between locations, and then use this information to generate random shocks that are cross-correlated. The benefit of these approaches is that they are practical for large-scale problem instances.

This research employs an ARMA model to capture temporal correlations and the method of Morales et al. (2010) to capture cross correlation between locations. The model for each location includes one AR term and one MA term, which simplifies to

$$y_t = \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t. \tag{6.34}$$

There is also strong seasonality for wind that should be accounted for. For example, many areas experience high wind in the mornings and low wind in the evenings. These daily patterns are captured here by introducing seasonal terms to the model. The updated model includes a 24-hour seasonal differencing term and a new moving average term to get

$$y_t = \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t + (y_{t-24} - \phi y_{t-25}) + (\Theta \varepsilon_{t-24} + \theta \Theta \varepsilon_{t-25}), \tag{6.35}$$

where Θ is a new parameter that captures the seasonal moving average. The reader may refer to Montgomery et al. (2008) for more details on these types of models. Using the nomenclature from that reference, equation (6.35) is classified as an (1,0,1)x(0,1,1)24 model. This particular formulation is chosen based on the goodness of fit and the distribution of residuals when applied to the data used in this study.

Torres, García, Blas, and Francisco (2005) show that hourly forecasts can be improved by transforming the wind speed data from Weibull to standard normal. This type

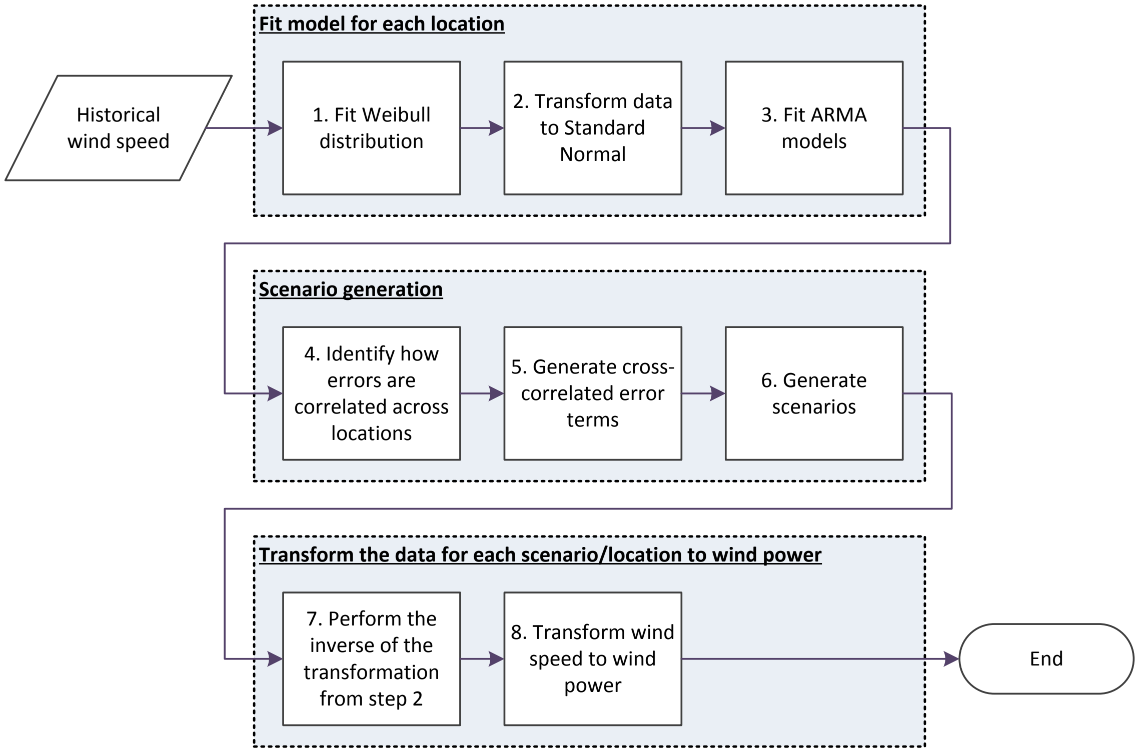


Figure 6.5: Flowchart for statistically generating wind scenarios. Steps 3 and 6 use the ARMA model in (6.35). Steps 4 and 5 capture spatial correlation using the algorithm of Morales et al. (2010).

of transformation is also used by Morales et al. (2010) and Papavasiliou et al. (2011). One benefit of working with normally distributed data is that wind predictions are never negative after performing the reverse transformation back to Weibull (Weibull random variables are nonnegative with probability one). The same strategy is employed here by transforming the historical wind speed data to approximate the normal distribution. The time series are fit to the transformed data before generating scenarios. The overall process is described in Figure 6.5. In the end, the data is transformed back to Weibull (to get wind speed) and then to power using a power curve like that in Figure 6.3.

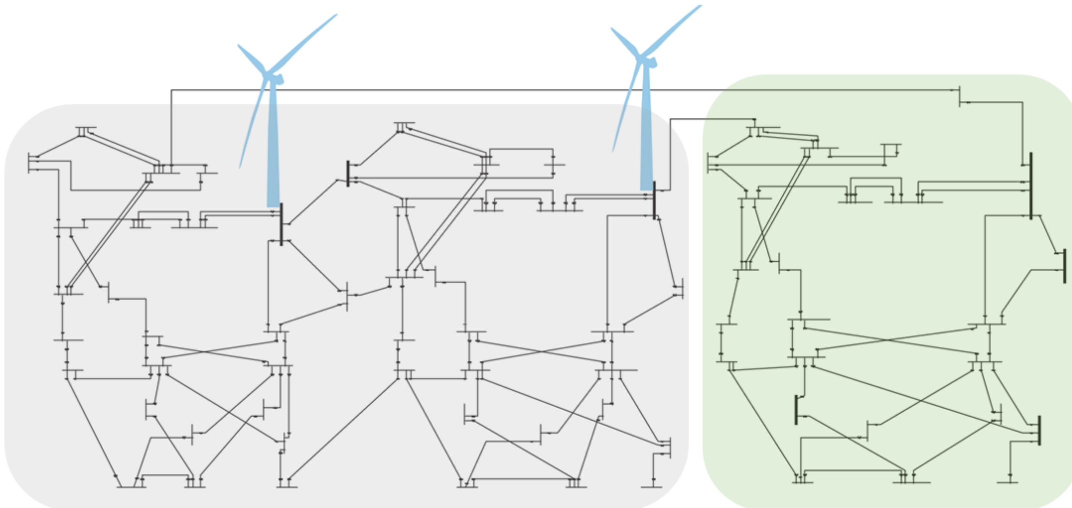


Figure 6.6: Wind locations (nodes 23 and 47) and contingency reserve zones.

6.6 Analysis and Results: IEEE 73-Bus (RTS 96) Test Case

The proposed algorithm is evaluated on the IEEE 73-bus (RTS 96) test case (Grigg et al., 1999). Modifications are made to the test case as described in Section 4.7. Wind farms are also added to the system as described in Section 6.6.1. The wind availability amounts to 33% of the total load on average.

This analysis uses 10-minute spinning reserves to cover generator contingencies. Contingency reserves are enforced using the two-reserve-zone requirement from Chapter 5 with a conservative reserve sharing parameter $\alpha = 0.5$. The contingency reserve zones are shown in Figure 6.6 along with the wind locations. Reliability against contingencies or any other non-wind uncertainties is not evaluated explicitly, but 10-minute reserve are forced to exceed the largest contingency by the respective reserve requirements. Load imbalances are allowed at a cost of $\nu =$\$10,000/MWh.

Testing is performed using CPLEX v12.5 on an 8-core 3.6GHz computer with 48GB of memory. SCUC is terminated after five minutes or upon reaching an optimality gap of 0.1% unless otherwise stated.

6.6.1 Design of the Analysis. Various policies are tested here to assess the proposed decomposition algorithm. Stochastic programming is used first to demonstrate the difficulty of solving problem (SUC) directly. This is contrasted by deterministic policies from the literature that allow for quick solution times and reasonable results when tuned properly. Finally, the capacity response sets are shown to improve expected costs in less than 20 iterations on average. The tested approaches are described in more detail below:

- Stochastic programming: This approach directly solves problem (SUC) using CPLEX. Scenario selection is applied to limit the problem size so it is computationally tractable. The scenario selection method of Gröwe-Kuska, Heitsch, and Römisch (2003) combines similar scenarios and recalculates the probabilities π accordingly. Using this method, a smaller set of 10 scenarios are identified out of 1000 candidates.

- Baseline: The baseline policy uses deterministic reserve requirements from the literature. The approach involves no scenario information but requires reserve to exceed a prescribed percentage of the wind forecast. Several variations are studied to gauge how well the baseline requirements can perform when properly tuned. Details on the baseline are provided in Section 6.6.1.

- Default response sets: The default response sets apply the capacity constraint (6.21) for 50 random scenarios. The response sets are initialized with all capacity qualified, i.e., all $\mathscr{C}^s_{gt} = \mathscr{U}^s_{gt} = P^{\max}_g$, and pruned using the decomposition algorithm. The tolerable regret is set to $\bar{R} = 10\%$ above a lower bound on the perfect forecast solution.

- Buffered response sets: This is an altered formulation using the same 50 training scenarios as the default response sets. Although the default approach works well for the training scenarios, it may allow additional load imbalance when evaluated out-of-sample. More conservative requirements may be warranted when the penalty ν is large to avoid load imbalances. The buffered response sets provide that conservatism by requiring an additional 5% capacity margin for each response set (by adjusting the right-hand side of (6.21)).

- Buffered response sets with scenario selection (SS): This policy uses scenario selection strategies from Papavasiliou et al. (2011) to identify training scenarios that have extreme wind shortages or deviations between hours. Modeling these scenarios helps to commit additional capacity that can protect against out-of-sample load imbalances.

The IEEE 73-bus test case does not identify fast generators, so some assumptions are needed for the analysis. This chapter first evaluates the system with no fast generators ($G^{\text{fast}} = \varnothing$) and then proceeds with 13% of capacity classified as fast. The results show that the baseline policies can perform much better when commitments are allowed to change in the evaluation stage. This result corroborates observations from the literature, e.g., Tuohy et al. (2009) find that fast generators can reduce the cost savings of stochastic programming from 0.6% to 0.25%. As it happens, response sets are found in this chapter to reduce expected costs by around 0.5% only when there are no fast generators. Response sets and the baseline policy provide similar results when there are fast generators.

Baseline Policies. Several heuristics have been proposed in the literature to deal with wind uncertainty (Holttinen et al., 2012). Many focus on mitigating short-term fluctuations on the order of 5 to 20 minutes. Some require reserve to exceed a certain percentage of the load or wind forecast at the hourly level (Sioshansi & Short, 2009; Papavasiliou & Oren, 2013). A similar baseline policy is used here:

$$\sum_{g\in G} r^{\text{slow}}_{gt} \geq W \sum_{n\in N} K^0_{nt}, \qquad \forall t \in T \tag{6.36}$$

$$p^0_{gt} - p^0_{g,t-1} + r^{\text{slow}}_{g,t-1} \leq R_g, \qquad \forall g \in G, t \in T \tag{6.37}$$

$$p^0_{gt} + r^0_{gt} + r^{\text{slow}}_{gt} \leq P^{\max}_g, \qquad \forall g \in G, t \in T. \tag{6.38}$$

The variable r^{slow}_{gt} represents *slow reserve* from generator g that is available within one scheduling period (60 minutes) and r^0_{gt} represents contingency reserves. Constraint (6.36) requires slow reserve to exceed a prescribed percentage of the wind forecast. Several values of W are tested in this section and then analyzed post-hoc.

For reserve products available in 60 minutes, it is important to consider hourly ramping limitations. Constraint (6.37) accounts for each generator's scheduled ramping (Sioshansi & Short, 2009). For example, this constraint forces slow reserve to zero when a generator is already scheduled to increase production at its maximum ramp rate. Constraint

(6.38) separates r and r^{slow} so that the slow reserve requirement is met without sacrificing contingency reserves.

Similar reserve requirements are used by Sioshansi and Short (2009) and Papavasiliou and Oren (2013). The major difference is that those authors also incorporate a percentage of the load forecast in the reserve requirement (6.36). The load component is omitted here because policies based only on wind are found to be dominant for this particular test case (which is unsurprising because this analysis assumes there is no load uncertainty). The baseline requirements are not enforced on a zonal level because the wind farms in this test case are centrality located in parts of the system that are generally export-constrained, making it easier to import reserves during wind shortages. The prime wind locations notwithstanding, local transmission constraints can still hinder reserve deliverability. A strong point of the proposed methodology is that it adapts to different congestion patterns across scenarios without relying on a single zone definition to work well under diverse conditions.

Wind Modeling. Wind modeling follows that of Lyon, F. Wang, et al. (2014). Wind farms are placed at nodes 23 and 47, which are central locations containing 660 MW of coal generation each. The wind locations are plotted in Figure 6.6. The wind is placed at central parts of the system because isolated areas are more likely to require additional transmission investments that are beyond the scope of this dissertation. The developed wind scenarios generate power up to 3340 MW and energy equal to 33% of the average load. Wind energy is assumed to have a production bid of zero and can be accepted up to the available amount.

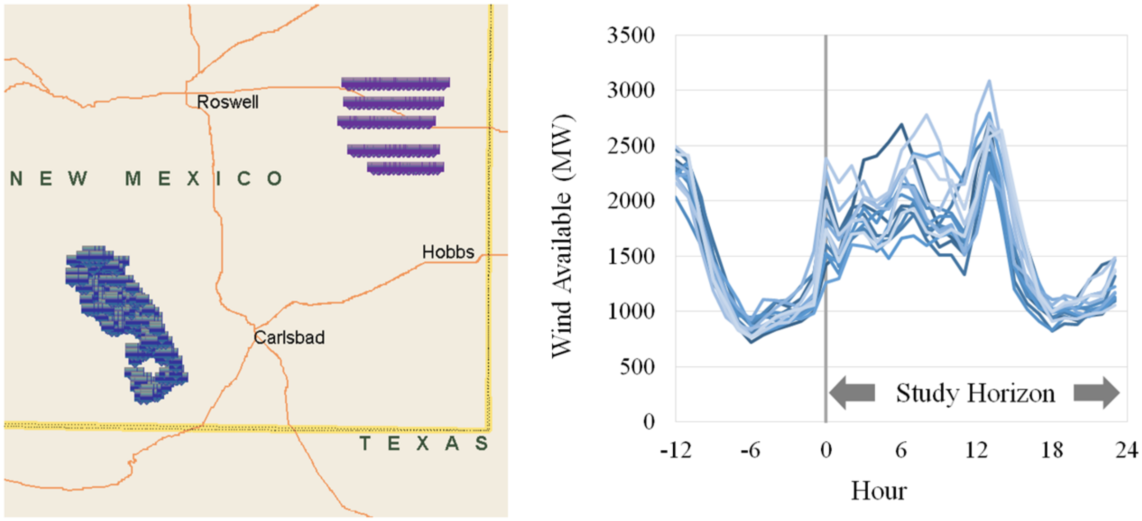


Figure 6.7: (a) Wind data locations. (b) Sample of 15 wind scenarios.

Scenarios are generated based on data from NREL's Western Wind dataset for the first three weeks of August 2005 (National Renewable Energy Laboratory, n.d.). Figure 6.7 (a) maps 510 turbines distributed across two clusters; these clusters are aggregated to create the two separate wind locations described above. Hourly ARMA models are fit for each location based on wind speed and the time series are later converted to power using an estimate of the aggregate power curves.

The adopted model (6.35) includes one AR and MA term for the most recent hour and a 24-hour MA term with seasonal differencing for daily seasonality per equation (6.35). The seasonality captures how wind varies by time of day: ramping up in the early morning and dropping off in the afternoon. The model's ability to capture temporal correlation is validated using the Ljung-Box test with significance level 0.10, which fails to reject the null hypothesis of independent errors for the first twenty time lags. The model does a reasonable job capturing the autocorrelation across most hours of the day.

Since scenarios are generated only from historical data (and not sophisticated forecasting models that consider weather modeling), the variance between samples tends to exceed the forecast errors seen in practice (Hodge et al., 2012). This bias is corrected by normalizing the sampled wind speeds with the average so that the power stays within 20% of the mean approximately 80% of the time. The adjustment is applied between steps seven and eight in Figure 6.5. Without this adjustment, the variability between scenarios is much greater than would be seen in practice. Figure 6.7 shows the available wind power for the first 15 scenarios generated using this approach. The forecast adopted in SCUC follows the mean value from 1000 simulated time series.

Test Days. Wind has the potential to influence the system in different ways across seasons. This analysis examines a portfolio of days across the year including a weekday (WD) and weekend (WE) from each season. The load profile for each day is shown in Figure 6.8. The same wind scenarios are used for each case, ranging in forecast from 28% of energy in peak days to 43% on off-peak days. The overall wind penetration is 33% of energy across the year. This exceeds the current renewable portfolio standards in the U.S. [5] but is below levels targeted by some European countries, e.g., Germany targets 45% renewable by 2025 and 55% renewables by 2035 (Lang & Mutschler, n.d.). Wind may exceed or fall short of 33% on individual days, which is reflected in the range of days tested.

[5] California is targeting 33% by the year 2020, but some of that will be met by other fuel types besides wind.

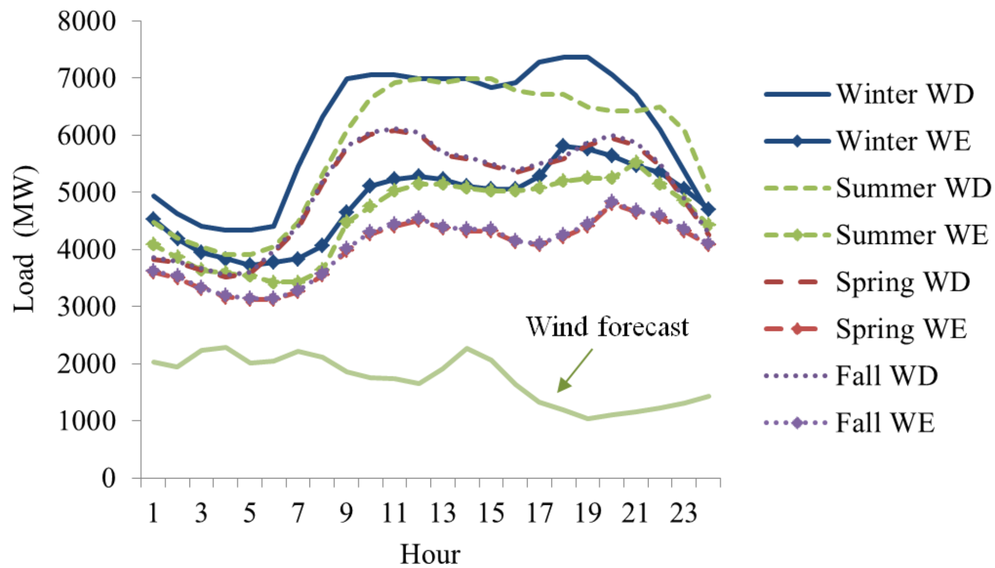


Figure 6.8: Load profile for the eight test days.

Evaluating Results. Given the first-stage commitments, solutions are evaluated by simulating dispatch across 100 random scenarios that are independent of the training scenarios. This analysis assumes the scenario is revealed upfront and solves a single 24-hour model. The corresponding results will be optimistic because, in reality, forecast errors still occur during the operating day. This assumption affects all of the evaluated models (they are compared in a consistent way) and, despite being an approximation, is common in the literature (Ruiz et al., 2009; Papavasiliou et al., 2011; Morales et al., 2009; Papavasiliou & Oren, 2013; Bertsimas et al., 2013; Jiang et al., 2014).

6.6.2 Analysis for No Fast Generators. The set of recourse actions are hereby restricted such that generator commitments cannot change in the second stage. The day-ahead problem attempts to find commitments that work well across a range of scenarios: allowing for low-cost generation while avoiding load imbalance.

Table 6.1: Stochastic programming results. The evaluation results are averaged across scenarios.

Day	(SUC) Solution: Time (min)	(SUC) Solution: Optimality Gap (%)	(SUC) Solution: Objective (millions \$)	Out-of-Sample Evaluation: Cost (millions \$)	Out-of-Sample Evaluation: Load Imbalance (MWh)	Out-of-Sample Evaluation: Regret (%)
Winter WD	120	0.45	1.14	1.15	1.89	3.61
Winter WE	120	0.25	0.54	0.54	0.38	2.44
Summer WD	120	0.19	0.93	0.96	2.16	4.20
Summer WE	120	0.16	0.48	0.47	0.00	2.77
Spring WD	120	0.20	0.63	0.69	6.04	12.11
Spring WE	49	0.10	0.33	0.33	0.04	4.79
Fall WD	103	0.10	0.64	0.67	3.37	7.22
Fall WE	120	0.14	0.34	0.34	0.40	5.54
Average	109	0.23	0.63	0.64	1.78	5.25

The stochastic programming results are summarized in Table 6.1. The solver terminates after two hours or upon reaching an optimality gap of 0.10%, which is met on 2/8 days. The solver averages an optimality gap of 0.23% despite only modeling 10 scenarios. The small sample size leads to many out-of-sample load imbalances because the full range of uncertainty is not captured. More scenarios would be preferable because the average regret exceeds 5%, but they would only exasperate the already long solution times. This points to the importance of scenario selection and decomposition algorithms because problem (SUC) is hard to solve directly.

The results of the baseline model are shown in Table 6.2 and Figure 6.9 for several values of W. A clear trade-off is shown where more reserve protects against load imbalance but comes at a cost. The baseline policy that is observed to minimize the expected cost is $W = 35\%$ of the wind forecast. This policy is the cheapest overall but can be far from the

Table 6.2: Results averaged across scenarios for baseline policies with $W = \{20, 25, 30, 35, 40\}\,\%$ of the wind forecast. The regret assumes $v = \$10{,}000$/MWh. The best results are highlighted for each day.

	Load Imbalance (MWh)					Regret (%)				
Day	20%	25%	30%	35%	40%	20%	25%	30%	35%	40%
Winter WD	4.6	0.2	0.0	0.0	0.0	5.3	1.8	1.8	2.2	2.6
Winter WE	16.3	7.5	0.0	0.0	0.0	32.4	15.9	2.0	2.2	2.4
Summer WD	2.2	0.8	0.0	0.0	0.0	4.1	3.0	2.2	3.0	3.3
Summer WE	40.0	0.1	0.1	0.1	0.0	89.4	2.7	2.7	2.9	3.2
Spring WD	0.8	0.1	0.0	0.0	0.0	3.2	2.4	2.4	2.8	3.5
Spring WE	35.9	14.7	6.6	0.0	0.0	116.4	49.7	25.0	4.4	4.9
Fall WD	2.2	0.0	0.0	0.0	0.0	5.1	1.9	2.4	2.8	3.3
Fall WE	39.3	20.2	7.4	0.0	0.0	125.4	65.5	26.8	4.4	4.5
Average	17.7	5.4	1.8	0.0	0.0	30.6	11.0	5.2	2.9	3.2

cheapest on individual days, suggesting that there is room for improvement by adapting a dynamic policy.

The same policy of $W = 35\%$ minimizes the expected cost for several load imbalance penalties v as shown in Figure 6.10. The more optimistic baseline policies only begin to perform as well when $v <$ \$2500/MWh. Such small values may be characteristic of systems with many flexible loads that are willing to reduce their consumption at relatively low price.

Response Sets. The proposed decomposition algorithm is evaluated across the same set of days and scenarios as the baseline. A set of 50 scenarios are included in the model. These in-sample scenarios are generated randomly and are independent from the scenarios used for evaluation.

The response set constraints are applied on top of slow reserve constraint (6.36) with $W = 20\%$. Table 6.3 and Figure 6.11 summarize the results. Every day experiences

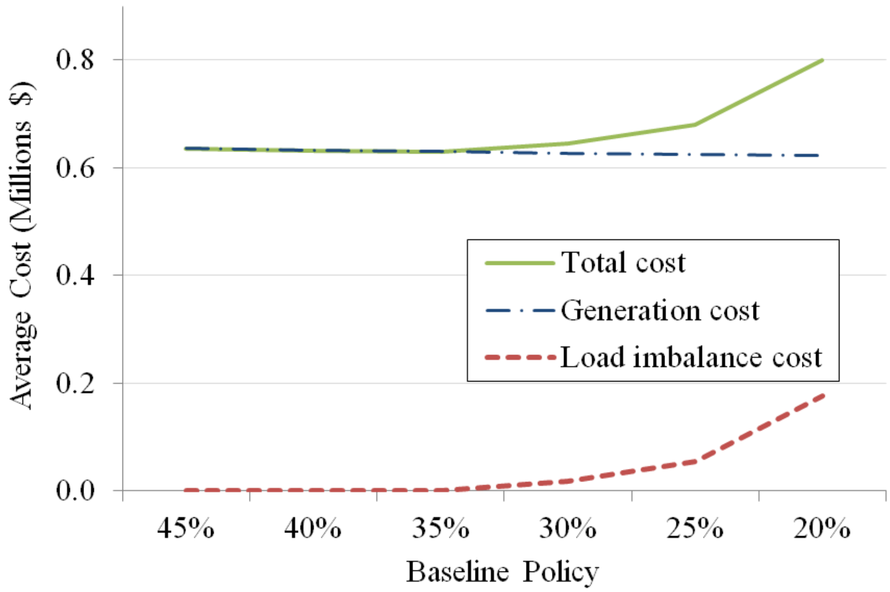


Figure 6.9: Generating production costs vs. load imbalance costs for different baseline policies. The respective policies require slow reserve to exceed the specified percentage of the wind forecast.

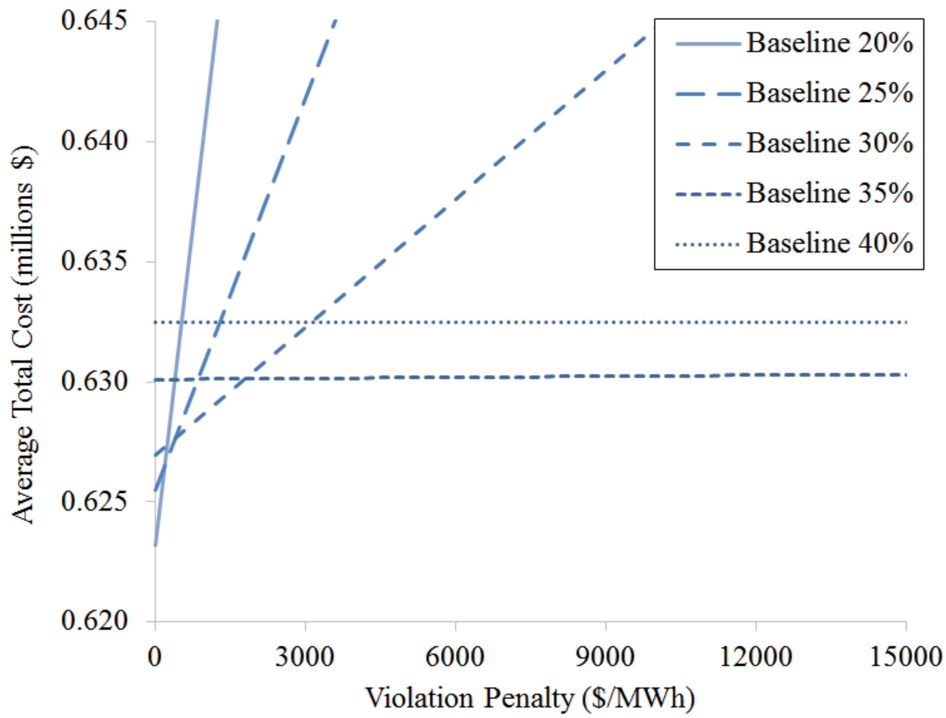


Figure 6.10: Total costs for the baseline policies assuming different violation penalties v.

load imbalance in the first iteration before the response sets are pruned. All imbalances in the training scenarios are eliminated through course of the decomposition algorithm. However, the performance is more modest when evaluated out-of-sample. The out-of-sample performance is heavily influenced by load imbalances in scenarios that are absent from the scheduling model. Much of these violations can be avoided by modifying the formulation as follows.

The response set requirements are made more strict by adding a 5% margin to the right-hand side of the capacity requirement (6.21) (this is the *buffered* policy in Table 6.3). This adjustment rarely hurts performance but can protect against shortages in scenarios that are absent from the model. The expected costs for this strategy are extrapolated for different imbalance penalties ν in Figure 6.11. The proposed model outperforms all baseline policies for \$2000/MWh $\leq \nu \leq$ \$6000/MWh. The performance is inferior for larger penalties ν because some out-of-sample load imbalances still exist as shown in Table 6.3.

Response Set Scenario Selection. One of the challenges of the proposed algorithm is selecting a proper set of scenarios to include in the model. The results in Table 6.3 show there are load imbalances during evaluation. The imbalances occur despite no imbalances in the training scenarios. Scenario selection techniques may help avoid such out-of-sample drops in performance.

Various scenario selection techniques are proposed in the literature. Some identify scenarios that are representative of a larger set, e.g., Gröwe-Kuska et al. (2003), and others

Table 6.3: Results averaged across scenarios for response set policies. The regret assumes v = \$10,000/MWh. The best results are highlighted for each day. SS = scenario selection.

Day	Load Imbalance (MWh)			Regret (%)		
	Default	Buffered	Buffered with SS	Default	Buffered	Buffered with SS
Winter WD	0.08	0.08	0.09	1.40	1.47	1.87
Winter WE	0.00	0.00	0.00	1.51	1.51	1.96
Summer WD	1.32	1.32	0.00	3.20	3.20	2.18
Summer WE	0.11	0.11	0.00	2.41	2.41	2.51
Spring WD	5.11	1.37	0.00	10.10	4.11	2.32
Spring WE	1.19	1.19	0.25	6.32	6.33	4.59
Fall WD	1.76	1.76	0.00	4.37	4.37	2.21
Fall WE	0.00	0.71	0.55	2.94	5.09	5.46
Average	1.20	0.82	0.11	3.73	3.14	2.51

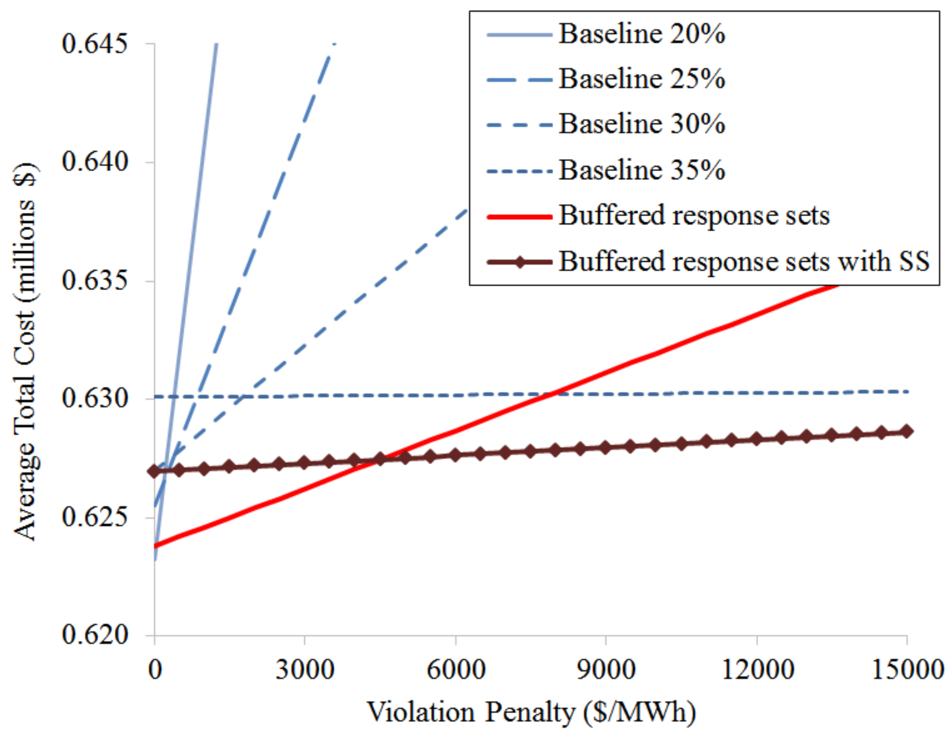


Figure 6.11: Total costs for response set policies with different violation penalties v. SS = scenario selection.

focus on capturing extreme scenarios, e.g., Papavasiliou et al. (2011). The latter approach is used here to hedge costly load imbalances. A small sample of scenarios is chosen to include:

1. the scenario with the smallest available wind energy overall,
2. the scenario with the smallest available wind energy in any one period,
3. the scenario with the biggest variance between periods,
4. the scenario with the biggest morning ramp,
5. the scenario with the biggest evening ramp,
6. the scenario with the biggest sum of absolute differences between periods,
7. the scenario with the biggest change in one period,
8. the scenario with the biggest difference between the maximum and minimum wind availability.

From 1000 randomly generated scenarios, seven extreme scenarios are found to satisfy one or more of the above criteria. These seven scenarios are fed into the decomposition algorithm to generate the scenario selection (SS) results in Table 6.3. The results show that load balance tends to improve due to the conservatism introduced by these extreme scenarios. The regret also improves because production costs increase relatively little compared to the smaller load imbalances. These results show that the way scenarios are selected can be more important than the total number of training scenarios included in the model [6] . It is advisable for more extreme scenarios to be included when the load imbalance penalty ν is high to further hedge against load imbalances.

The performance is extrapolated for different load imbalance penalties ν in Figure 6.11. The proposed model has the potential to improve efficiency if coupled with effective scenario selection. The cost improvement over the baseline policies ranges from 0.5% to 0.2% for \$5000/MWh $\leq \nu \leq$ \$15,000/MWh. The crossover point where the best baseline policy has a lower cost is $\nu =$ \$35,000/MWh, which is higher than most value of lost load estimates (Sullivan et al., 2009). Systems with higher load imbalance penalties should specify an uncertainty set with more extreme scenarios to further hedge against load imbalances.

A statistical analysis is applied to compare response sets with scenario selection to the baseline policy with $W = 35\%$. The average cost difference is approximately normally distributed and is analyzed here using the t-distribution (Montgomery & Runger, 2010) due to the statistically small sample size of eight days. The 95% confidence interval for the daily cost improvement is \$(640, 4671) for $\nu =$ \$5000/MWh. The results indicate that the proposed policy will provide some cost benefit over the long run compared to the best baseline policy.

Computation. The computational bottleneck in the proposed algorithm is the need to solve SCUC once every iteration. The evaluation and pruning stages solve relatively fast and

[6]The stochastic programming model can also be replicated using these scenarios. Doing so yields an average regret of 3.04%, which is lower than the scenario selection technique of Gröwe-Kuska et al. (2003) in Table 6.1 but higher than the best baseline and response set policies.

may analyze scenarios in parallel. The wall clock computation times are summarized in Table 6.4. A theoretical time for parallel computing is also computed by replacing the total evaluation and pruning times with the slowest scenario from each stage. The potential time reduction from parallel computing is modest (around 10%) because most of the time is spent solving SCUC. However, parallel computing would allow the model to consider many more scenarios without greatly increasing computation times.

The number of iterations are also shown in Table 6.4. One iteration is roughly comparable to the total solution time when using a baseline policy. The proposed algorithm converges in as little as two iterations and as many as twenty-eight. This is slower than the baseline models but much faster than solving problem (SUC) directly. It is also faster than results reported for the fast stochastic programming method of Papavasiliou and Oren (2013). Their method also solves SCUC at each iteration (it actually solves SCUC several times in parallel), so a reasonable comparison can be made by looking at the iteration counts. They stop their algorithm after 120 iterations, which is four times longer than the response set decomposition algorithm. Some of the additional iterations can be attributed to Papavasiliou and Oren (2013) using a larger test case [7] , but the difference in iteration count is still arguably large.

Twenty-eight iterations is too many for application in most day-ahead markets. However, there may be sufficient time to apply the algorithm as an out-of-market correction (OMC) tool to repair unreliable solutions. This would be a natural first application while

[7] Papavasiliou and Oren (2013) test on the WECC 225-bus test case, which has roughly 30% more generators and three times more nodes and lines than this study.

Table 6.4: Iterations and computing time of the decomposition algorithm (theoretical times for parallel computing are provided in parenthesis).

	Iteration Count			Time (min)		
Day	Default	Buffered	Buffered with SS	Default	Buffered	Buffered with SS
Winter WD	25	10	10	94 (81)	50 (46)	47 (45)
Winter WE	5	3	6	18 (14)	7 (5)	17(16)
Summer WD	8	11	13	20 (16)	12 (10)	21 (20)
Summer WE	2	2	21	17 (16)	14 (13)	35 (32)
Spring WD	11	11	31	88 (80)	91 (84)	37 (34)
Spring WE	10	10	6	63 (57)	39 (36)	37 (36)
Fall WD	11	5	28	80 (73)	40 (36)	122 (117)
Fall WE	11	20	27	69 (63)	115 (105)	140 (135)
Average	10.4	9.0	17.8	56 (50)	46 (42)	57 (54)

the computational performance is being improved. The above analysis does not explicitly evaluate OMCs but may be interpreted as a proof of concept for the proposed algorithm. The algorithm may be particularly well-suited to replace or inform the manual commitments made by operators on a regular basis in ISO-NE (Patton, LeeVanSchaick, & Chen, 2012). An automated approach such as this may locate capacity better and avoid the potential perception of biased decisions from market operators.

Future research may investigate techniques to improve convergence times. An obvious approach would be to pre-disqualify capacity ahead of time so that fewer iterations are needed. CAISO already allows for capacity disqualification in (6.18) (CAISO, 2010), and it presumably has experience with setting that value. Computational performance may also be improved by tweaking the pruning algorithm so that larger, more restrictive, cuts are generated in each iteration (or just during earlier iterations perhaps). Larger cuts can be superficially achieved by restricting constraints in problem (ICC), and future research can develop strategies that have stronger justification or empirical performance.

6.6.3 Analysis for Fast Generators. Fast generators are defined as those that can change their commitment during scenario response. The number of fast generators varies between systems. The IEEE 73-bus test case does not specify which generators are qualified as fast. However, generators with small minimum down times can generally be started on short notice and may be reasonably placed in G^{fast} [8] . Thirty-nine small generators (13% of conventional generator capacity) are selected here because they have a minimum down time of four hours or less as shown in Table 6.5. These generators provide additional flexibility because their commitment status may change during scenario response.

The benefit provided by these fast generators is significant. The results in Table 6.6 show that good solutions can now be found using modest baseline reserve requirements. The average regret is 1.22%, halving the best results from the previous section where there were no fast generators. There is little room for improvement here because the system is very flexible at the hourly level [9] . Response sets are also applied to achieve a similar, but slightly worse, regret of 1.28%. They perform no better than the already strong baseline policies in this case.

[8] The actual classifications depend on other information not provided for the IEEE 73-bus system. For example, ISO-NE classifies fast generators as those with minimum up and down times of one hour or less and starting lead times of 30 minutes or less (ISO-NE, 2014b). These generators are automatically considered by the software. However, other generators may also be manually turned on or off when "conditions warrant" through the *reserve adequacy assessment* process (ISO-NE, 2014c)

[9] In addition to the subset of fast generators, every generator in the IEEE 73-bus test case can ramp from its minimum to its maximum output within one hour. Both of these things contribute to there being much dispatching flexibility at the hourly level.

Table 6.5: Conventional generator mix for the IEEE 73-bus test case. Generator type u50 (hydro) is not designated as fast because it is always committed due to it being inexpensive and having a minimum output of zero MW.

Type	Classification	Capacity (MW)	Ramp Rate (MW/hr)	Marginal Cost ($/MW)	Min Down Time (hr)	Count
u20	Fast	20	20	163.0	1	12
u12	Fast	12	12	94.74	2	15
u76	Fast	76	76	19.64	4	12
Fast Total:		1,332	1,332	55.62[a]	3.2[a]	39
u50	Not fast	150	150	0	0	6
u100	Not fast	100	100	75.64	8	9
u155	Not fast	155	155	15.46	8	12
u197	Not fast	197	180	74.75	10	9
u350	Not fast	350	240	15.89	48	3
u400	Not fast	400	400	5.46	48	6
Total:		10,215	9,732	32.62[a]	20.5[a]	84

[a]Capacity-weighted average

Recall that response sets were previously applied with a maximum regret of 10%. This cost bound is no longer suitable because it is easy to satisfy due to the additional flexibility from fast generators. Table 6.6 shows results from response sets with a maximum regret of $\bar{R} = 5\%$ applied alone with no other slow reserve requirements. The proposed algorithm leads to nearly the same expected costs as the baseline policies. The computational performance is volatile however: ranging from one iteration to being stopped after a two hour time limit due to the maximum regret remaining unsatisfied for some scenario. This shows that the computational performance is also sensitive to the prescribed constraint on regret (6.25). The solution time can be significantly improved by relaxing the constraint on regret at the risk of obtaining more expensive solutions.

Table 6.6: Regret for baseline and response set policies when fast generators are available assuming ν = \$10,000/MWh. There is no load imbalances for these policies. The best results are highlighted for each day.

Day	Baseline 10%	Baseline 15%	Baseline 20%	Buffered response sets with SS
Winter WD	1.36	1.00	1.13	1.30
Winter WE	1.13	1.13	1.13	1.17
Summer WD	1.41	1.36	1.25	1.37
Summer WE	1.85	1.86	1.85	1.78
Spring WD	1.11	1.10	1.19	1.08
Spring WE	2.13	1.67	1.26	1.57
Fall WD	0.88	0.88	0.96	0.93
Fall WE	1.74	1.23	2.18	1.24
Average	1.37	1.22	1.28	1.28

The policy implication is that recourse flexibility trumps the use of more advanced scheduling software. Therefore, systems with few fast generators but growing wind should investigate ways to increase dispatch flexibility during the operating day. Efforts could encourage flexible technologies or amend practices to make better use of existing resources. For example, ISO-NE only looks one hour ahead during intra-day scheduling by default (ISO-NE, 2013a), and it may be beneficial to expand their model horizons to facilitate intra-day commitments from slower generators. The proposed capacity response sets may continue to be applied after more fast generators are available, but the results in this study suggest that well-tuned baseline policies can perform just as well when that flexibility exists.

6.7 Conclusion

Power systems face greater uncertainty from growing penetrations of wind generation. The intermittency of wind implies that more aggressive rebalancing will be necessary to maintain system frequency. Capacity requirements are used by industry today to mitigate uncertainty, but the literature lacks ways to determine capacity requirements. This work employs response sets and provides an automated way to determine capacity requirements based on location, ramping capability, and cost.

Testing on the IEEE 73-bus system shows the algorithm's ability to reduce scenario costs compared to other deterministic policies from the literature. The approach is particularly beneficial when day-ahead commitments are fixed, providing expected cost improvements around 0.5%. Response sets provide an efficient way to anticipate what capacity will be available without solving a stochastic model. Future work may investigate disqualifying capacity based on scenario emissions to help ensure operators meet their emissions targets. Future work may also focus on other uncertainties, e.g., to support new reserve products being developed for short-term wind fluctuations (in the range of 5–20 minutes) (Holttinen et al., 2012).

Chapter 7

CONCLUSIONS AND FUTURE RESEARCH TOPICS

7.1 Conclusions

Power system operators adopt approximate models to find solutions in an acceptable amount of time. They mitigate uncertainty by embedding flexibility requirements into their scheduling models. These requirements approximate how much spare generation or transmission is needed to allow for disturbances. Such policies provide protection but do not guarantee reliability unless they are applied in a conservative way. Over time, this conservatism can cost millions of dollars and distort market outcomes by procuring reserves that are unavailable due to congestion. Practical benefits can come from scheduling models that better account for uncertainty.

Uncertainty and congestion in power systems are growing due to the adoption of more intermittent renewable resources. Wind supplied only 3% of the total energy in 2011 but is the fastest growing renewable resource in the US (Energy Information Administration, 2013b). The wind penetration in some areas is growing faster than investments in the transmission infrastructure (ERCOT, 2007; Gu et al., 2012), and managing congestion well through operations can postpone the need for expensive new transmission capacity. Advanced scheduling models are envisioned as a way to handle growing congestion and uncertainty in an efficient manner.

Much of the literature investigates reserve requirements while assuming all reserves are deliverable. This assumption can be improved using reserve zones, but is generally not guaranteed because recourse decisions are subject to a complex set of power flows. Several papers on stochastic and robust optimization attempt to address this issue by modeling transmission constraints on recourse decisions. However, stochastic models can only be implemented in a limited way due to the computational challenges they currently face. This dissertation contributes practical amendments to security-constrained unit commitment (SCUC) to improve economics without solving a complex stochastic model.

The first contribution, in Chapter 4, is congestion-based reserve requirements that mitigate congestion within reserve zones. The model predicts what recourse actions will be performed for different scenarios and procures more reserve if these actions stress the transmission network when applied over the baseline solution. The additional reserve allows for alternative recourse actions to circumvent congestion. More than just determining proper reserve quantities, the results show that this approach can provide efficient solutions by preemptively reducing the flow on critical lines when it is inexpensive to do so.

Congestion-based reserve requirements can be combined with many approaches from the literature to help justify the assumption that reserves are deliverable within zones. The approach is computationally tractable but does not guarantee reliability. A limitation is that the formulation must be tuned to the particular system where it is being applied. This dissertation provides an interpretation for each user-defined parameter to help guide and justify these tuning decisions. The number of parameters (two in the formulation presented) does not greatly exceed that of other reserve requirements already used in practice. Nevertheless, the approach should be evaluated on a case-by-case basis to verify if the operational benefits are worthwhile.

The second contribution builds upon reserve disqualification practices found in industry. The existing practice allows operators to tell the model what reserves should be treated as unavailable due to anticipated congestion. The work in Chapter 5 develops a mathematical structure for making such decisions. The approach generalizes existing methods by disqualifying reserves on a per-scenario basis. This creates a response set of prime reserves for each scenario. The response sets are determined by iterating between SCUC and reserve disqualification. The algorithm can find practical application by informing or replacing the manual reserve disqualifications that are performed today when the market solution is unreliable. However, the algorithm would be more impactful when used within the day-ahead market itself. Application in day-ahead markets would allow reserve disqualification to influence more commitment decisions and help the market model identify and properly compensate scarce resources.

The response set policies allow for a new market settlement scheme that rewards generators at prime locations. Locational pricing contrasts the conventional approach of one price per zone. The new settlement scheme decomposes prices by scenario to distinguish reserves that are deliverable for some scenarios but not for others. Testing shows that this scheme directs payments toward resources that can circumvent congestion during critical scenarios. The proposed model only differs from zonal pricing when reserves are disqualified, and it preserves normal settlements when the baseline SCUC solution is reliable. This is important because the locational pricing addresses undeliverable reserves without overhauling the entire settlement structure. Therefore, the barriers for implementation should be low compared to other new pricing schemes being investigated for stochastic programming (Pritchard et al., 2010; J. Zhao et al., 2013).

Lastly, response sets are applied in Chapter 6 to generalize the capacity requirements used to mitigate day-ahead wind forecast errors. The capacity requirements are especially useful when the operator has little flexibility to turn on additional units moving closer to real time. The proposed response sets are based on scenario cost, i.e., capacity is disqualified when it is expensive to produce or it exasperates constraints with high shadow prices. For example, capacity may be disqualified if it has a high operating cost or its dispatch encourages congestion or load curtailment. This approach is shown to reduce load imbalances and improve expected costs relative to other deterministic policies from the literature. The approach is particularly viable when real-time costs vary a lot due to high amounts of uncertainty, such as from deep penetrations of wind power generation.

A limitation of the response set algorithms is that they work in an iterative manner. This implies solving unit commitment multiple times over the decomposition algorithm. Some day-ahead markets may not provide enough time for convergence. In such cases, out-of-market corrections may be used (as they are today) to repair unreliable solutions. Response sets will most likely be introduced initially as a means to determine these out-of-market corrections. As the computational performance improves, the response set updates may be gradually moved into the day-ahead market as time permits. The benefits will be more efficient commitment decisions and prices coming from the day-ahead market.

Both response sets and congestion-based reserves are heuristics for addressing congestion. Both account for how congestion patterns may change between scenarios and across periods. Both improve upon existing policies in some situations. Testing demonstrates that the proposed approaches sometimes perform much better and rarely perform worse than traditional approaches. Based on empirical evidence, the benefit may fall somewhere around 0.5%–1.0% of total expected operating costs. The proposed approaches generally solve slower than traditional models but faster than existing stochastic models. Therefore, operators who are interested in stochastic models may use the proposed methods as an alternative that is more computationally tractable.

In summary, the models used to schedule power systems are limited in their ability to anticipate and mitigate congestion under uncertainty. Furthermore, uncertainty is expected to grow as systems embrace deeper penetrations of intermittent renewable resources. Stochastic models are proposed in the literature to help fix this problem but currently lack the scalability needed for practical application. This dissertation stakes a middle ground that develops deterministic models to provide better approximations in a tractable manner. The proposed approaches collect low-hanging fruit in a manner that is computationally feasible. This spares precious computing time that can be used to facilitate other modeling improvements not covered in this dissertation.

7.2 Future Work

There is an ongoing challenge to make accurate models more tractable [1] . Given the large size of practical problem instances, practical advancements in the near future are likely to come from heuristics and modifications to existing deterministic polices. This section outlines how model accuracy and solution times may be improved using variants of the proposed reserve requirements. Future work is outlined for congestion-based reserve requirements in Section 7.2.1 and for response sets in Section 7.2.2. Potential extensions for use with stochastic models are covered in Section 7.2.4.

Model development should address operational needs. Some important uncertainties have fallen outside the scope of this dissertation. While Chapter 6 focuses on day-ahead wind uncertainty, practitioners have emphasized existing needs to address short-term uncertainties (forecast errors from minutes to hours ahead of time). Many operators are considering changing 5-minute reserve requirements to mitigate wind uncertainty (Holttinen et al., 2012). CAISO and MISO are also investigating new reserve products available for dispatch with 15 to 20 minutes (Xu & Tretheway, 2012; Navid & Rosenwald, 2013). Future work may extend the research in this dissertation to better address various other forms of uncertainty, from short-term wind fluctuations to transmission contingencies to uncertain fuel availability.

[1] Even the existing deterministic models take long to solve on occasion (Chen, X. Wang, & Wang, 2014).

7.2.1 Congestion-Based Reserve Requirements.

Anticipating Recourse Decisions. Congestion-based reserve requirements make use of simplified recourse policies that anticipate when reserves are not deliverable. The adopted affine policy is based on pre-defined participation factors for how generators will respond to contingencies. More complex affine policies have been used in the literature to simplify stochastic problems (Bertsimas et al., 2011), and a similar path may be taken for congestion-based reserve requirements.

Future work may develop new methods to better predict the location of dispatched reserves. One potential affine policy is developed by Warrington et al. (2013) for real-time power operations. The authors caution their model is imprecise when applied more than a few hours ahead. However, some imprecision may be permissible because the user-defined parameters are tunable to make congestion-based reserve requirements more conservative. If the anticipated actions are infeasible, then more reserve can be procured to enable alternative actions. The recognition that imperfect recourse predictions can still be valuable under this scheme motivates new developments that may have previously been overlooked.

The Transmission Stress Function. Congestion-based reserve requirements are tied to a function of transmission stress. The implicit assumption is that marginal increments in transmission stress are counteracted by the additional reserve allowing for alternative recourse actions. This assumption can be poor in the worst case because increments and decrements in reserve may come from different locations. It is recognizable that acquiring more reserve tends to help operators circumvent congestion [2] , but it is difficult to derive a relationship describing just how much additional reserve is necessary. Future research that makes progress in this area would allow for improvements to congestion-based reserve requirements.

7.2.2 Response Sets (Reserve and Capacity Disqualification).

Pruning Response Sets Based on AC Analysis. One advantage of reserve disqualification is that the pruning model for revising response sets need not be convex. Therefore, it is possible to adapt the model to consider AC power flows. This would facilitate greater model accuracy. It would also provide a mechanism to disqualify reserves that are not available due to voltage limitations. The pruning algorithm would likely remain computationally tractable as an AC model because each contingency can be evaluated independently and in parallel.

This is an example of how taking one step back may allow two steps forward. The adoption of reserve disqualification as a heuristic means optimality is not guaranteed. However, it also affords the opportunity to achieve better computational times while reducing the reliance on imperfect linear power flow approximations. Such linear approximations can have errors upward of 20% for individual transmission lines (Stott et al., 2009). Eliminating these errors means that fewer manual adjustments are needed to improve reliability. When-

[2] The baseline results in Figure 4.13 (a) show that additional reserve tends to improve reliability, but the relationship is not monotonic. Stricter reserve requirements can make reliability worse due to the combinatorial nature of unit commitment. Research has shown that near optimal solutions can have drastically different combinations of committed generators (Sioshansi et al., 2008).

ever possible, it is preferable for the scheduling model to be more accurate so resources are coordinated in an effective way.

Multiple Contingencies and Probabilistic Pricing. Some ISOs may wish to protect against extreme events beyond N-1. Much of the literature considers probability when protecting against rarer scenarios such as multiple contingencies [3] . Such models are useful to avoid spending inordinate amounts protecting against unlikely scenarios. A common realization of this approach is to constrain or penalize the expected value of lost load caused by insufficient reserves.

Future work can apply response sets for scenarios with multiple contingencies. A probabilistic interpretation may be gained by adding a shortage variable to each requirement. Each shortage may be penalized in the objective according to the scenario probability times the value of lost load (VOLL): reserve requirements would then be relaxed when reserve is more expensive than the expected cost of curtailing load. These relaxations would filter down to the reserve prices and carry a probabilistic interpretation with them.

A similar probabilistic approach is used by Bai et al. (2006) [4] . The benefit that would be provided by response sets is improved reserve locations and pricing for multiple-contingency events. This would be accomplished through a probabilistic framework. A limitation is that the number of scenarios would grow rapidly if enumerated for all combina-

[3] See the literature review in Sections 3.1.1 and 3.1.2

[4] Bai et al. (2006) also develop a market settlement scheme for their probabilistic requirements. The scheme penalizes generators that are likely to have an outage and also penalizes reserves providers that are unlikely to perform when called upon (such as demand response resources). These strategies may be preserved by future work, if desired, because they take the reserve clearing price as an input.

tions of contingencies. Therefore, scenario selection would be a necessary and an important aspect of this work to keep the problem to a reasonable and manageable size.

Transmission Contingencies. Generator outages are a natural application for response sets because the minimum reserve quantities are well defined (reserve must exceed what that failed generator produced before the outage). Unfortunately, it is not obvious how much reserve is necessary to alleviate congestion following a transmission outage. The main barrier for applying response sets to transmission outages is the need to identify positive and negative reserve quantities.

The changes in power flows that occur immediately after a transmission outage are straightforward to calculate using line outage distribution factors (LODFs) (Wood & Wollenberg, 1996). Reserve may then be dispatched to relieve overloaded lines. The appropriate reserve quantities are sensitive to the reserve locations. For example, relatively little reserve may be needed if it can be dispatched from locations that have large effects (shift factors) on the overloaded lines. Recourse must also be careful to not overload any new lines. There are no obvious ways to accurately predict how much reserve may be exercised without employing a stochastic model or contingency analysis. Therefore, approximations seem necessary when specifying minimum reserve quantities for transmission outages.

Response sets can be applied to any scenarios that have their own reserve requirements. Once minimum reserve quantities are specified, response sets will help ensure the reserve is held at effective locations. In the case of transmission contingencies, the pruning algorithm should be updated to account for multiple reserve products (up and down) and the fact that the disturbance is no longer isolated to a single node. These changes are manageable but should be handled carefully to retain the desired properties.

Alternative Scenarios. The work in this dissertation may be adapted to mitigate other types of uncertainty besides contingencies and wind availability. For example, ISO-NE faces the risk of natural gas (NG) fuel shortages that can force generators to temporarily shut down (ISO-NE, 2014a). Fuel disruptions may occur after clearing the day-ahead energy market because many generators only procure their fuel in a NG market held later in the day, where they must compete with other industries for utilization of limited pipeline capacity [5] . Natural gas shortages are infrequent but have the potential to be catastrophic if not properly prepared for because almost half of the ISO-NE load is served by NG generation (ISO-NE, 2014a). The capacity response sets from Chapter 6 may be adapted to help locate capacity that protects against these types of scenarios. Such automated approaches may be particularly beneficial because the potential disturbances (NG shortages) are distributed across the grid in ways that can transcend traditional reserve zones.

Controlling Emissions. Cost and reliability are not the only meaningful criteria to power system operators. Another important consideration is emissions. Improving emissions is a primary goal behind efforts for renewable integration. However, minimizing cost can encourage higher emissions even when clean generators are inexpensive (e.g., curtailing wind can sometimes hurt cost but improve emissions overall by removing the need for

[5] Natural gas is extensively used for heating in the Northeastern US.

generators that can respond quickly to variants in wind production (Deng, Hobbs, & Renson, 2014)). Deterministic unit commitment models have consequently been developed to constrain or charge for emissions (Gjengedal, 1996; Catalão, Mariano, Mendes, & Ferreira, 2010), but the stochastic versions of these problems remain difficult to solve. Future work may improve emissions across scenarios via response sets, e.g., by disqualifying reserve or capacity that cannot contribute to both low costs *and* low emissions for different scenarios.

Sub-Tree Reserve Disqualification. The proposed algorithm for disqualifying reserves has a similar structure to Benders' Decomposition. Both methods can be applied by solving unit commitment (the master problem), evaluating scenarios, and applying cuts when scenario performance is unsatisfactory. The computational bottleneck in this process is the need to solve unit commitment at each iteration. The other steps are less time consuming because the subproblems are simpler and can be solved in parallel across scenarios.

Computational performance may be improved by reducing the number of times the master problem is solved. In some instances, researchers have sped up Benders' Decomposition by applying it at localized portions of the branch and bound tree of the MILP solver (Rubin, 2011; L. Wu, 2013). This approach can save time by eliminating the need to rebuild the tree at each iteration. However, there is a trade-off because this practice may duplicate efforts identifying the same cuts at different parts of the tree. Future research may investigate sub-tree reserve disqualification to see if a consistent improvement to computational performance can be found.

It should be noted that Alstom (which provides the market software for several ISOs) has proprietary heuristics that can quickly identify good upper bounds for SCUC. This capability can greatly reduce the size of the branch and bound tree, to the point that the solver will often terminate at the root node (Chen, X. Wang, & Wang, 2014). Obviously, applying cuts to localized parts of the tree would not be beneficial in such instances. Any benefit would come from when the solver spends time exploring the branch and bound tree, which still happens using Alstom's software but to a rarer extent (Chen, X. Wang, & Wang, 2014).

Pre-Disqualification and Re-Qualification. Another way to reduce the number of iterations is to disqualify reserves or capacity ahead of time through off-line procedures. Such approaches would not necessarily solve for the final response sets but try to improve initiation conditions for the decomposition algorithms from Chapters 5 and 6. New tools may be developed to ensure pre-disqualifications work well across a range of potential day-types.

When working with response sets defined ahead of time, it would be useful to have a way to re-qualify reserve or capacity as needed. Such a mechanism would loosen restrictions on the initial response sets because they could be modified in either direction (currently the response sets have to start off large because they are made progressively smaller). It would also allow for reserves or capacity to be re-qualified if they become more deliverable after some iterations or moving closer to real time.

7.2.3 Short-Term Scheduling.

While this work has focused on day-ahead scheduling, the developed methodologies may also be applied on shorter time scales. Moving closer to real time, operators may wish to amend reserve requirements to incorporate new information about the operating state. For example, operators could revise the recourse actions predicted by congestion-based reserve requirements. This would be particularly effective in places like MISO, where the initial contingency response is not determined by an optimization model but by participating factors derived from the current operating state (Chen, Gribik, & Gardner, 2014) [6] .

Future research can compare the downstream effects of the proposed reserve requirements to those of stochastic models. Operators that use stochastic day-ahead models may experience problems if they revert back to deterministic for short-term operations. Simply reverting back to deterministic models may undermine reliability and create a systematic divergence between the day-ahead and real-time markets. This would be less of an issue for the proposed methodologies because the new reserve requirements are easy to reapply in downstream applications. The general impact of day-ahead scheduling policies on short-term operations would be an interesting topic for future work.

Quick-Reserve Products. Various regions are updating their energy portfolios to include more renewable generation. Meta-analysis by Holttinen et al. (2012) indicates that most wind integration studies focus on uncertainty at the 10–30 minute range. For example,

[6] More precisely, MISO immediately starts ramping units following a contingency in proportion to the amount of reserve cleared by the real-time market (Chen, Gribik, & Gardner, 2014).

CAISO and MISO are considering requirements for 5–20 minute reserve products that can respond to short-term wind fluctuations (CAISO, 2013; Navid & Rosenwald, 2013). Reserve deliverability is not guaranteed because these products are procured on a zonal basis and are subject to the same limitations as zonal contingency reserves. Future research may extend the proposed policies to improve the deliverability of other quick-reserve products.

7.2.4 Hybrid Stochastic Models. It is common for stochastic models to retain some deterministic reserve requirements to cover scenarios that are not given stochastic treatment. For example, none of the stochastic models for wind in Section 3.3 consider contingencies and vice versa. There is potential for advanced reserve requirements to improve stochastic models and make them more scalable. Future research may investigate how the proposed approaches can be combined with stochastic models to identify economical solutions in less time.

Benders' Decomposition. One hybrid approach would be to integrate reserve disqualification with Benders' Decomposition. The two methods would combine naturally because both essentially have the same master problem; the main difference lies is in how cuts are generated. Some situations would warrant Benders' cuts over reserve disqualification because:

1. Benders' cuts are readily applicable for different uncertainties (e.g., transmission outages).

2. Benders' cuts are valid inequalities, which do not remove optimal solutions.

Reserve disqualification has been shown to quickly improve reliability when the incumbent solution has many contingency violations. Therefore, a reasonable strategy may be to use reserve disqualification when there is a large gap between the master and slave problems. Such an approach would leverage reserve disqualification early on to eliminate large portions of the feasible space and then apply Benders' cuts to refine the solution. Optimality would not be guaranteed, but such an approach could reduce the number of iterations and avoid the bloat that can sometimes occur after many iterations of adding Benders' cuts (Ryan, Wets, Woodruff, Silva-Monroy, & Watson, 2013).

Parallel Decomposition. Parallel decomposition methods such as progressive hedging and Lagrangian relaxation have been applied to stochastic unit commitment (Takriti, Birge, & Long, 1996; L. Wu et al., 2007, 2008; Papavasiliou et al., 2011; Papavasiliou & Oren, 2013). These models relax the coupling constraints between scenarios and solve the scenarios as separate subproblems. They can incorporate many scenarios because the subproblems are independent and can be solved in parallel. However, it may take an unsatisfactory number of iterations to converge to a solution. LaBove and Hedman (personal communication, June 24th, 2013) apply dynamic reserve requirements as a unifying force between scenarios, and preliminary results suggest that this approach can improve the convergence of progressive hedging. Reserve requirements may also address uncertainties that are not explicitly captured by stochastic scenarios. For example, they may improve N-1 reliability across *wind* scenarios. The policies in this dissertation may contribute to new hybrid models that improve reliability and solution times for parallel decomposition algorithms.

A potential downside is that meaningful computational improvements may necessitate strict reserve requirements. Such requirements are more likely to remove optimal and near-optimal solutions from the feasible set. Care should be taken to avoid these situations and avoid requirements that exceed necessary conditions for optimality because such requirements may lead to uneconomical solutions.

Necessary Conditions for Optimality. The heuristics in this section combine reserve requirements with established stochastic approaches. They would seize to be heuristics if the reserve requirements correspond to necessary conditions for optimality, i.e., if they do not remove all optimal solutions from the stochastic formulation's feasible space. There may be opportunities for future work to develop meaningful reserve requirements that are guaranteed to preserve optimal solutions. This could represent a strong theoretical step for making stochastic models more scalable without compromising optimality.

Scenario Selection. Scenario selection is the process of determining which scenarios will be given stochastic treatment. It has been used to reduce the size of stochastic models so they can be solved more quickly. Different scenario selection techniques have been developed to ensure a range of scenarios is represented (Gröwe-Kuska et al., 2003; Papavasiliou et al., 2011). To my knowledge, no scenario selection techniques in the literature attempt to identify which scenarios do not need stochastic treatment due to being adequately addressed by reserve requirements. Likewise, reserve requirements are not designed to reduce the number of scenarios receiving stochastic treatment. Existing techniques instead try to get the most from the respective approach when it is applied by itself. However, scenario selection and reserve requirements do not need to both cast wide nets when they are combined as part of the same algorithm. For example, one application would be to use response sets for generator contingencies and stochastic scenarios for transmission contingencies. Other opportunities may be currently overlooked because scenario selection is still a young topic for stochastic unit commitment. The present state of this topic is described well by Ruiz et al. (2009): "There is no obvious rule for the selection of optimal reserve requirements in a stochastic formulation. Substantially more research is needed in this area." The techniques and approaches developed in this dissertation may be useful for focusing scenario selection in future work.

REFERENCES

Al-Abdullah, Y., Khorsand, M. A., & Hedman, K. W. (2013, August). Analyzing the impacts of out-of-market corrections. In *IREP Symp.* (pp. 1–10). Rethymno.

Abiri-Jahromi, A., Fotuhi-Firuzabad, M., & Abbasi, E. (2007, November). Optimal scheduling of spinning reserve based on well-being model. *IEEE Trans. Power Syst. 22*(4), 2048–2057.

Alaywan, Z., Wu, T., & Papalexopoulos, A. D. (2004, October). Transitioning the California market from a zonal to a nodal framework: An operational perspective. In *IEEE Power Syst. Conf. and Exp.* (Vol. 2, pp. 862–867).

Aminifar, F., Fotuhi-Firuzabad, M., & Shahidehpour, M. (2009, February). Unit commitment with probabilistic spinning reserve and interruptible load considerations. *IEEE Trans. Power Syst. 24*(1), 388–397.

Anstine, L. T., Burke, R., Casey, J., Holgate, R., John, R. S., & Stewart, H. G. (1963, October). Application of probability methods to the determination of spinning reserve requirements for the Pennsylvania–New Jersey–Maryland Interconnection. *IEEE Trans. Power App. Syst. 82*(68), 726–735.

Audomvongseree, K. & Yokoyama, A. (2004, September). Risk based TRM evaluation by probabilistic approach. In *Int'l. Conf. on Prob. Methods Applied to Power Syst.* (pp. 254–259). Ames, Iowa.

Bai, J., Gooi, H. B., Xia, L. M., Strbac, G., & Venkatesh, B. (2006, August). A probabilistic reserve market incorporating interruptible load. *IEEE Trans. Power Syst. 21*(3), 1079–1087.

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011, August). Theory and applications of robust optimization. *Siam Rev. 53*(3), 464–501.

Bertsimas, D., Litvinov, E., Sun, X. A., Zhao, J., & Zheng, T. (2013, February). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Trans. Power Syst. 28*(1), 52–63.

Bertsimas, D. & Sim, M. (2004, February). The price of robustness. *Operations Research*, *52*(1), 35–53.

bin Othman, M. M., Mohamed, A., & Hussain, A. (2008, November). Determination of transmission reliability margin using parametric bootstrap technique. *IEEE Trans. Power Syst. 23*(4), 1689–1700.

Blumsack, S., Hines, P., Patel, M., Barrows, C., & Sanchez, E. C. (2009, July). Defining power network zones from measures of electrical distance. In *IEEE PES General Meeting* (pp. 1–8). Calgary, AB.

Bouffard, F., Galiana, F. D., & Conejo, A. J. (2005, November). Market-clearing with stochastic security – Part I: Case studies. *IEEE Trans. Power Syst. 20*(4), 1827–1835.

Bouffard, F. & Galiana, F. D. [Franscisco D.]. (2004, February). An electrical market with a probabilistic spinning reserve criterion. *IEEE Trans. Power Syst. 19*(1), 300–307.

CAISO. (2006, April). Real-time (intra-zonal) congestion. Retrieved from http://www.caiso.com/Documents/Chapter6_Real-time_Intra-Zonal_Congestion.pdf

CAISO. (2007, April). Intra-zonal congestion. Retrieved from https://www.caiso.com/1bb7/1bb77b241b920.pdf

CAISO. (2009, February). *Market parameter settings for MRTU market launch*. Retrieved from www.caiso.com/2354/2354107423420.pdf

CAISO. (2010, January). Technical bulletin: Minimum online commitment constraint. Retrieved from http://www.caiso.com/Documents/TechnicalBulletin-MinimumOnlineCommitmentConstraint.pdf

CAISO. (2013, April). 2012 annual report on market issues & performance. Retrieved from http://www.caiso.com/Documents/2012AnnualReport-MarketIssue-Performance.pdf

California Public Utilities Commission. (2013). California renewables portfolio standard (RPS). Accessed: 2013-8-1. Retrieved from http://www.cpuc.ca.gov/PUC/energy/Renewables/index.htm

Carrión, M. & Arroyo, J. M. (2006, August). A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. *IEEE Trans. Power Syst. 21*(3), 1371–1378.

Catalão, J., Mariano, S., Mendes, V., & Ferreira, L. (2010, March). A practical approach for profit-based unit commitment with emission limitations. *Elect. Power and Energy Syst. 32*(3), 218–114.

Chakraborty, S. & Funabashi, T. (2010, October). Security constrained unit commitment strategy for wind/thermal units using Lagrangian relaxation based particle swarm optimization. In *IPEC Conf. Proc.* (pp. 549–554). Singapore.

Chattopadhyay, D. & Baldick, R. (2002, August). Unit commitment with probabilistic reserve. In *IEEE PES Winter Meeting* (Vol. 1, pp. 280–285).

Chen, Y., Gribik, P., & Gardner, J. (2014, March). Incorporating post zonal reserve deployment transmission constraints into energy and ancillary service co-optimization. *IEEE Trans. Power Syst. 29*(2), 537–549.

Chen, Y., Wang, X., & Wang, Q. (2014, June). Overcoming computational challenges on large scale security constrained unit commitment (SCUC) problems – MISO and Alstom's experience with MIP solver. Presented at FERC technical conference on increasing real-time and day-ahead market efficiency through improved software. Retrieved from http://www.ferc.gov/CalendarFiles/20140623080505-M1%20-%201%20-%20FERC2014_Chen_M1_06172014.pdf

Cheung, K. W., Ma, X., & Sun, D. (2006, October). Functional design of ancillary service markets under the framework of standard market design for ISO New England. In *Int'l. Conf. on Power System Tech.* (pp. 1–7). Chongqing.

Cotilla-Sanchez, E., Hines, P. D. H., Barrows, C., Blumsack, S., & Patel, M. (2013, November). Multi-attribute partitioning of power networks based on electrical distance. *IEEE Trans. Power Syst. 28*(4), 4979–4987.

Daneshi, H. & Srivastava, A. K. (2011, July). ERCOT electricity market: Transition from zonal to nodal market operation. In *IEEE PES General Meeting* (pp. 1–7). San Diego, CA.

Das, D. (2006). *Electrical power systems*. Daryaganj, Delhi: New Age International.

Deng, L., Hobbs, B. F., & Renson, P. (2014). What is the cost of negative bidding by wind? A unit commitment analysis of cost and emissions. *IEEE Trans. Power Syst.* , accepted for publication.

Dillon, T. S., Edwin, K. W., Kochs, H.-D., & Taud, R. J. (1978, November). Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination. *IEEE Trans. Power App. Syst. PAS–97*(6), 2154–2166.

Doherty, R. & O'Malley, M. (2005, May). A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Trans. Power Syst. 20*(2), 587–595.

Dorvlo, A. S. (2002, November). Estimating wind speed distribution. *Energy Conv. and Mgmt. 43*(17), 2311–2318.

Ela, E., Milligan, M., & Kirby, B. (2011, August). *Operating reserves and variable generation*. National Renewable Energy Laboratory. Retrieved from http://www.nrel.gov/docs/fy11osti/51978.pdf

Ela, E. & O'Malley, M. (2012, August). Studying the variability and uncertainty impacts of variable generation at multiple timescales. *IEEE Trans. Power Syst. 27*(3), 1324–1333.

Energy Information Administration. (2013a). Existing transmission capacity by high-voltage size, 2011. Accessed: 2013-8-1. Retrieved from http://www.eia.gov/electricity/annual/html/epa_08_10_a.html

Energy Information Administration. (2013b, January). Summary electricity statistics 2001–2011. Accessed: 2013-8-1. Retrieved from http://www.eia.gov/electricity/data.cfm#summary

Entriken, R. & DePillis, M. S. (2009, July). Preliminary evaluation of co-optimized market design for the ISO New England. In *IEEE PES General Meeting* (pp. 1–8). Calgary, AB.

ERCOT. (2007, December). Report on existing and potential electric system constraints and needs. Retrieved from http://www.ercot.com/news/presentations/2008/35171_ERCOT_2007_Transmission_Constraints_Needs_Report.pdf

ERCOT. (2010a, July). *ERCOT protocols section 7: Congestion management*. Retrieved from http://www.ercot.com/content/mktrules/protocols/current/07-070110.doc

ERCOT. (2010b). *Setting the shadow price caps and power balance penalties in security constrained economic dispatch v0.15*. Retrieved from www.ercot.com/content/meetings/natf/keydocs/2010/0930/09_ercot_busPract_shadow_price_caps_power_balance_penal.doc

Ewing, B. T., Kruse, J. B., Schroeder, J. L., & Smith, D. A. (2007, March). Time series analyis of wind speed using VAR and the generalized impulse response technique. *Jour. of Wind Eng. and Ind. Aerod. 95*(3), 209–219.

FERC. (2005, July). Docket no. RM01-12-000 – Order terminating proceeding. Retrieved from http://www.ferc.gov/eventcalendar/Files/20050719123006-RM01-12-000.pdf

FERC. (2011, June). Docket no. ER11-2794-000 – Order conditionally accepting tariff revisions – MISO. Retrieved from http://www.ferc.gov/EventCalendar/Files/20110628160939-ER11-2794-000.pdf

FERC. (2012a, April). Docket no. ER09-1063-004 – PJM. Retrieved from http://www.pjm.com/~/media/documents/ferc/2012-orders/20120419-er09-1063-004.ashx

FERC. (2012b, July). Docket no. ER12-668-000 – Order conditionally accepting tariff revisions – MISO. Retrieved from http://www.ferc.gov/eventcalendar/Files/20120723084844-ER12-668-000.pdf

FERC. (2014a, March). Docket no. AD10-12-005 – Notice of technical conference: Increasing real-time and day-ahead market efficiency through improved software. Retrieved from http://www.ferc.gov/CalendarFiles/20140328151623-AD10-12-005TechConf.pdf

FERC. (2014b, March). The strategic plan FY 2009–2014. Retrieved from http://www.ferc.gov/about/strat-docs/FY-09-14-strat-plan-print.pdf

Ferris, M., Liu, Y., & Zhao, F. (2014, June). Modeling and computation of security-constrained economic dispatch with multi-stage rescheduling. Presented at FERC technical conference on increasing real-time and day-ahead market efficiency through improved software. Retrieved from http://www.ferc.gov/CalendarFiles/20140624105700-T4B%20-%204%20-%20SCEDslides_ferc.pdf

Fisher, R. C. (2006). *State & local public finance* (3rd ed.). South-Western College Pub.

Flynn, M., Sheridan, P., Dillan, J. D., & O'Malley, M. J. (2001, February). Reliability and reserve in competitive electricity market scheduling. *IEEE Trans. Power Syst. 16*(1), 78–87.

Galiana, F. D. [Francisco D.], Bouffard, F., Arroyo, J. M., & Restrepo, J. F. (2005, November). Scheduling and pricing of coupled energy and primary, secondary, and tertiary reserves. *Proc. IEEE*, *93*(11), 1970–1983.

Geoffrion, G. M. (1972). Generalized Benders decomposition. *Jour. of Opt. Theory and Apps. 10*(4), 237–260.

Gjengedal, T. (1996, March). Emission constrained unit-commitment (ECUC). *IEEE Trans. Energy Conv. 11*(1), 132–138.

Glover, J. D. & Sarma, M. (1994). *Power system analysis and design* (2nd ed.). Boston, Massachusetts: PWS Publishing Company.

Gooi, H. B., Mendes, D. P., Bell, K. R. W., & Kirschen, D. S. (1999, November). Optimal scheduling of spinning reserve. *IEEE Trans. Power Syst. 14*(4), 1485–1492.

Grigg, C. et al. (1999, August). The IEEE reliability test system – 1996. *IEEE Trans. Power Syst. 14*(3), 1010–1020.

Gröwe-Kuska, N., Heitsch, H., & Römisch, W. (2003, June). Scenario reduction and scenario tree construction for power management problems. In *IEEE Power Tech* (pp. 1–7). Bologna, Italy.

Gu, Y., McCalley, J. D., & Ni, M. (2012, October). Coordinating large-scale wind integration and transmission planning. *IEEE Trans. Sustain. Energy*, *3*(4), 652–659.

Guan, X., Luh, P. B., & Prasannan, B. (1996, May). Power system scheduling with fuzzy reserve requirements. *IEEE Trans. Power Syst. 11*(2), 864–869.

Hedman, K. W. (2009, May). *Flexible transmission in the smart grid* (Doctoral dissertation, Univ. of California, Berkeley).

Hedman, K. W., Ferris, M. C., O'Neill, R. P., Fisher, E. B., & Oren, S. S. (2010, May). Co-optimization of generation unit commitment and transmission switching with N-1 reliability. *IEEE Trans. Power Syst. 25*(2), 1052–1063.

Hejazi, H. A., Mohabati, H. R., Hosseinian, S. H., & Abedi, M. (2011, August). Differential evolution algorithm for security-constrained energy and reserve optimization considering credible contingencies. *IEEE Trans. Power Syst. 26*(3), 1145–1155.

Hodge, B.-M., Florita, A., Orwig, K., Lew, D., & Milligan, M. (2012, May). *A comparison of wind power and load forecasting error distributions*. National Renewable Energy Laboratory. Retrieved from http://www.nrel.gov/docs/fy12osti/54384.pdf

Hogan, W. W. (2009, October). A model for a zonal operating reserve demand curve. Harvard Univ. Retrieved from http://www.hks.harvard.edu/fs/whogan/Hogan_MIT_ORC_101509.pdf

Holttinen, H. et al. (2012, October). Methodologies to determine operating reserves due to increased wind power. *IEEE Trans. Sustain. Energy*, *3*(4), 713–723.

Hong, Y.-Y., Chang-Chien, C.-N., Wu, K.-L., & Yang, M.-S. (2002, June). Determination of congestion zones in deregulated electricity markets using fuzzy clustering. In *Proc. Power Systems Computational Conf.* (pp. 1–7). Seville.

Huang, Y., Zheng, Q. P., & Wang, J. (2014, November). Two-stage stochastic unit commitment model including non-generation resources with conditional value-at-risk constraints. *Electric Power Systems Res. 116*, 427–438.

Imran, M. & Bialek, J. W. (2008, December). Effectiveness of zonal congestion management in the European electricity market. In *Int'l. Conf. on Power and Energy* (pp. 7–12). Johor Bahru.

ISO-NE. (2011, March). *ISO New England operating procedure no. 8, operating reserve and regulation v8*. Retrieved from http://www.iso-ne.com/rules_proceds/operating/isone/op8/op8_rto_final.pdf

ISO-NE. (2012, June). *SOP-RTMKTS.0120.0030 Implement transmission remedial action v20*. Retrieved from http://www.iso-ne.com/rules_proceds/operating/sysop/rt_mkts/sop_rtmkts_0120_0030.pdf

ISO-NE. (2013a, October). Manual m-11 for market operations v44. Retrieved from http://www.iso-ne.com/static-assets/documents/2014/09/m_11_market_operations_revision_47_10_06_13.doc

ISO-NE. (2013b, February). *SOP-RTMKTS.0060.0020 Monitor system security v57*. Retrieved from http://www.iso-ne.com/rules_proceds/operating/sysop/rt_mkts/sop_rtmkts_0060_0020.pdf

ISO-NE. (2014a, January). 2014 regional electricity outlook. Retrieved from http://www.iso-ne.com/aboutiso/fin/annl_reports/2000/2014_reo.pdf

ISO-NE. (2014b). ISO New England Inc. transmission, markets, and services tariff section i: General terms and conditions. Retrieved from http://www.iso-ne.com/static-assets/documents/regulatory/tariff/sect_1/sect_i.pdf

ISO-NE. (2014c, May). *SOP-RTMKTS.0050.0010 Perform reserve adequacy assessment v53*. Retrieved from http://www.iso-ne.com/rules_proceds/operating/sysop/rt_mkts/sop_rtmkts_0050_0010.pdf

Jiang, R., Wang, J., Li, G., & Guan, Y. (2012, May). Robust unit commitment with wind power and pumped storage hydro. *IEEE Trans. Power Syst. 27*(2), 800–810.

Jiang, R., Wang, J., Zhang, M., & Guan, Y. (2013, August). Two-stage minimax regret robust unit commitment. *IEEE Trans. Power Syst. 28*(3), 2271–2282.

Jiang, R., Zhang, M., Li, G., & Guan, Y. (2014, May). Two-stage network constrained robust unit commitment problem. *Eur. Jour. of Oper. Res. 234*.

Kirby, B. J. & Dyke, J. W. V. (2002, June). *Congestion management requirements, methods and performance indices*. Oak Ridge National Laboratory. Retrieved from http://certs.lbl.gov/pdf/congestion.pdf

Kirschen, D. & Strbac, G. (2004). *Fundamentals of power systems economics*. West Sussex, England: John Wiley and Sons Ltd.

Korad, A. & Hedman, K. W. (2013, November). Robust corrective topology control for system reliability. *IEEE Trans. Power Syst. 28*(4), 2013.

Landberg, L. (2001, March). Short-term prediction of local wind conditions. *Jour. of Wind Eng. and Ind. Aerod. 89*(3–4), 235–245.

Lang, M. & Mutschler, U. (n.d.). Overview renewable energy sources act. Accessed: 2014-11-5. German Energy Blog. Retrieved from http://www.germanenergyblog.de/?page_id=283

Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., & Yan, Z. (2009, May). A review of the forecasting of wind speed and generated power. *Renewable & Sustain. Energy Reviews*, *13*(4), 915–920.

Lerner, J., Grundmeyer, M., & Garvert, M. (2009). The importance of wind forecasting. *Renewable Energy Focus*, *10*(2), 64–66.

Liu, H., Tesfatsion, L., & Chowdhury, A. A. (2009, July). Locational marginal pricing basics for restructured wholesale power markets. In *IEEE PES General Meeting* (pp. 1–9). Calgory, AB.

Liu, M. & Gross, G. (2004, May). Role of distribution factors in congestion revenue rights applications. *IEEE Trans. Power Syst. 19*(2), 802–810.

Lojowska, A. (2009, August). *Wind speed modeling* (Master's thesis, Delf Univ. of Tech.).

Lyon, J. D., Hedman, K. W., & Zhang, M. (2014, January). Reserve requirements to efficiently manage intra-zonal congestion. *IEEE Trans. Power Syst. 29*(1), 251–258.

Lyon, J. D., Wang, F., Hedman, K. W., & Zhang, M. (2014, June). Market implications and pricing of dynamic reserve policies for systems with renewables. *IEEE Trans. Power Syst.* , accepted for publication.

Lyon, J. D., Zhang, M., & Hedman, K. W. (2014, March). Locational reserve disqualification for distinct scenarios. *IEEE Trans. Power Syst.* , accepted for publication.

Makarov, Y. V., Guttromson, R. T., Huang, Z., Subbarao, K., Etingov, P. V., Chakrabarti, B. B., & Ma, J. (2010, January). *Incorporating wind generation and load forecast uncertainties into power grid operations*. U.S. Department of Energy. Retrieved from http://www.pnl.gov/main/publications/external/technical_reports/PNNL-19189.pdf

Maslennikov, S. & Litvinov, E. (2009, May). Adaptive emergency transmission rates in power system and market operation. *IEEE Trans. Power Syst. 24*(2), 923–929.

McCalley, J. et al. (2004, June). Probabilistic security assessment for power system operations. In *IEEE PES General Meeting* (Vol. 1, pp. 212–220).

MISO. (2008). Operating reserve zone modeling, September–November 2008. Printed copy of meeting minutes.

MISO. (2012a, January). *Miso energy and operating reserve markets, business practices manual, BPM-002-r11*. Retrieved from https://www.misoenergy.org/_layouts/MISO/ECM/Redirect.aspx?ID=19178

MISO. (2012b, January). Spinning reserve demand curve construct. Retrieved from https://www.misoenergy.org/_layouts/MISO/ECM/Redirect.aspx?ID=123000

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting* (1st ed.). New York: John Wiley and Sons, Inc.

Montgomery, D. C. & Runger, G. C. (2010). *Applied statistics and probability for engineers* (5th ed.). New York: John Wiley and Sons, Inc.

Morales, J. M., Conejo, A. J., & Pérez-Ruiz, J. (2009, May). Economic valuation of reserves in power systems with high penetration of wind power. *IEEE Trans. Power Syst. 24*(2), 900–910.

Morales, J. M., Mínguez, R., & Conejo, A. J. (2010, March). A methodology to generate statistically dependent wind speed scenarios. *Applied Energy*, *87*(3), 843–855.

Motto, A. L., Arroyo, J. M., & Galiana, F. D. (2005, August). A mixed-integer LP procedure for the analysis of electric grid security under disruptive threat. *IEEE Trans. Power Syst. 20*(3), 1357–1365.

Motto, A. L., Galiana, F. D., Conejo, A. J., & Arroyo, J. M. (2002, August). Network-constrained multiperiod auction for a pool-based electricity market. *IEEE Trans. Power Syst. 17*(3), 646–653.

National Renewable Energy Laboratory. (n.d.). Western wind resource dataset. Accessed December, 2013. Retrieved from http://wind.nrel.gov/Web_nrel/

Navid, N. & Rosenwald, G. (2013, July). Ramp capability product design for MISO markets. MISO. Retrieved from https://www.misoenergy.org/Library/Repository/Communication%20Material/Key%20Presentations%20and%20Whitepapers/Ramp%20Product%20Conceptual%20Design%20Whitepaper.pdf

Nemhauser, G. & Wolsey, L. (1999). *Integer and combinatorial optimization*. John Wiley and Sons, Inc.

NERC. (1996, June). *Available transfer capability definitions and determination*. NERC. Retrieved from http://www.westgov.org/wieb/wind/06-96NERCatc.pdf

NERC. (2007, December). Definition of "adequate level of reliability". Retrieved from http://www.nerc.com/docs/pc/Definition-of-ALR-approved-at-Dec-07-OC-PC-mtgs.pdf

NERC. (2014a, February). Standard BAL-001-0 – Real power balancing control performance. Retrieved from http://www.nerc.com/pa/Stand/Reliability%20Standards%20Complete%20Set/RSCompleteSet.pdf

NERC. (2014b, February). Standard BAL-002-0 – Disturbance control performance. Retrieved from http://www.nerc.com/pa/Stand/Reliability%20Standards%20Complete%20Set/RSCompleteSet.pdf

NPCC. (2012, October). *Regional reliability reference directory #5 reserve*. Retrieved from https://www.npcc.org/Standards/Directories/Directory_5-Full%20Member%20Approved%20Clean-20120914-GJD%20October%2018%202012.pdf

NYISO. (2013, December). *Ancillary service manual v4*. Retrieved from http://www.nyiso.com/public/webdocs/markets_operations/documents/Manuals_and_Guides/Manuals/Operations/ancserv.pdf

O'Neill, R. P., Sotkiewicz, P. M., Hobbs, B. F., Rothkopf, M. H., & Stewart Jr., W. R. (2005, July). Efficient market-clearing prices in markets with nonconvexities. *Eur. Jour. of Oper. Res. 164*(1), 269–285.

Ortega-Vazquez, M. A. & Kirschen, D. (2007, February). Optimizing the spinning reserve requirements using a cost/benefit analysis. *IEEE Trans. Power Syst. 22*(1), 24–33.

O'Sullivan, J. W. & O'Malley, M. J. (1999, May). A new methodology for the provision of reserve in an isolated power system. *IEEE Trans. Power Syst. 14*(2), 519–524.

Ou, Y. & Singh, C. (2002, May). Assessment of available transfer capability and margins. *IEEE Trans. Power Syst. 17*(2), 463–468.

Ozturk, U. A., Mazumdar, M., & Norman, B. A. (2004, August). A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Trans. Power Syst. 19*(3), 1589–1598.

Pandžić, H., Qiu, T., & Kirschen, D. S. (2013, July). Comparison of state-of-the-art transmission constrained unit commitment formulations. In *IEEE PES General Meeting* (pp. 1–5). Vancouver, BC.

Papavasiliou, A. & Oren, S. S. (2013, May). Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research*, *61*(3), 578–592.

Papavasiliou, A., Oren, S. S., & O'Neill, R. P. (2011, November). Reserve requirements for wind power integration: A scenario-based stochastic programming framework. *IEEE Trans. Power Syst. 26*(4), 2197–2206.

Papavasiliou, A., Oren, S. S., & Rountree, B. (2014). Applying high performance computing to transmission-constrained stochastic unit commitment for renewable energy integration. *IEEE Trans. Power Syst.* , accepted for publication.

Papic, M., Vaiman, M. Y., Vaiman, M. M., & Povolotskiy, M. (2007, July). A new approach to constructing seasonal nomograms in planning and operations environments at Idaho Power Co. In *Power tech*. Lausanne.

Parent, C. (2013a, December). Effects of minimum commitment constraints to the day-ahead and real-time markets. ISO-NE. Retrieved from http://www.iso-ne.com/committees/comm_wkgrps/mrkts_comm/mrkts/mtrls/2013/nov13142013/a03_iso_presentation_12_19_13.ppt

Parent, C. (2013b, November). Net commitment period compensation (NCPC) LSCPR cost allocation. Market Committee Presentation. ISO-NE. Retrieved from http://www.iso-ne.com/committees/comm_wkgrps/mrkts_comm/mrkts/mtrls/2013/oct892013/a07_iso_presentation_10_09_13.ppt

Patton, D. B., LeeVanSchaick, P., & Chen, J. (2012, June). 2011 assessment of the ISO New England electricity markets. Potomic Economics. Retrieved from http://www.iso-ne.com/markets/mktmonmit/rpts/ind_mkt_advsr/emm_mrkt_rprt.pdf

PJM. (2012). Reserve requirements in PJM. Printed copy of meeting minutes.

PJM. (2013, November). *PJM manual 10: Pre-scheduling operations v26*. Retrieved from http://www.pjm.com/~/media/documents/manuals/m10.ashx

Pritchard, G., Zakeri, G., & Philpott, A. (2010, April). A single-settlement, energy-only electric power market for unpredictable and intermittent participants. *Operations Research*, *58*(4-part-2), 1210–1219.

Purchala, K., Haesen, E., Meeus, L., & Belmans, R. (2005, October). Zonal network model of European interconnected electricity network. In *CIGRE/IEEE PES Int'l Symp.* (pp. 362–369). New Orleans, LA.

Radi, K. M. & Fox, B. (1991, July). Power system economic loading with flexible emergency reserve provision. *Proc. Inst. Elect. Eng. C*, *138*(4), 257–262.

Rajan, D. & Takriti, S. (2005, December). *Minimum up/down polytopes of the unit commitment problem with start-up costs*. IBM Research. Retrieved from http://www.research.ibm.com/people/d/dpkrjn/

Ramezani, M., Singh, C., & Haghifam, M.-R. (2009, May). Role of clustering in the probabilistic evaluation of TTC in power systems including wind power generation. *IEEE Trans. Power Syst. 24*(2), 849–858.

Rothwell, G. & Gómez, T. (2003). *Electricity markets: Regulation and deregulation*. IEEE Press Power Systems Engineering. John Wiley and Sons, Inc.

Rubin, P. (2011, October). Benders decomposition then and now. Retrieved from http://orinanobworld.blogspot.com/2011/10/benders-decomposition-then-and-now.html

Ruiz, P. A., Philbrick, C. R., Zak, E., Cheung, K. W., & Sauer, P. W. (2009, May). Uncertainty management in the unit commitment problem. *IEEE Trans. Power Syst. 24*(2), 642–651.

Ryan, S. M., Wets, R. J.-B., Woodruff, D. L., Silva-Monroy, C., & Watson, J.-P. (2013, July). Toward scalable, parallel progressive hedging for stochastic unit commitment. In *IEEE PES General Meeting* (pp. 1–5). Vancouver, BC.

Salmeron, J., Wood, K., & Baldick, R. (2004, May). Analysis of electric grid security under terrorist threat. *IEEE Trans. Power Syst. 19*(2), 905–912.

Salmeron, J., Wood, K., & Baldick, R. (2009, February). Worst-case interdiction analysis of large-scale electric power grids. *IEEE Trans. Power Syst. 24*(1), 96–104.

Savage, L. T. (1951, March). The theory of statistical decision. *Jour. of Amer. Statist. Assoc. 46*(253), 55–67.

Shields, M., Boughner, M., Jones, R., & Tackett, M. (2007, August). Market subcommittee minutes/ASM market design. Printed copy. MISO.

Sioshansi, R., O'Neill, R., & Oren, S. S. (2008, May). Economic consequences of alternative solution methods for centralized unit commitment in day-ahead electricity markets. *IEEE Trans. Power Syst. 23*(2), 344–352.

Sioshansi, R. & Short, W. (2009, May). Evaluating the impacts of real-time pricing on the usage of wind generation. *IEEE Trans. Power Syst. 24*(2), 516–524.

Smith, A. (1869). *An inquiry into the nature and causes of the wealth of nations*. Clarendon Press.

S´ooder, L. (2004, September). Simulation of wind speed forecast errors for operation planning of multi-area power systems. In *Int'l. Conf. on Prob. Methods Applied to Power Syst.* (pp. 1–6). Ames, Iowa.

SPP. (2013, November). *Market protocols: SPP integrated marketplace, v17*. Retrieved from http://www.spp.org/publications/Integrated%20Marketplace%20Protocols%2017.0.pdf

Stott, B., Jardim, J., & Alsaç, O. (2009, August). DC power flow revisited. *IEEE Trans. Power Syst. 24*(3), 1290–1300.

Sullivan, M. J., Mercurio, M., Schellenberg, J., & Freeman, Sullivan & Co. (2009, June). *Estimated value of service reliability for electric utility customers in the United States*. Lawrence Berkeley National Laboratory. Retrieved from certs.lbl.gov/pdf/lbnl-2132e.pdf

Takriti, S., Birge, J. R., & Long, E. (1996, August). A stochastic model for the unit commitment problem. *IEEE Trans. Power Syst. 11*(3), 1497–1508.

Törnqvist, L., Vartia, P., & Vartia, Y. O. (1985, February). How should relative changes be measured? *The Amer. Statist. 39*(1), 43–46.

Torres, J., García, A., Blas, M. D., & Francisco, A. D. (2005, July). Forecasting of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, *79*(1), 65–77.

Tseng, C.-L., Oren, S. S., Svoboda, A. J., & Johnson, R. B. (1999, May). Price-based adaptive spinning reserve requirements in power system scheduling. *Elect. Power and Energy Syst. 21*, 137–145.

Tuohy, A., Meibom, P., Denny, E., & O'Malley, M. (2009, May). Unit commitment for systems with significant wind penetration. *IEEE Trans. Power Syst. 24*(2), 592–601.

Univ. of Washington. (1999, June). Power systems test case archive. Retrieved from http://www.ee.washington.edu/research

U.S. Government Accountability Office. (2011). Federal oversight of electricity markets and infrastructure. Accessed: 2013-9-30. Retrieved from http://www.gao.gov/key_issues/federal_oversight_of_electricity_markets_and_infrastructure/issue_summary

Vos, K. D. & Driesen, J. (2014, August). Dynamic operating reserve strategies for wind power integration. *IEEE Trans. Power Syst. 8*(6), 598–610.

Wan, Y.-H., Ela, E., & Orwig, K. (2010, May). *Development of an equivalent wind plant power-curve*. National Renewable Energy Laboratory. Retrieved from http://www.nrel.gov/docs/fy10osti/48146.pdf

Wang, F. & Hedman, K. W. (2012, September). Reserve zone determination based on statistical clustering methods. In *North Amer. Power Symp.* (pp. 1–6). Champaign, IL.

Wang, F. & Hedman, K. W. (2014). Dynamic reserve zones for day-ahead unit commitment with renewable resources. *IEEE Trans. Power Syst.* , accepted for publication.

Wang, J., Shahidehpour, M., & Li, Z. (2008, August). Security-constrained unit commitment with volatile wind power generation. *IEEE Trans. Power Syst. 23*(3), 1319–1327.

Wang, J., Shahidehpour, M., & Li, Z. (2009, August). Contingency-constrained reserve requirements in joint energy and ancillary services auction. *IEEE Trans. Power Syst. 24*(3), 1457–1468.

Wang, Q., Jean-Paul, & Guan, Y. (2013, August). Two-stage robust optimization for N-k contingency-constrained unit commitment. *IEEE Trans. Power Syst. 28*(3), 2366–2375.

Wang, Q., Wang, X., Cheung, K., Guan, Y., & Bresler, F. S. (2013, June). A two-stage robust optimization for PJM look-ahead unit commitment. In *IEEE Power Tech* (pp. 1–6). Grenoble.

Warrington, J., Goulart, P., Mariéthoz, S., & Morari, M. (2013, November). Policy-based reserves for power systems. *IEEE Trans. Power Syst. 28*(4), 4427–4437.

Weaver, S. (2012, June). Assessment of resource performance and operating reserve requirements. ISO-NE. Retrieved from www.iso-ne.com/committees/comm_wkgrps/relblty_comm/relblty/mtrls/2012/jun202012/a7_tmnsr_requirement_analysis.pptx

WECC. (2008, October). *Standard BAL-002-WECC-1 – Contingency reserves*. Retrieved from http://www.nerc.com/files/BAL-002-WECC-1.pdf

Wood, A. J. & Wollenberg, B. F. (1996). *Power generation, operation, and control* (2nd ed.). New York: John Wiley and Sons, Inc.

Wu, H. & Gooi, H. B. (1999, January). Optimal scheduling of spinning reserve with ramp constraints. In *IEEE PES Winter Meeting* (Vol. 2, pp. 785–790).

Wu, L. (2013, November). An improved decomposition framework for accelerating LSF and BD based methods for network-constrained UC problems. *IEEE Trans. Power Syst. 28*(4), 3977–3986.

Wu, L., Shahidehpour, M., & Li, T. (2007, May). Stochastic security-constrained unit commitment. *IEEE Trans. Power Syst. 22*(2), 800–811.

Wu, L., Shahidehpour, M., & Li, T. (2008, August). Cost of reliability analysis based on stochastic unit commitment. *IEEE Trans. Power Syst. 23*(3), 1364–1374.

Xia, L. M., Gooi, H. B., & Bai, J. (2005, May). A probabilistic reserve with zero-sum settlement scheme. *IEEE Trans. Power Syst. 20*(2), 993–1000.

Xiao, F. & McCalley, J. D. (2009, February). Power system risk assessment and control in a multiobjective framework. *IEEE Trans. Power Syst. 24*(1), 78–85.

Xiao, J., Hodge, B.-M. S., Pekny, J. F., & Reklaitis, G. V. (2011, September). Operating reserve policies with high wind power penetration. *Computers and Chem. Eng. 35*(9), 1876–1885.

Xie, K. & Billinton, R. (2009, April). Considering wind speed correlation of WECS in reliability evaluation using time-shifting technique. *Electric Power Systems Res. 79*(4), 687–693.

Xu, L. & Tretheway, D. (2012, October). Flexible ramping products. Retrieved from http : / / www . caiso . com / documents / secondreviseddraftfinalproposal - flexiblerampingproduct.pdf

Zhao, J., Zheng, T., Litvinov, E., & Zhao, F. (2013, June). Pricing schemes for two-stage market clearing models. Presented at FERC technical conference on increasing real-time and day-ahead market efficiency through improved software. Retrieved from http://www.ferc.gov/CalendarFiles/20140411125647-M3%20-%20Zhao.pdf

Zhao, L. & Zeng, B. (2012a, January). *An exact algorithm for power grid interdiction problem with line switching*. Unpublished. Retrieved from http://www.optimization-online.org/DB_HTML/2012/01/3309.html

Zhao, L. & Zeng, B. (2012b, January). *An exact algorithm for two-stage robust optimization with mixed integer recourse problems*. Unpublished. Retrieved from http://imse.eng.usf.edu/faculty/bzeng/MOChA_group/papers/optimization/2SRO_MIP.pdf

Zheng, Q. P., Wang, J., & Liu, A. L. (2014, September). Stochastic optimization for unit commitment – A review. *IEEE Trans. Power Syst.* , accepted for publication.

Zheng, T. & Litvinov, E. (2008, May). Contingency-based zonal reserve modeling and pricing in a co-optimized energy and reserve market. *IEEE Trans. Power Syst. 23*(2), 277–286.

APPENDIX A

ACRONYMS

BA	Balancing authority, such as an ISO or vertically integrated utility
CAISO	California ISO
DAM	Day-ahead market
DCOPF	Direct current optimal power flow
EENS	Expected energy not served (MWh)
ERCOT	Electric Reliability Council of Texas (ISO)
FERC	Federal Energy Regulatory Commission
GSF	Generation shift factor
ISO	Independent system operator
ISO-NE	ISO New England
LMP	Locational marginal price of energy ($/MWh)
LODF	Line outage distribution factor
LOLP	Loss of load probability
LSE	Load-serving entity
MISO	Midcontinent ISO
MW	Megawatt (unit of power)
MWh	Megawatt hour (unit of energy)
N-1	Failure of one generator or transmission branch
N-k	Simultaneous failure of k resources
NERC	North American Electric Reliability Corporation

NYISO	New York ISO
OMC	Out-of-merit correction
PJM	Pennsylvania – Jersey – Maryland (ISO)
PTDF	Power transfer distribution factor
RMP	Reserve marginal price for a particular reserve product ($/MW)
RTM	Real-time market
SCUC	Security-constrained unit commitment
SCOPF	Security-constrained optimal power flow
TRM	Transmission reliability margin
VOLL	Value of lost load ($/MW or $/MWh)

APPENDIX B

NOMENCLATURE

B.1 Sets

G	Generators and reserve providers.
$G(n) \subseteq G$	Generators located at node n.
$\mathscr{G}(k) \subseteq G$	Generators located in zone k.
H	Transmission stress intervals.
$I(g)$	Production intervals for generator g.
L	Transmission lines (also including transformers).
$LO(n) \subseteq L$	Transmission lines out of node n (beginning at node n).
$LI(n) \subseteq L$	Transmission lines into node n (ending at node n).
N	Nodes; $n(g) \in N$ is the location of resource g.
$\mathscr{P}$	Reserve products. Each product has a time component and a type of reserve (e.g., 10-minute spinning reserve).
$\mathscr{P}(\text{sp}) \subseteq \mathscr{P}$	Spinning reserve products.
$\mathscr{P}(\text{nsp}) \subseteq \mathscr{P}$	Non-spinning reserve products.
S	Wind scenarios.
$\tilde{S} \subseteq S$	Wind scenarios for which to disqualify capacity.
T	Time periods.
Z	Reserve zones; $z(c) \in Z$ is the zone where contingency c occurs.

B.2 Parameters

In the following, index t represents the scheduling period.

a_{lt}^{g}	Operational generation shift factor on line l following failure of generator g.
d_{lt}^{k}	Operational line outage distribution factor on line l following failure of line k.
B_l	Electrical susceptance of line l.
c_g^i	Marginal production cost of generator g in interval i.
c_g^{NL}	Fixed cost to operate generator g (no-load cost).
c_g^{SU}	Start-up cost of generator g.
c_g^{SD}	Shut-down cost of generator g.
$\mathscr{C}_{gt}^{s}$	Committed capacity from generator g that belongs to the response set for scenario s.
$\mathscr{U}_{gt}^{s}$	Uncommitted capacity from generator g that belongs to the response set for scenario s.
$\underline{C}^s$	Perfect-forecast solution bound for scenario s.
D_{nt}	Demand for power at node n (load).
F_l	Capacity of line l during normal operating conditions.
$\tilde{F}_l^{10}$	Capacity of line l imposed 10 minutes after a disturbance; $\tilde{F}_l^{10} \geq F_l$. This is the limit used during contingency analysis (see Figure 2.5 and discussion).
$\hat{I}_n$	Net power injection at node n prior to re-dispatch (for a particular scenario-period).

K_{nt}^{s} Wind power available at node n during scenario s.

$LODF_{l}^{k}$ Sensitivity of flow on line l to the outage of line k, i.e., $\Delta f_{lt} = LODF_{l}^{k} f_{kt}$.

n Factor discounting transmission stress for unutilized lines.

$\tilde{P}_{g}^{i}$ Size of operating interval i for generator g.

$P_{g}^{\max}$ Maximum operating level of generator g.

$P_{gt}^{\max,s}$ Maximum operating level of generator g in scenario s, which fluctuates for wind generators.

$P_{g}^{\min}$ Minimum operating level of generator g.

$PTDF_{nl}$ Sensitivity of flow on line l to injections at node n or from generator n.

$\hat{P}_{gt}^{s}$ Portion of capacity from generator g in the response set for scenario s.

Q_{t}^{s} Required capacity for scenario s.

R_{g} Ramping capability of generator g over a normal scheduling period.

R_{g}^{SU} Ramping capability of generator g over a start up period; $R_{g}^{SU} \geq R_{g}$.

R_{g}^{SD} Ramping capability of generator g over a shut down period; $R_{g}^{SD} \geq R_{g}$.

$\dot{R}_{g}^{\pi}$ Ramping capability of resource g for reserve product π (e.g., 10-minute ramping).

$\hat{R}_{g}$ Reserve held by resource g (for a particular scenario-period).

$\bar{R}$ Prescribed maximum percent regret for any wind scenario.

UT_{g} Minimum up time for generator g.

DT_{g} Minimum down time for generator g.

W Slow reserve requirement percent of the wind forecast.

α Conservativity factor for congestion-based reserve requirements.

β Presumed reserve needed to alleviate a marginal increase in predicted line utilization.

Γ_{gt}^{c} Contingency response set membership; $\Gamma_{gt}^{c} \in [0,1]$.

$\hat{\Gamma}_{g}$ Contingency response set membership of resource g from the most recent iteration (for a particular scenario-period).

ν Assumed value of lost load balance.

ϕ_{gt} Reserve payment to generator g.

φ^{s} Slope of piecewise linear segment s for congestion-based reserve requirements.

ω^{s} Width of piecewise linear segment s for congestion-based reserve requirements.

ρ_{it}^{c} Assumed participation of resource i in the response to outage c.

B.3 Variables

In the following, index t represents the scheduling period.

b_{nt} Flow imbalance at node n.

e_{g} Exercised reserve from resource g (for a particular scenario-period).

f_{lt} Power flow on line l.

i_{nt} Net power injection at node n.

k_{nt}^{s} Wind production at node n during scenario s.

m_g Undeliverable reserve from resource g (for a particular scenario-period).

m_{gnt}^{s} Costly capacity from resource g moved to node n as a dummy injection for scenario s.

p_{gt} Power produced by generator g.

$\tilde{p}_{gt}^{i}$ Power produced by generator g in operating interval i.

r_{gt} Reserve from resource g.

$\dot{r}_{gt}^{\pi}$ Reserve from resource g of a specific product π.

$\tilde{r}_{kt}^{x}$ Reserve in zone k classified as deliverable to zone x or contingency x.

r_{gt}^{slow} Reserve from resource g available within one scheduling period (e.g., 60-minute reserve).

S_{kt}^{j} Reserve sharing limit from zone k to zone j (may also be specified as a parameter).

u_{gt} Binary commitment indicator for generator g.

v_{gt} Start up indicator for generator g.

w_{gt} Shut down indicator for generator g.

γ_g Proportion of reserve from resource g that is not deliverable (for a particular scenario-period).

γ_{gt}^{s} Proportion of capacity from resource g that is costly for scenario s.

δ_{lt}^{+} Predicted largest increase of flow on line l after re-dispatch for any scenario.

δ_{lt}^{-} Predicted largest decrease of flow on line l after re-dispatch for any scenario.

η_t Minimum reserve margin.

θ_{nt} Voltage angle at node n.

λ_{kt}^{c} Dual variable for the constraint classifying reserve in zone k as deliverable for contingency c.

μ_{lt} Utilization of line l.

$\tilde{\mu}_{lt}^{s}$ Utilization of line l across segment s.

APPENDIX C

POWER FLOW SCHEDULING MODELS

C.1 Unit Commitment

The formulation below is for the unit commitment problem, which serves as a base for the methods proposed in this dissertation. The unit commitment problem determines which generators (units) to turn on (commit) in order to minimize cost subject to physical and operational constraints. A description of the model components is provided in Sections 2.2.2, 2.2.3, and 2.3.1. The nomenclature is described in Appendix B.

$$\text{Minimize: } \sum_{g\in G}\sum_{t\in T}\left(\sum_{i\in I(g)} c_g^i \tilde{p}_{gt}^i + c_g^{\mathrm{NL}} u_{gt} + c_g^{\mathrm{SU}} v_{gt} + c_g^{\mathrm{SD}} w_{gt}\right) \tag{C.1}$$

Subject to:

Piecewise production segments:

$$\sum_{i\in I} \tilde{p}_{gt}^i = p_{gt}, \qquad \forall g \in G, t \in T \tag{C.2}$$

$$0 \le \tilde{p}_{gt}^i \le \tilde{P}_g^i, \qquad \forall g \in G, i \in I(g), t \in T \tag{C.3}$$

Transmission constraints:

$$\sum_{l\in LI(n)} f_{lt} - \sum_{l\in LO(n)} f_{lt} + \sum_{g\in G(n)} p_{gt} = D_{nt}, \qquad \forall n \in N, t \in T \tag{C.4}$$

$$f_{lt} = B_l\left(\theta_{nt} - \theta_{mt}\right), \qquad \forall l = (m,n) \in L, t \in T \tag{C.5}$$

$$-F_l \le f_{lt} \le F_l, \qquad \forall l \in L, t \in T \tag{C.6}$$

Generator capacity and ramping constraints:

$$P_g^{\min} u_{gt} \leq p_{gt} \leq P_g^{\max} u_{gt}, \qquad \forall g \in G, t \in T \tag{C.7}$$

$$p_{gt} - p_{g,t-1} \leq R_g u_{g,t-1} + R_g^{\mathrm{SU}} v_{gt}, \qquad \forall g \in G, t \in T \tag{C.8}$$

$$p_{gt} - p_{g,t-1} \geq -R_g u_{gt} - R_g^{\mathrm{SD}} w_{gt}, \qquad \forall g \in G, t \in T \tag{C.9}$$

Minimum up and down times:

$$v_{gt} - w_{gt} = u_{gt} - u_{g,t-1}, \qquad \forall g \in G, t \in T \tag{C.10}$$

$$\sum_{q=t-UT_g+1}^{t} v_{gq} \leq u_{gt}, \qquad \forall g \in G, t \in T \tag{C.11}$$

$$\sum_{q=t-DT_g+1}^{t} w_{gq} \leq 1 - u_{gt}, \qquad \forall g \in G, t \in T \tag{C.12}$$

Reserve availability:

$$\dot{r}_{gt}^{\pi} \leq P_g^{\max} - p_{gt}, \qquad \forall g \in G, \pi \in \mathscr{P}(\mathrm{sp}), t \in T \tag{C.13}$$

$$\dot{r}_{gt}^{\pi} \leq \dot{R}_g^{\pi} u_{gt}, \qquad \forall g \in G, \pi \in \mathscr{P}(\mathrm{sp}), t \in T \tag{C.14}$$

$$\dot{r}_{gt}^{\pi} \leq \dot{R}_g^{\pi} (1 - u_{gt}), \qquad \forall g \in G, \pi \in \mathscr{P}(\mathrm{nsp}), t \in T \tag{C.15}$$

Variable bounds:

$$u_{gt} \in \{0, 1\}, \qquad \forall g \in G, t \in T \tag{C.16}$$

$$0 \leq v_{gt}, w_{gt} \leq 1, \qquad \forall g \in G, t \in T \tag{C.17}$$

The problem is labeled security-constrained unit commitment (SCUC) when reserve requirements are adopted to improve reliability. The most basic reserve requirement is for

10-minute reserves (spinning plus non-spinning) to exceed the largest generator contingency:

$$\sum_{h \in G \setminus g} \left(\dot{r}_{ht}^{10(\mathrm{sp})} + \dot{r}_{ht}^{10(\mathrm{nsp})} \right) \geq p_{gt}, \qquad \forall g \in G, t \in T. \tag{C.18}$$

This requirement is a necessary condition for N-1 reliability. More complex and restrictive requirements are discussed throughout this dissertation to help ensure enough reserves are deliverable. For simplicity and generality, the available reserve is usually referenced as r_{gt} without specifying the type of reserve product(s) involved. The equivalent to (C.18) under the simplified notation is

$$\sum_{h \in G \setminus g} r_{ht} \geq p_{gt}, \qquad \forall g \in G, t \in T. \tag{C.19}$$

Reserve requirements help ensure there is a feasible contingency response. A model for contingency response is formulated in the next section. The model is presented as a stand-alone problem for contingency analysis, but it can also be combined with unit commitment to create a two-stage stochastic model.

C.2 Contingency Analysis

The formulation below evaluates recourse actions for scenarios with one or more contingencies. The set C contains the contingency scenarios. Contingency response balances supply and demand within 10 minutes while minimizing the sum of transmission violations s^{+} and s^{-} [1] . The parameter A_x^c indicates the availability of resource x during contingency $c \in C$ (it takes the value one if the resource is available and zero otherwise). The decision variables are also indexed by scenario, where $c = 0$ represents the base conditions before any contingency occurs. Generator ramping is constrained in (C.25) and (C.26) relative to the pre-contingency conditions. On-line generators can ramp by $\dot{R}^{10(\mathrm{sp})}$ and off-line generators can ramp by $\dot{R}_g^{10(\mathrm{nsp})}$. Non-spinning reserves may be omitted or ignored by setting all $\dot{R}_g^{10(\mathrm{nsp})} = 0$, which disallows participation from units that were not committed prior to the contingency.

[1] The problem is feasible whenever the network is connected (there are no islands) and the reserve quantity exceeds the injections lost from generator contingencies.

$$\text{Minimize: } \sum_{c \in C} \sum_{t \in T} \sum_{l \in L} \left(s_{lt}^{c+} + s_{lt}^{c-} \right) \tag{C.20}$$

Subject to:

Transmission constraints:

$$\sum_{l \in LI(n)} f_{lt}^{c} - \sum_{l \in LO(n)} f_{lt}^{c} + \sum_{g \in G(n)} A_g^c p_{gt}^{c} = D_{nt}, \qquad \forall c \in C, n \in N, t \in T \tag{C.21}$$

$$f_{lt}^{c} = A_l^c B_l \left(\theta_{nt}^{c} - \theta_{mt}^{c} \right), \qquad \forall c \in C, l = (m,n) \in L, t \in T \tag{C.22}$$

$$-\tilde{F}_l^{10} \leq f_{lt}^{c} + s_{lt}^{c-} - s_{lt}^{c+} \leq \tilde{F}_l^{10}, \qquad \forall c \in C, l \in L, t \in T \tag{C.23}$$

Generator capacity and reserve ramping constraints:

$$P_g^{\min} u_{gt}^{c} \leq p_{gt}^{c} \leq P_g^{\max} u_{gt}^{c}, \qquad \forall c \in C, g \in G, t \in T \tag{C.24}$$

$$p_{gt}^{c} - p_{gt}^{0} \leq \dot{R}_g^{10(\mathrm{sp})} u_{gt}^{0} + \dot{R}_g^{10(\mathrm{nsp})} \left(1 - u_{gt}^{0} \right), \qquad \forall c \in C, g \in G, t \in T \tag{C.25}$$

$$p_{gt}^{c} - p_{gt}^{0} \geq -\dot{R}_g^{10(\mathrm{sp})}, \qquad \forall c \in C, g \in G, t \in T \tag{C.26}$$

Variable bounds:

$$s_{lt}^{c+}, s_{lt}^{c-} \geq 0, \qquad \forall c \in C, l \in L, t \in T \tag{C.27}$$

$$u_{gt}^{c} \in \{0,1\}, \qquad \forall c \in C, g \in G, t \in T \tag{C.28}$$

APPENDIX D

SMALL TEST CASE FORMULATION

This appendix describes the test case used in Sections 4.6 and 5.6 to illustrate congestion-based reserve requirements and reserve disqualification. The system has three nodes, three generators, and three transmission lines. Figure D.1 provides a graphical interpretation of the network. Tables D.1 and D.2 contain the generator and transmission attributes. All 60 MW of load is at the same location as G1. The load can be served by this generator alone, but the cheapest way to operate the system is to import power from generators G2 and G3.

Table D.1: Generator characteristics of the three-node system.

Generator	$P^{\min}$ (MW)	$P^{\max}$ (MW)	R^{10} (MW)	Cost/MWh
G1	0	60	60	\$100
G2	0	45	15	\$50
G3	0	45	15	\$75

Table D.2: Transmission characteristics of the three-node network.

Line	Rating F (MW)	Rating F^{10} (MW)	PTDFs		
			G1	G2	G3
L1	30	33	0	$2/3$	$1/3$
L2	30	33	0	$1/3$	$2/3$
L3	999	999	0	$1/3$	$-1/3$

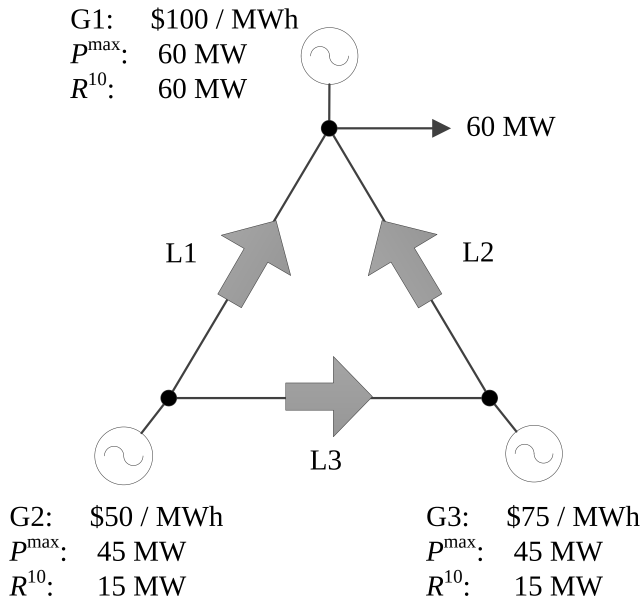


Figure D.1: Three-node network (the arrows indicate reference directions).

A single-period DCOPF is used to schedule the system. There is a reliability criterion that the system must be able to restore the balance between supply and demand within ten minutes following a generator contingency. The failure of G2 and G3 are trivial because both can be replaced by G1 without increasing the flow on any constrained line [1] . This example does not consider transmission contingencies. The deterministic formulation is provided by equations (D.1)–(D.19) below. The recourse decisions for the outage of G1 are described by (D.20)–(D.19). The N-1 reliable space is determined by taking all of these constraints together.

[1] Lost injection from G2 or G3 can always be substituted by p_1 because G1 has enough capacity to serve all of the load and can ramp to full capacity during the contingency re-dispatching period. These substitutions always decrease the right-hand-side of (D.6)–(D.7), but never below zero. Therefore, the flow on L1 and L2 will not be constrained. Furthermore, the flow on L3 is never constrained because its rating exceeds the total load. Therefore, there is always a sufficient amount of deliverable reserve available to respond to the failure of G2 or G3. This is an intuitive result because G1 provides a lot of reserve at a prime location.

$$\text{Minimize: } 40p_1 + 100p_2 + 75p_3 \tag{D.1}$$

Subject to:

Transmission constraints:

$$i_1 = p_1 - 60 \tag{D.2}$$

$$i_2 = p_2 \tag{D.3}$$

$$i_3 = p_3 \tag{D.4}$$

$$i_1 + i_2 + i_3 = 0 \tag{D.5}$$

$$f_1 = \frac{2}{3}i_2 + \frac{1}{3}i_3 \tag{D.6}$$

$$f_2 = \frac{1}{3}i_2 + \frac{2}{3}i_3 \tag{D.7}$$

$$f_3 = \frac{1}{3}i_2 - \frac{1}{3}i_3 \tag{D.8}$$

$$-30 \leq f_1 \leq 30 \tag{D.9}$$

$$-30 \leq f_2 \leq 30 \tag{D.10}$$

$$-100 \leq f_3 \leq 100 \tag{D.11}$$

Generator constraints:

$$0 \leq p_1 \leq 60 \tag{D.12}$$

$$0 \leq p_2 \leq 45 \tag{D.13}$$

$$0 \leq p_3 \leq 45 \tag{D.14}$$

$$r_2 \leq 45 - p_2 \tag{D.15}$$

$$r_3 \leq 45 - p_3 \tag{D.16}$$

$$r_2 \leq 15 \tag{D.17}$$

$$r_3 \leq 15 \tag{D.18}$$

Reserve requirement with no locational component:

$$r_2 + r_3 \geq p_1 \tag{D.19}$$

Recourse constraints for the outage G1:

$$i_1^1 = -60 \tag{D.20}$$

$$i_2^1 = p_2^1 \tag{D.21}$$

$$i_3^1 = p_3^1 \tag{D.22}$$

$$i_1^1 + i_2^1 + i_3^1 = 0 \tag{D.23}$$

$$f_1^1 = \frac{2}{3} i_2^1 + \frac{1}{3} i_3^1 \tag{D.24}$$

$$f_2^1 = \frac{1}{3} i_2^1 + \frac{2}{3} i_3^1 \tag{D.25}$$

$$f_3^1 = \frac{1}{3} i_2^1 - \frac{1}{3} i_3^1 \tag{D.26}$$

$$-33 \leq f_1^1 \leq 33 \tag{D.27}$$

$$-33 \leq f_2^1 \leq 33 \tag{D.28}$$

$$-999 \leq f_3^1 \leq 999 \tag{D.29}$$

$$0 \leq p_2^1 \leq 45 \tag{D.30}$$

$$0 \leq p_3^1 \leq 45 \tag{D.31}$$

$$p_2^1 \leq p_2 + 15 \tag{D.32}$$

$$p_3^1 \leq p_3 + 15 \tag{D.33}$$

$$p_2^1 \geq p_2 - 15 \tag{D.34}$$

$$p_3^1 \geq p_3 - 15. \tag{D.35}$$

The dual of the recourse problem is useful for the analysis in Section 5.6.3 that applies Benders' Decomposition. It is convenient to simplify the primal problem by substituting for the variables i and f. The problem can also be simplified by noting that generators G2 and G3 will not be limited when ramping down and that lines L1 and L2 will not be constrained in the negative direction. The resulting primal and dual formulations are provided below with the dual variable for each respective constraint shown in parenthesis.

$(Primal)$:

$$\text{Minimize: } 0 \tag{D.36}$$

Subject to:

$$p_2^1 + p_3^1 = 60 \qquad (\pi_1) \tag{D.37}$$

$$p_2^1 \leq 45 \qquad (\pi_2) \tag{D.38}$$

$$p_3^1 \leq 45 \qquad (\pi_3) \tag{D.39}$$

$$p_2^1 \leq p_2 + 15 \qquad (\pi_4) \tag{D.40}$$

$$p_3^1 \leq p_3 + 15 \qquad (\pi_5) \tag{D.41}$$

$$\frac{2}{3}p_2^1 + \frac{1}{3}p_3^1 \leq 33 \qquad (\pi_6) \tag{D.42}$$

$$\frac{1}{3}p_2^1 + \frac{2}{3}p_3^1 \leq 33 \qquad (\pi_7) \tag{D.43}$$

$$p_2^1, p_2^2 \geq 0 \tag{D.44}$$

$(Dual)$:

$$\text{Maximize: } 60\pi_1 + 45(\pi_2 + \pi_3) + (p_2 + 15)\pi_4$$

$$+ (p_3 + 15)\pi_5 + 33(\pi_6 + \pi_7) \tag{D.45}$$

Subject to:

$$\pi_1 + \pi_2 + \pi_4 + \frac{2}{3}\pi_6 + \frac{1}{3}\pi_7 \leq 0 \qquad (p_2^1) \tag{D.46}$$

$$\pi_1 + \pi_3 + \pi_5 + \frac{1}{3}\pi_6 + \frac{2}{3}\pi_7 \leq 0 \qquad (p_3^1) \tag{D.47}$$

$$\pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7 \leq 0 \tag{D.48}$$