Enhanced Power System Operational Performance

with Anticipatory Control under Increased Penetration of Wind Energy

by

David Ganger

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

Approved July 2016 by the
Graduate Supervisory Committee:

Vijay Vittal, Co-Chair
Junshan Zhang, Co-Chair
Kory Hedman
John Undrill

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

As the world embraces a sustainable energy future, alternative energy resources, such as wind power, are increasingly being seen as an integral part of the future electric energy grid. Ultimately, integrating such a dynamic and variable mix of generation requires a better understanding of renewable generation output, in addition to power grid systems that improve power system operational performance in the presence of anticipated events such as wind power ramps. Because of the stochastic, uncontrollable nature of renewable resources, a thorough and accurate characterization of wind activity is necessary to maintain grid stability and reliability. Wind power ramps from an existing wind farm are studied to characterize persistence forecasting errors using extreme value analysis techniques. In addition, a novel metric that quantifies the amount of non-stationarity in time series wind power data was proposed and used in a real-time algorithm to provide a rigorous method that adaptively determines training data for forecasts. Lastly, large swings in generation or load can cause system frequency and tie-line flows to deviate from nominal, so an anticipatory MPC-based secondary control scheme was designed and integrated into an automatic generation control loop to improve the ability of an interconnection to respond to anticipated large events and fluctuations in the power system.

ACKNOWLEDGMENTS

I express my profound gratitude to Dr. Vijay Vittal and Dr. Junshan Zhang, Professors in The School of Electrical, Computer and Energy Engineering, Arizona State University for their time, guidance, constant support in this project, encouraging me to consistently further my research. I also greatly appreciate Dr. Kory Hedman and Dr. John Undrill's presence on my committee, complete with their energy, advice, and guidance through these projects.

TABLE OF CONTENTS

Page

APPENDIX Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

GLOSSARY

List of Symbols

α_{ev}	Tail parameter of distribution
κ	Increasing scaling function
m_n	Number of samples in POT analysis
μ	Location parameter of distribution
n	Number of total samples in time series
R	Number of bootstraps per run in determination of x_{min}
R^2	R squared, coefficient of determination
s_n	Sequence of thresholds for empirical data in POT analysis
σ	Scale parameter of distribution
u_k	*k*-th order log empirical moment
w_k	*k*-moment ratio estimator
WR	Set of wind ramp events
$X_{(i)}$	Descending order statistics
x_{min}	Threshold for POT analysis
AR(p)	Autoregressive model with p lags
ξ	Shape parameter in extreme distributions, defined as 1/ α_{ev}
z	Difference statistic estimate for MSE($\widehat{1/\alpha_{ev}}$)
$k(x_i, x)$	Kernel function
$H(\omega, t)$	Hilbert amplitude spectrum value at frequency ω and time t
DS(ω)	Degree of stationarity for some frequency ω

List of Symbols (continued)

$n(\omega)$	Average amplitude within the designated frequency bin centered at ω
n_{sum}	Sum of the average amplitudes across the spectrogram
EDNS_T	EDNS for some length of data T
ε	Epsilon threshold of EDNS for dynamic window algorithm
γ	Gaussian radial basis function kernel parameter
$\mathcal{F}$	Feature space for a nonlinear mapping in SVR
Φ	Nonlinear mapping from
L	Lagrange function
ξ_i, ξ_i^*	Penalties in the loss function of SVR
C	Cost parameter in SVR
X_{st}	Stationary signal in EDNS example
X_{rw}	Random walk signal in EDNS example
c_n	nth IMF
$h_{m,n}$	mth proto-IMF for the nth IMF
$O(\cdot)$	Worst-case complexity
T	Training window length
T^*	Optimal training window for a statistical forecast
T^*_{main}, T^*_{aux}	Optimal training window for the main and auxiliary wind farm
B	Frequency bias factor in units of MW/0.1 Hz
f_{sys}	Power system frequency
Δf_{sys}	Deviation from the nominal system frequency

List of Symbols (continued)

f_{set}	Frequency set point adjustment
$\dot{f}_{set}$	Change in frequency set point
f_{set}^{a}	Frequency set-point adjustment in external area a
f_{ref}	Area frequency reference
P_{tie_ref}	Net scheduled interchange from an area
P_{tie}	Tie-line flow
ΔP_{tie}	Difference between net actual interchange and net scheduled interchange
P_{tie}^{m}	Measured net tie-line flow from an area
P_{turb}^{i}	Turbine power output at bus i
N	Number of states in the state-space representation of a system
M	Number of inputs in the state-space representation of a system
A	N x N matrix describing a system in state-space representation
B	N x M matrix describing the effect of an input in a system
X_i	N x 1 vector of states at time i
U_i	M x 1 vector of manipulated variables at time i
$\dot{U}_i$	1st derivative of control actions at time i
d_i	Actual disturbance in power system at time i
$\hat{d}_i$	Forecast of disturbance input at time i
$\hat{d}_i^a$	Forecast of disturbance input at time i in area a
p	Prediction horizon of MPC controller
m	Control horizon of MPC controller

List of Symbols (continued)

K_p^{expi}	Proportional gain of external PI controller
K_i^{expi}	Integral gain of external PI controller
ε_d	Error in forecast of disturbance input
ε_{ptie}	Error in calculated tie-line flow
B_d	Coefficient representing the effects of d_i on the system
T_s	Time step of discrete control method, such as MPC
J	Cost function for MPC
Q_f	Scalar penalty for the frequency deviation in MPC
Q_{ptie}	Scalar penalty for the frequency deviation in MPC
R_f, S_f	Penalty for the frequency set point and change in frequency set point
u_{min}	Lower bound for the manipulated frequency set-point
u_{max}	Upper bound for the manipulated frequency set-point
Q_{kal}	Process noise covariance matrix in the Kalman algorithm
R_{kal}	Measurement noise covariance matrix in the Kalman algorithm
σ_d	Noise parameter for the disturbance forecast error in Q_{kal}
σ_{ptie}	Noise parameter for the tie-line error in Q_{kal}
$Pt_{size,i}$	Turbine power output capacity in system base at the ith bus
ω_{set}^i	Governor speed reference for generation at the ith bus
P_{tie-ij}	Power flow between buses i and j
K	Stiffness coefficient of transmission line
H	Inertia constant of equivalent rotating machine

List of Symbols (continued)

$\Delta\omega_i$	Speed deviation of rotating machine at bus i from nominal
ΔP_L	Change in load, in system base
δ_i	Voltage phase angle at bus i
q_f	Weight for monitored bus
K_p^{expi}	Proportional gain of external PI controller
K_i^{expi}	Integral gain of external PI controller
Gen_{res}	Fraction of responsive generation in the area
D	Load damping value in system
R	Governor droop for responsive generation in system
w_d	White noise input to drive disturbance estimate in MPC model
w_{ptie}	White noise input to drive tie-line flow estimate in MPC model

List of Acronyms

ACE	Area control error
ACOPF	AC optimal power flow
AEMO	Australian Energy Market Operator
AIC	Akaike information criterion
AGC	Automatic generation control
AMS	Annual maxima series
AMSE	Asymptotic mean squared error
ANOVA	Analysis of variance

List of Acronyms (continued)

AR	Autoregressive
ARIMA	Autoregressive integrated moving average
BA	Balancing authority
BIC	Bayesian information criterion
BPA	Bonneville Power Administration
CDF	Cumulative distribution function
DOE	Department of Energy
DFIG	Doubly-fed induction generator
DS	Degree of stationarity
eACE	Effective area control error
EDNS	Ensemble degree of non-stationarity
EFT	Error of frequency over time
EMD	Empirical mode decomposition
ERCOT	Electric Reliability Council of Texas
EVA	Extreme value analysis
FACTS	Flexible alternating current transmission system
GEV	Generalized extreme value
GPD	Generalized Pareto distribution
HHT	Hilbert-Huang transform
i.i.d.	Independent and identically distributed
IMF	Intrinsic mode function
ISO	Independent system operator

List of Acronyms (continued)

KKT	Karush-Kuhn-Tucker
KPSS	Kwiatkowski, Phillips, Schmidt, and Shin test
MAE	Mean absolute error
MPC	Model predictive control
MSE	Mean squared error
NERC	North American Electric Reliability Corporation
NREL	National Renewable Energy Laboratory
NWP	Numerical weather prediction
OPF	Optimal power flow
PDF	Probability distribution function
POT	Peaks over threshold
PSR	Priestly-Subba Rao test
PSLF	Positive sequence load flow
RBF	Radial basis function
RMSE	Root mean square error
RTO	Regional transmission organization
SVM	Support vector machine
SVR	Support vector regression
VaR	Value at risk
VAr	Volt-ampere reactive
VAR	Vector autoregression
WECC	Western Electricity Coordinating Council

List of Acronyms (continued)

WPP	Wind power prediction
WSS	Wide-sense stationarity

1. INTRODUCTION

Renewable energy resources, such as wind power, are a substantial component of the future generation portfolio. According to the U.S. Department of Energy (DOE), global generation capacity of renewable energy has drastically increased during the past decade, and the U.S. consumption of renewable fuel is projected to continue its increase by 1.6 percent per year until the year 2040 [1]. Specifically regarding wind power, a study by the DOE and Sandia National Laboratories shows that 20% wind energy penetration is a possibility by 2030 [2, 3]. Ultimately, integrating such a dynamic and variable mix of generation requires the development of tools and practices capable of coordinating the many renewable resources, such as wind and solar energy, to be added to the electric grid. With the increase in renewable energy penetration comes an increasingly complex power system, with a need to address stability and reliability through improved control. This study aims to contribute to stability of the smart grid and future power system by contributing to three interconnected focus areas: 1) characterizing the extreme deviations in wind farm power output, 2) quantifying non-stationarity in training data used in statistical forecasting, and 3) integrating prior knowledge of events into the control loop by the use of anticipatory secondary control.

Extreme wind power ramp characterization is helpful for wind power modeling, which is used in power systems planning and research. A method for quantifying non-stationarity for wind power forecasts would ultimately provide improved short-term forecasts, used in power system operations. Both improved modeling and improved short-term forecasts can help create a better understanding of wind power output over time, which is used in anticipatory secondary control.

This dissertation groups each main work into its own chapter. Chapter 1 outlines the structure of the dissertation and details the motivation for the studies as well as some prior work in literature. Chapter 2 presents the characterization of the extreme wind ramp events, and Chapter 3 presents a novel method for quantifying non-stationarity within potential sets of training data for short-term wind power forecasting. Chapter 4 presents work on anticipatory secondary control within the power system, with a given forecast of a disturbance. Chapter 5 concludes by detailing the impact of this work along with providing possible future research directions for this work.

1.1. Motivation

This section provides a brief motivation for each section of the work. Individual chapters detail the motivation for individual topics.

Rare events in the power system may be extremely difficult to predict, but they have extreme economic and social impact. Examples include the Northeast U.S. blackout of 2003, where estimates for cost of lost service range from $7-10 billion for the approximately 16 hours of lost service, or the Bellingham pipeline rupture in 1999, resulting in 3 deaths and a loss of $45 million [4, 5]. In the context of wind power, rare

events include sudden fluctuations called wind ramps that lead to errors in persistence forecasts for wind farm power outputs. Modeling wind power output with an appropriate representation of rare events helps to maintain stability in a power system, as the penetration of renewable energy may introduce many rare but catastrophic events such as sudden power generation drops or extended periods of low power generation.

In addition to characterizing rare events in wind power, improvements in short-term wind power forecasts would lower power system operating costs and increase reliability. As renewable energy generation capacity increases across the globe, the economic importance of improved renewable generation forecasts grows [6]. The U.S. Department of Energy, the National Oceanic and Atmospheric Administration, and many large private industry members propose that the uncertainties in state-of-the-art short-term (0 to 6 hour-ahead) wind forecasts add unnecessary increased costs and risks to the U.S. electrical grid [7, 8]. Improved forecasts can lower power system operating costs by decreasing the necessary amounts of spinning reserve and improve the reliability of the power system [9, 10]. Short-term forecast improvements can also improve the performance or computation speed of power systems operations tools, such as through scenario reduction of wind farm power outputs in stochastic power system operations studies [11]. The variability of renewable energy is directly linked to the variance in the forecast, and smaller forecasting errors would allow a balancing authority (BA) to carry fewer reserves, lowering the overall cost of energy to the power system. One aspect in short-term forecasting that is frequently overlooked is the importance of the stationarity in the training data for short-term wind power forecasts. Quantifying the amount of non-stationarity in the training data, ultimately leading to the determination of a suitable training window, is presented in this work.

Renewable energy is inherently stochastic, as renewable energy sources depend almost solely on weather-related events such as wind speeds. Thus, replacing conventional, dispatchable generation with renewable resources may require careful planning, including the design of new control systems that may not have existed before. Improvements in the secondary control of a power system may be desirable, such that rare generation or transmission events can be handled more effectively. The proposed improvement centers on the use of an external anticipatory controller, augmented onto the existing AGC framework through a modification of the area control error (ACE) signal. This anticipatory controller would use an event forecast and a linearized model of the power system to proactively minimize the impact of sudden, predictable events.

1.2. Literature Review

This section provides a literature review of the works presented in this dissertation. Prior work surrounding the three subjects of extreme wind ramp characterization, determining non-stationarity in wind power output time series, and anticipatory secondary control will be explored.

1.2.1. Extreme Wind Ramp Events

Accounting for wind ramp events covers a broad set of topics, including defining a wind ramp, linking wind speeds to wind power outputs, and creating techniques to accurately forecast wind in the short-term or long-term. A general overview of these topics will be presented in the following sections.

(i) Wind Ramp Definitions

Defining a wind ramp is not trivial, as it is not always clear what timescale and magnitude impact the definition should encompass. Realizing the end time of the wind ramp is also important, especially in short-term forecasting that takes place in resolutions of less than an hour. To illustrate the difficulty of defining a ramp, Figure 1-1 shows a generic ramp with the main variables of interest for a wind ramp. The start and end point of the wind ramp is currently point B and point D, respectively, based on the marked duration, ramp rate, and power swing. However, point A and point C are another pair of viable start and end points with different duration, ramp rate, and power swing. Choosing these exact points is subjective, usually only specific to the context in which this wind ramp definition would be used.

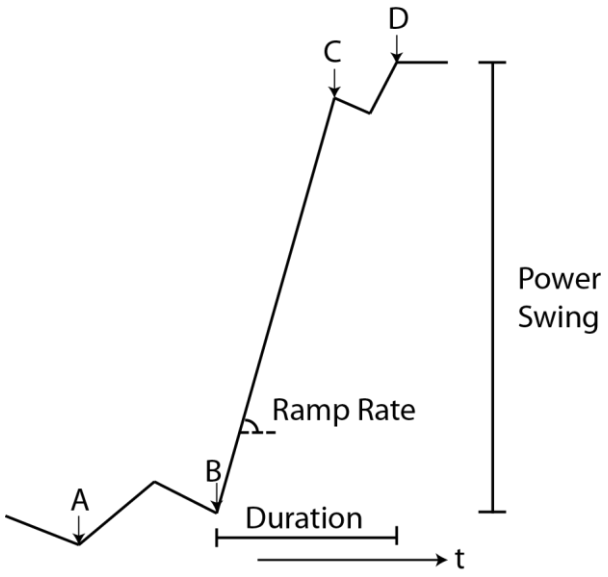


Figure 1-1 Generic Wind Power or Speed Curve

Recent work in this area includes [12], which developed rules and algorithms to detect wind ramps in a set of data, offline. The work uses dynamic programming to find ramp events that meet certain criteria such as whether a ramp is above a certain threshold,

and provides a framework to add more rules to for wind ramp definition and detection. Also, Argonne National Laboratory has performed a fairly thorough review of ramp definitions, compiling information on varying methods to both define and use information on wind ramps [13]. Some other work defines ramps by a very static, specific set of rules, such as [14] which defines a ramp as a change in wind farm power output that is greater than 50% of the capacity within four hours, or [15], defining it as a 20% change in less than an hour.

(ii) Present Efforts to Quantify Wind Variance in Speed and Power

Wind energy output is highly correlated with the wind speeds around the wind turbines of interest, due to the fact that wind turbines are powered by the kinetic energy force of wind. A power curve is shown in Figure 1-2, mapping turbine output to wind speed through the blades of a variable pitch GE turbine [16]. Other turbines with variable blade pitch have power curves with the same shape, while turbines with static blades have an optimal wind speed at which rated output is reached. In addition, it is important to note that all commercial turbines have wind speed cut-off points, where the turbine is shut off and power output is quickly brought to zero to prevent high wind speeds from damaging the turbine. The existence of these power curves would imply that the vast amount of work in meteorology and weather prediction could be directly applied to this wind power prediction problem, allowing for minimal variation between forecasted and actual wind outputs in a farm.

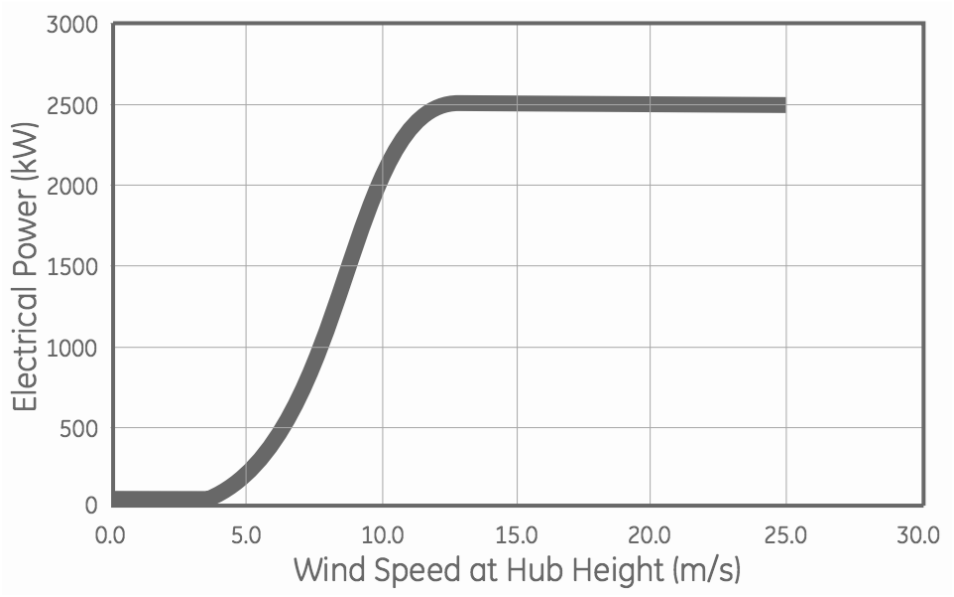


Figure 1-2 Power vs. Speed Curve for GE Turbine 2.5-103 [16]

However, due to the great variations that wind speeds can have even over the distance of a few hundred meters, a resolution much greater than that of existing weather forecasts would have to exist to accurately predict the output of a particular wind turbine. It is not economical and often times impossible to measure wind speeds at each turbine, so mapping the many wind forecasts to exact power output at a given wind farm is not trivial. There is much work focused on using wind speeds to predict behavior of power systems through spatial or temporal correlation [17, 18, 19].

Because of the stochastic nature of wind, forecasting is an invaluable tool in wind power scheduling. The power system operator must know the expected power output of the wind farm to be able to dispatch the conventional generation appropriately, and smaller deviations in predicted and actual wind power outputs is desired. Numerical weather prediction (NWP) algorithms exist, but because power system economic dispatch operates on a much smaller timescale, NWP algorithm results are generally not as useful for short-

term wind power forecasts. In small geographical regions, such as wind farms, persistent wind forecasts are most commonly utilized for very short-term (less than 30 minutes ahead) forecasts. This method simply takes the existing wind speed or power output statistic and extrapolates the same value into the next time slot, assuming that no changes will occur between the time slots. Although this is an elementary approach to prediction, the simplicity and relative accuracy of persistence predictions on small timescales have made this a popular choice for short-term forecasting and benchmark tests for experimental forecasts [20].

For day-ahead forecasts used in unit commitment, there are numerous wind speed and power forecasting services that are available to industry, including 3TIER, Aeolis, and AWS. For long-term planning purposes, the forecast offerings are fairly sparse, though 3TIER mentions the tools they have for this as well.

The statistics of the errors on wind predictions on the economic dispatch timescale are important to characterize for both economic reasons and for power system reliability. A very low variance in errors, for example, may signal that only a small amount of reserve needs to be allocated for compensation of the generation variation, while considerable amounts of large deviations may suggest that more fast-ramping reserves should be kept online to maintain power system security. Load curtailment scenarios must be a last-resort action to preserve reliability, as this is an expensive option for the power system [21, 22]. Wind ramps and wind errors on the short timescale has been assumed to be Gaussian in distribution simply because of the small amounts of wind generation power systems. However, many characteristics of wind do not have Gaussian properties, and persistence forecasting methods show non-Gaussian error statistics [21]. It is important to determine

the distribution of errors for purposes of reserve planning as well as wind forecasting efforts. The distribution of errors may also determine how often the generation/load balance is upset, indicating the frequency of a large secondary control action to correct this generation/load imbalance.

(iii) Extreme Events and Applications in Power Systems

Extreme value analysis (EVA) has been used in the engineering field in applications ranging from power estimation in VLSI circuits to statistically characterizing sonar reverberation [22, 23]. However, little work has been published on characterizing wind ramp events using EVA, but there has been related work on the power system in general referencing extreme values.

Many studies detail the impacts that extreme events have on power systems, offering different ways to tackle or circumvent such problems, such as an adaptive learning method or load shedding algorithm [24, 25]. In addition, EVA has been studied in relation to the power bidding market, using price spikes as extreme events [26]. Another group established a tool that would allow for the prediction of an extreme event given a wind power prediction (WPP) model by simply applying ramp detection algorithms [27].

It is important to include any work in wind power forecasting as well, as it has the same end goal of reducing uncertainty with stochastic generation. The authors in [28] perform a basic review of the forecasting studies used globally, including numerical weather prediction (NWP) models, statistical methods, artificial networks, and hybrid forecasting models, and discusses the differences between wind power forecasting and wind speed forecasting. Learning schemes with fuzzy logic or neural network approaches are used, with demonstrated cases of performing better than persistence forecasting [29,

30]. Use of spatio-temporal analysis also shows promising results for short-term wind forecasting [31]. Characterization of extreme wind ramps would help construct or validate these other forecasting methods.

1.2.2. Optimal Training Windows for Wind Power Forecasts

Though many forecasting works contain some notion of finding stationary time periods in the data set, very few works explicitly calculate or quantitatively explore stationarity. Typically, non-stationary features such as trends are considered, where the time series is differenced or filtered such that the time series has properties that signal a stationary signal [32]. There are also methods that are used to test for stationarity, but they are not meant for use in finding training data sets for time series forecasting. Lastly, the notion of stationarity within the data set may not be mentioned, as the performance of the model is used to justify the fact that the assumptions underlying the data (that the time series is stationary) is met. However, the results of the tests involve some subjectivity and may in-fact conflict with each other, as mentioned in Chapter 3. In one wind power forecasting work, an approximately stationary epoch is determined by combining periods of wind generation with similar PDFs over different days in a month [31]. This provided improved forecasting results, but it was a heuristic method which may not apply to all types of wind farm forecasts.

1.2.3. Controller Design for AGC

The automatic generation control (AGC) function for the power system involves both economic dispatch of the committed generation units (by simply using the participation factors assigned to each set of responsive generation) and load frequency

control [33]. With knowledge of the wind ramp error, a controller could be designed to assist the power system by ramping generators before extreme events to prevent overloading transmission lines or shedding load. An assessment of AGC was performed by N. Jaleeli et al., describing the regular operation of automatic generation control as well as its inherent limits [34]. A more recent assessment has been done in 2005 in [35], detailing the modern advancements in the AGC and area control error (ACE) as well as the integration of flexible alternating current transmission system (FACTS) devices and renewable technologies.

Specific to wind-related applications, [36] introduces another automatic generation control system for individual wind farms, separate from a system-wide AGC. A study was also done on the effects of wind ramps on power system operation, with the conclusion that ramping capabilities of the AGC and generation are the limiting factor [37]. It was also found in [38] with a simple model that there is a limit in wind penetration in a modern power system with AGC to maintain a frequency within limits, namely that a power fluctuation of 5% of the total thermal plant capacity may be tolerated without exceeding 1% frequency deviation.

One notable study that performed the integration of wind power forecasting error into AGC was from [39], where any deviations from predicted or scheduled wind outputs is supplemented by controllable generation. Specifically, hydro generation was assumed to have enough ramping capability to account for any errors in the wind output, and any deviations from the estimated wind output was added to the ACE value, which would be integrated over a set timeframe to establish any changes in generation set points that would

be needed. An overview of some downfalls of present AGC setups with high penetrations of wind power is discussed in [40], focusing on economic optimization.

Recent research in improvements to AGC include: 1) creating accurate, non-linear power system models with deadband effects [41, 42, 43], 2) applying newer, nonlinear control techniques such as genetic algorithms or neural networks [44, 45, 46], and 3) introducing forms of storage into the AGC framework [47, 48, 49]. Optimal control theory has also been applied to AGC, where the frequency control is shown to be tighter for certain cost matrices in the optimal controller's optimization function [50, 51]. Notably, works involving AGC in the context of renewables integration include dynamic load dispatch, in which the participation factor of each generator changes dynamically, and the application of a fuzzy logic PI controller [52, 53].

The application of MPC in secondary control has been seen in other works as well, though it is implemented through a replacement of the conventional AGC system rather than as an add-on module, and does not take into account any anticipated disturbance. The most comprehensive work involves the use of MPC for load frequency control in a simplified model of the Nordic power system, where limitations on tie-line power flow, generation capacity, and generation rate of change were taken into account [54]. The work used a centralized MPC controller as opposed to distributed MPC, which was explored through an example power system simulation in another work [55]. Simulations using MPC for frequency control in an isolated wind-aluminum power system also found that MPC has benefits when using it to control load in addition to generation in a small power system [56].

This work plans to build off of the prior work presented in this section. Original work on wind ramp characterization, determining non-stationarity in wind power time series, and anticipatory secondary control will be discussed in the next few chapters.

2. CHARACTERIZATION OF EXTREME WIND RAMP EVENTS

2.1. Background and Motivation

Rare events in the power system such as blackouts may be extremely difficult to predict, but they have extreme economic and social impact. Examples include the Northeast U.S. blackout of 2003, where estimates for cost of lost service range from $7-10 billion for the approximately 16 hours of lost service, or the Bellingham pipeline rupture in 1999, resulting in 3 deaths and a loss of $45 million [4, 5]. The scientific community has attributed these rare events to physical issues, where a downed power line or sudden generation loss resulted in cascading blackouts, and software issues, where simple network elements failed to respond. In the context of wind power, rare events include sudden fluctuations called wind ramps that lead to errors in persistence forecasts for wind farm power outputs. In a system with a large wind penetration, such a collective wind ramp can drastically change the operating points for generators in the system, possibly overloading certain system elements.

So far, existing wind power forecasting models aim for the best average forecast error, that is, close to zero error when averaged over a large number of forecasts. Small numbers of extreme errors hardly affect averaged errors, but can be costly or even catastrophic to the power system. Extreme power system errors are not necessarily extreme meteorological events, so even correctly modeling all extreme variations in the weather

may not account for the errors in wind power output. Maintaining a model of rare events is an integral part of maintaining stability in a complex power system, as the penetration of renewable energy may prove to introduce many rare but catastrophic events such as sudden power generation drops or extended periods of low power generation.

2.1.1. Potential Operational Problems with Wind Energy

A critical concern about renewable energy is the inability to precisely control its output, as balancing the load and generation is necessary to ensure power quality and reliability to the load. In the past, power systems were planned with the assumption of a controllable, unidirectional power flow from the generation sources to the loads [57]. Conventional generation is almost completely controllable, with known constraints on the ramp rate limits and operating points of the generation. Load patterns introduce some randomness, but is highly cyclical and follows general trends. For example, demand is generally higher in the daytime and the early evening in peak load hours, and are generally lowest in the middle of the night during off-peak [58]. This leads to controllable generation following the load in both unit commitment and economic dispatch, which operate on different timeframes.

Unit commitment solves the optimization problem of which generation units to turn on during some period of time, typically decided at least one day in advance. If too few units are committed to generating power given a level of load, the power system operator may have to purchase excessive reserves. In extreme cases, the power system may have to shed load, as the system would not be able to provide reliable power to all loads. If too many units are committed, the system would still have to pay for this unused power because of no-load costs incurred by the generation. Economic dispatch is defined by the U.S.

Energy Policy Act of 2005 in Section 1234 as the "operation of generation facilities to produce energy at the lowest cost to reliably serve consumers, recognizing any operational limits of generation and transmission facilities" [59]. In modern energy systems, economic dispatch is performed in real-time, compensating for any mismatches between the scheduled and real generation requirements. In the present market, this can mean anywhere from a 5-minute to a 15-minute period for its iterative optimization process.

In the future, power flow in the grid may not be unidirectional, but in diverse, time-varying directions with varying magnitudes. Power flows in all directions as a consequence of the integration of renewable energy and storage [60]. In addition, the future grid incorporates a large suite of sensors, creating a communication network layer overlay on the already complex power system.

This complex grid of power and communication introduces many problems over the traditional power grid, one of which is the ability to perform real-time load balancing. Economic dispatch depends on the ability of the generation to be flexible enough to meet changes in demand over a small timeframe, and existing techniques for dispatch depend on both spinning and non-spinning reserve in conventional generation. Modern day dispatch assumes scheduled generation to be completely controllable, and any generation scheduled in a day-ahead market would be the baseline from which to establish any load imbalance to be met.

However, with the introduction of stochastic, uncontrollable generation such as wind and solar, the generation base becomes increasingly stochastic. Given two independent random variables X and Y, the variance of the sum of the variables is simply the sum of the variances. As applied to the load balancing problem, adding the variation in

load and the variation in wind will give a larger variance, which necessitates more reserves to be available to the system operator.

So far, existing wind power forecasting models aim for the best aggregate forecast error, that is, close to zero error when averaged over a large number of forecasts. Small numbers of extreme errors barely affect averaged errors, but can be very costly or catastrophic to the power system. Worse yet, extreme power system errors are not necessarily extreme meteorological events, so even correctly modeling all extreme variations in the weather may not account for the errors in wind power output. With increasing penetrations of wind energy in the power system, power system planning and economic dispatch would benefit with a characterization of the extreme variations in wind power outputs.

2.1.2. Wind Ramps

Wind power ramps are defined as quick and large changes in wind power output [13]. With persistence forecasting methods for wind power output, wind ramps constitute the error between the forecasted and real power output. Wind ramps have been defined using many different sets of rules, but in this study, a wind ramp is only defined as an increase or decrease of wind power during a set duration in time, *dt*. There is no minimum threshold of power output ramping in this case.

A mathematical definition for a wind ramp down as used in this work follows

$$\{\big(X(t) - X(t - dt)\big) \in WR \mid X(t) - X(t - dt) < 0\} \qquad (2\text{-}1)$$

where *X* is the time series data and *WR* is the set of wind ramps. Likewise, a wind ramp up would involve the difference of the time series data being positive.

2.2. Extreme Value Analysis

Extreme value analysis (EVA) shifts focus from what classical statistics characterizes as the average behavior of a stochastic process. The central limit theorem, for example, motivates much of classical statistics, with outliers generally ignored in the creation of the model for a process. However, low-probability, high-impact events can result in disasters, breaking down a financial system or otherwise dramatically affecting human society. EVA has been developed over the years to understand and characterize these events.

2.2.1. Historical Background on Extreme Value Analysis

Extreme value analysis (EVA) has seen an explosion of development recently due to the financial market, but the first recorded work on extreme values in a statistical context was made as early as the 19th century. In the 1920s, L. H. C. Tippett noticed a pattern in the yarn breakage rates used in weaving, linking the strength of a thread to the strength of its weakest fiber. Both R. A. Fisher and Tippett published a paper on the extremes of a distribution, with the patterns seen in data [61]. In this paper, the distribution of the largest member of a sample from a normal population was fit to a curve, utilizing a shape parameter α_{ev}. Soon after in 1943, B. Gnedenko published a paper detailing the generalization and unification of the theory dealing with maxima [62]. E. J. Gumbel then published a book, including details on the statistics of the extremes [63]. Further research was performed in extremes, and Pickands, Balkema, and de Haan found ways to characterize the asymptotic tail distribution of a random variable [64, 65]. This work spawned two separate approaches to characterizing the extremes of a distribution, based

on what is now known as the Fisher-Tippett-Gnedenko theorem and the Pickands-Balkema-de Haan theorem.

Uses of EVA have historically revolved around the insurance and financial markets, as well as weather-related phenomena [66]. For example, Value at Risk (VaR) is a common risk measure used by financial risk management experts to determine the probability of an extreme loss [67]. Meteorology has used this tool primarily in natural disaster assessments, such as the 100-year flood measure that predicts a flood event that has a 1% probability of occurring in any given year [68].

2.2.2. Extreme Value Theory

The main result of the Fisher-Tippett-Gnedenko theorem (also referred to as the extreme value theorem) is most easily described as an analogue of the well-known central limit theorem. That is, if a random variable X_i is independent and identically distributed (i.i.d.), then

$$\bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \qquad (2\text{-}2)$$

The central limit theorem in (2-2) addresses the results of summing a series of i.i.d. random variables, but does not address maxima or minima. An analogue specific to extreme value theory must try to relate $max\ (X_1, \ldots, X_n)$ to a distribution. First, note that X_i for $i = 1,2, \ldots, n$ must be i.i.d. random variables.

Define $M_n \coloneqq \max(X_1, \ldots, X_n)$, a set of maxima for many random variables. It is of interest to prove that M_n or some transformation of M_n gives a stable distribution. This means that any linear combinations of two independent random variables M_n will give the same distribution. Extreme value theory has shown that if there are sequences

$\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty}$ with $\frac{(M_n - b_n)}{a_n} \xrightarrow{d} Z$, then Z has a maximum stable distribution. The minima of a set of i.i.d. random variables can also be represented by such a distribution for Z, shown as a simple extension of the proof for the maximum.

In the realm of extreme value theory are three types of distributions that Z can converge to, labeled as Type I, Type II, and Type III, all with different properties and, importantly, different tails. These compose the generalized extreme value (GEV) distribution family:

Type I (Gumbel Distribution)

This is the most basic distribution of the extreme value distribution family, with two parameters, μ as the location parameter and β as the scale parameter (β must be positive). The cumulative distribution function (CDF) of the Gumbel distribution is:

$$F(x) = e^{-e^{-(x-\mu)/\beta}} \qquad (2\text{-}3)$$

This distribution is considered to have a light upper tail, and is positively skewed. The tail is in reference to values that are far away from the median of the distribution. The Gumbel distribution's light tail indicates that the extreme or tail values display a lower probability of occurrence than in the case of the normal distribution, or that the extreme portion of the distribution spreads out less than that of the normal distribution. Figure 2-1 shows the probability density function (PDF) of the distribution over a range of parameters.

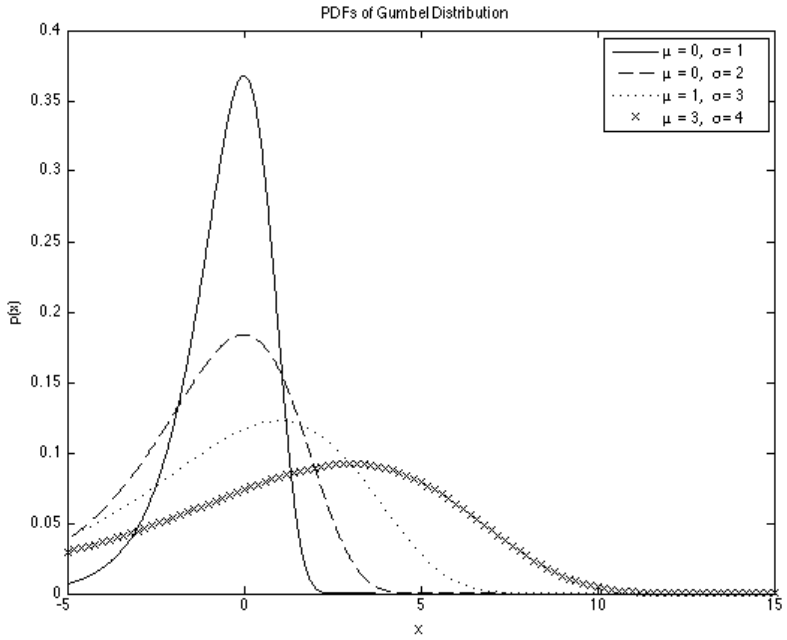


Figure 2-1 Type I Gumbel Distribution PDFs with Various Parameters

Type II (Fréchet Distribution)

The Fréchet distribution has three parameters, μ as the location parameter, β as the scale parameter, and α_{ev}, the shape parameter. Note that by convention, $\xi = 1/\alpha_{ev}$, where ξ is also referred to as the shape parameter. The CDF of the Fréchet distribution is:

$$F(x) = e^{-\left(\frac{x-\mu}{\beta}\right)^{-\frac{1}{\xi}}} \qquad (2\text{-}4)$$

Any value $x \leq \mu$ gives $F(x) = 0$. This distribution is considered to have a heavy or 'fat' upper tail, and also has a minimum value specified by the location parameter. The Fréchet distribution's light tail indicates that the extreme or tail values display a high probability of occurrence than in the case of the normal distribution, or that the extreme portion of the distribution spreads out more than that of the normal distribution. Figure 2-2 shows the PDF of the distribution over a range of parameters.

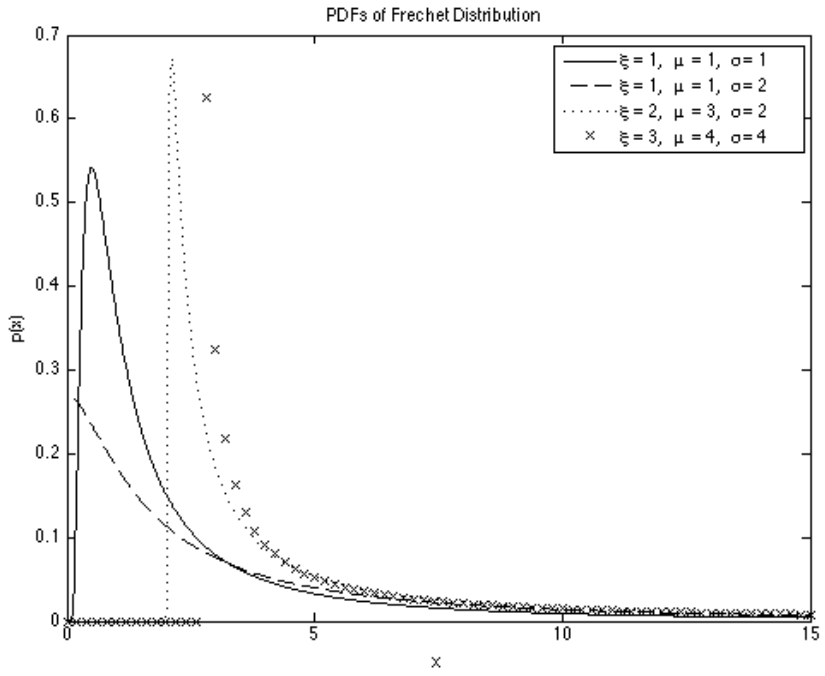


Figure 2-2 Type II Fréchet Distribution PDFs with Various Parameters

Type III (Reversed Weibull Distribution)

The reversed Weibull distribution is similar to the standard Weibull distribution used in classical statistics, but has one extra parameter and is reversed such that the distribution has an absolute maximum (no tail). This distribution has three parameters, μ as the location parameter, β as the scale parameter, and α_{ev}, the shape parameter. Note that similar to the Fréchet distribution, $\xi = -1/\alpha_{ev}$, where ξ is also referred to as the shape parameter. The CDF of the reversed Weibull distribution is:

$$F(x) = e^{-\left(\frac{-(x-\mu)}{\beta}\right)^{-\frac{1}{\xi}}} \quad (2\text{-}5)$$

Any value $x \geq \mu$ gives $F(x) = 1$, and this distribution has no upper tail though some refer to its tail as light or thin. This distribution is not common in applications of extreme value theory because of this inherent limitation of a set maximum. For example,

in many applications it is difficult to verify that the distribution has a meaningful and predictable maximum value. Figure 2-3 shows the PDF of the distribution over a range of parameters.

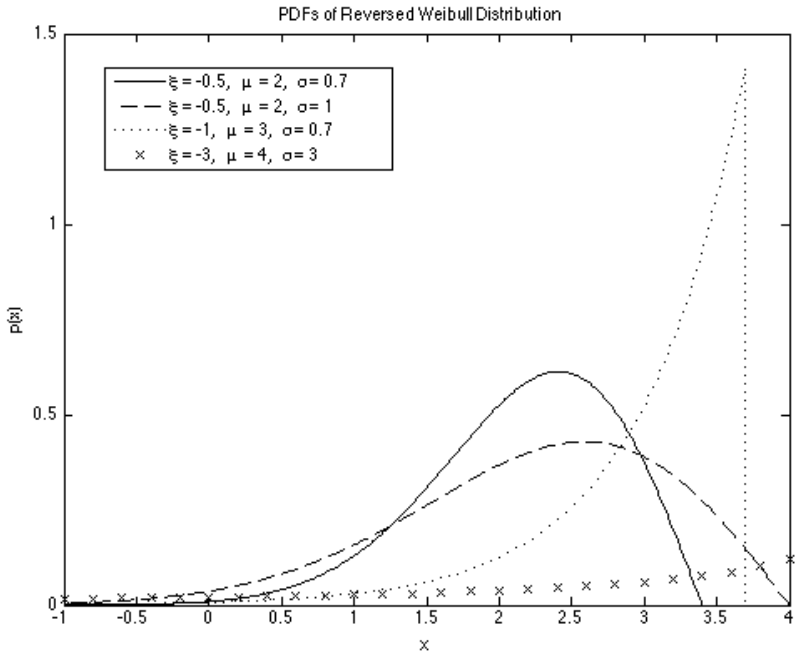


Figure 2-3 Type III Reversed Weibull PDFs with Various Parameters

These three types of distributions all fall under the generalized extreme value distribution family, and apply to the two different theorems in EVA.

First Theorem of Extreme Value Theory

Also known as the Fisher-Tippett-Gnedenko theorem, this theory gives a general result of the asymptotic distribution of extreme order statistics.

If $M_n \coloneqq \max(X_1, \dots, X_n)$,

and $\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty}$ with $\frac{(M_n - b_n)}{a_n} \xrightarrow{d} Z$

Then the limit distribution of Z belongs to the GEV family.

Second Theorem of Extreme Value Theory

Also known as the Pickands-Balkema-de Haan theorem, this theory gives an asymptotic tail distribution of a random variable X when the true underlying distribution of X is unknown. In other words, we're interested in estimating the tail distribution of a random variable X, defined as the conditional excess distribution function:

$$F_u(y) = P(X - x_{min} \leq y | X > x_{min}) = \frac{F(x_{min}+y)-F(x_{min})}{1-F(x_{min})} \qquad (2\text{-}6)$$

In (2-6), x_{min} is known as the threshold, and y is the new random variable of interest. The range of this distribution is $0 \leq y \leq x_F - x_{min}$, where x_F is the final (maximum or minimum) endpoint of the underlying distribution. To determine $F_u(y)$, the theorem shows that $F_u(y)$ is well approximated by the generalized Pareto distribution (GPD), with very similar structure to that of the GEV family (Type I, Type II, and Type III). The same location, scale, and shape parameters apply to the GPD family, and in fact, the value of the shape factor ξ is the same when applying either theorem to the same set of empirical data.

2.2.3. Data Analysis Methods using EVA

The origin of EVA stemmed from applications within manufacturing and optimization, and continues to thrive through its use in financial risk analysis. It is therefore important to explore the ways empirical data can be characterized with the theories in EVA. Two main methods exist for analyzing empirical data with EVA, the annual maxima series (AMS) method and the peaks over threshold (POT) method. While AMS analysis has been in place for longer than the more modern POT analysis, both approaches have distinct advantages.

Annual Maxima Series (AMS)

The implementation of the AMS method stems directly from the conclusions in the first theorem of extreme value theory. The first theorem of extreme value theory states that a set of maxima taken from a set of random variables will conform to one of three GEV distributions. To actually fit a set of data to a distribution, it is necessary to define a 'block' representing a random variable. This block must be assumed independent of all other blocks, such that they may represent a series of random variables. The maximum of each block will represent the values $M_n \coloneqq \max(X_1, \dots, X_n)$, and standard maximum likelihood estimation or generalized method of moments can be applied to estimate the parameters in this statistical model.

Peaks Over Threshold (POT)

The POT method relies on the second theorem of EVA, namely that the asymptotic tail distribution of any given distribution will fall under the GPD family. That is, defining $F_u(y) = P(X - x_{min} \leq y | X > x_{min})$, the distribution of $F_u(y)$ falls under the GPD family. This method requires x_{min} to be chosen, though there is presently no analytical method to determine this value. Estimation of GPD parameters in the POT method utilizes maximum likelihood estimators or the generalized method of moments, similar to the AMS method.

Uses and Limitations of AMS vs. POT

Both methods are used in modern characterization of extreme events, as both have their strengths. The AMS method is by far the oldest method, and also the most

straightforward to implement. The POT method is fairly new and less straightforward to implement, but has shown to create higher resolution models.

Models using AMS only need to divide the data into blocks of similar size, typically throughout time, to determine block maxima or minima. Choosing the block size in AMS can be important, as there is a tradeoff between bias and variance [69]. If the blocks are too small, then the approximation of the distribution is poor, leading to bias in estimation. On the other hand, very large blocks generate fewer data for analysis, leading to a large estimation variance. However, standard practice in literature sets the block size to something easily conceptualized, such as taking hourly, daily, or yearly data. These choices do not seem to affect the validity of the modeling.

In the POT method, all data is taken above a threshold, usually resulting in many more data points than the AMS method. However, the threshold determination in this method is less straightforward, though like the block size in the AMS method, there is a tradeoff between bias and variance. A low threshold includes too many points, introducing bias because of the inclusion of points outside the tail of the distribution, while a high threshold includes too few observations and results in high variance. It is important to recognize that in the POT method, too low of a threshold will dramatically affect the estimation, as only the asymptotic tail of the distribution is captured by the POT theory. For this reason, it is popular to assume a threshold slightly above what one would normally, intuitively assume, to be conservative.

As both methods have a parameter that balances the bias and variance of the theoretical fit, the model is only as good as the determination of this parameter.

2.2.4. Determination of x_{min} in POT Method

Because of the effects that the determination of x_{min} has on the parameters of the POT method, many different ways to determine x_{min} have been explored [70]. One of the simplest methods involve plotting the data or the frequency of the data on a logarithmic plot and noting when the tail begins, which under many conditions would be a clear linear section of the graph. The choice of x_{min} is then the 'elbow' of this log-log chart, which is when the tail begins. Another popular approach to estimate the threshold is by using the Hill plot, which is simply the Hill estimator of the tail index plotted against varying threshold values [71, 72]. The threshold is then chosen to be the lowest threshold after which the tail index stabilizes. The intuition behind this method is that since the tail parameters depend on the data subset chosen, the parameter should 'converge' to a stable value once the proper tail data has been identified.

It is known that for a given sample size n, there exists a unique sequence of thresholds $s_n = x_{min,n}$ that allow for the bias and variance of a tail parameter to decrease at the same rate as n increases [73]. However, estimation of s_n would require the mean squared error (MSE) of $\widehat{\frac{1}{\alpha_{ev}}}$, an estimation of the MSE of $1/\alpha_{ev}$, which is itself affected by the choice of s_n. A more comprehensive and thorough approach to threshold estimation has been explored by J. Danielsson and C. G. de Vries [71, 72]. They solve this problem by using the idea behind control variates in Monte Carlo estimation, reducing the variance of a variable (in this case, $\widehat{1/\alpha_{ev}}$) by subtracting another carefully constructed parameter with known errors and variance.

An overview of the theory behind this procedure follows. The goal is to find a suitable estimate $s_n = x_{min}$ for the threshold which will allow for parameter estimates with the least bias and variance. Because minimizing the bias and variance are conflicting goals given a set sample size n, it will be sufficient to find a s_n such that, as $n \to \infty$, the bias and variance of $\widehat{1/\alpha_{ev}}$ converge to nothing at the same rate. With the vision of using sub-sample bootstrapping and control variate techniques, it is important to show what the *k*-moment ratio tail index estimators are, and for that the definition of the conditional k^{th} order log empirical moment must be defined. Given a sample $X_1, \dots, X_n$ of n i.i.d. observations from the distribution $F(x)$:

$$u_k(m_n) \equiv \frac{1}{m_n}\sum_{i=1}^{m_n}\left(\log\frac{X_{(i)}}{s_n}\right)^k,\ s_n = X_{m_n+1} \qquad (2\text{-}7)$$

where $X_{(i)}$ are the descending order statistic and m_n is the number of order statistics used (number of statistics above s_n). The *k*-moment ratio estimator is defined as

$$w_k(s_n) \equiv \widehat{\frac{1}{\alpha_{ev}}} = \frac{u_k(s_n)}{k u_{k-1}(s_n)} \qquad (2\text{-}8)$$

where k is a positive integer, the moment number. After some extensions of the proof in [72], the asymptotic mean squared error (AMSE) of $\widehat{1/\alpha_{ev}}$ is

$$\text{AMSE}\big(w_k(s_n)\big) \approx \frac{\kappa(k)s_n^{\alpha}}{a\alpha_{ev}^2 n} + \frac{b^2\beta^2\alpha_{ev}^{2k-4}s_n^{-2\beta}}{(\alpha_{ev}+\beta)^{2k}} \qquad (2\text{-}9)$$

where a, b are scaling constants, α_{ev} and β are the shape and scale factors of the extreme value distribution, and $\kappa(k)$ is an increasing function

$$\kappa(k) \equiv \frac{(2k)!}{(k!)^2} + \frac{(2k-2)!}{\big((k-1)!\big)^2} - \frac{2(2k-1)!}{k!(k-1)!} \qquad (2\text{-}10)$$

There is a unique sequence s_n that asymptotically balances the two terms in (2-9) as $n \to \infty$, which is derived from the first order condition that $\frac{\delta\text{AMSE}}{\delta s_n} = 0$.

The minimization of the subsample bootstrap MSE utilizes control variate techniques, where the variance of the estimator is minimized by adding or subtracting a term for which many of the parameters or errors are known. A difference statistic is used, such that the AMSE has the same convergence rate:

$$z(s_n) \equiv w_2(s_n) - w_1(s_n) \tag{2-11}$$

The minimization of the subsample bootstrap MSE is then

$$\min_{s_{n_1}} \frac{1}{R} \sum_{r=1}^{R} \left[\left(z(s_n) \right)^2 \right] \tag{2-12}$$

where R is the number of bootstrap resamples. As the AMSE$[z]$ has the same order of magnitude as the AMSE$[w_k]$, (2-12) can be used to find the optimal threshold s_n, calculated under two different sample sizes m_{n_1} and m_{n_2}. To arrive at this consistent estimator, calculate:

$$\widehat{m}_n(w_2) = \frac{\left(m^*_{n_1}(z) \right)^2}{m^*_{n_2}(z)} \left[\frac{\sqrt{2} \log m^*_{n_1}(z)}{2 \log n_1 - 2 \log m^*_{n_1}(z)} \right]^{f_u} \tag{2-13}$$

and $m^*_{n_1}$, $m^*_{n_2}$ are optimal number of points in two different bootstraps and where f_u is defined as

$$f_u = \frac{2 \log n_1 - 2 \log m^*_{n_1}(z)}{\log n_1} \tag{2-14}$$

The procedure for finding this threshold would involve first choosing a bootstrap size $n_1 < n$, then drawing R bootstrap resamples of size n_1. Then calculate the bootstrap MSE of the difference statistic z at each m_{n_1} and find the optimal $m^*_{n_1}$ such that the bootstrap is minimized. Repeat the entire procedure for a smaller resample size $n_2 = \frac{(n_1)^2}{n}$, yielding $m^*_{n_2}$. Then calculate $\widehat{m_n}(w_2)$ using (2-13), and estimate the parameter $\xi = \frac{1}{\alpha_{ev}}$ using this threshold. The difficulty in this procedure lies in determining how many values

of n_1 to sweep over, as ideally all values would be examined. It will be shown in the following section that this would take a prohibitively long time given a large enough data set size and a standard computer, and so a grid over which n_1 would be used must be defined.

2.3. **Characterization of Wind Ramps Results**

Characterizing wind ramps initially involved fitting the data with a Gaussian distribution. However, after this was shown to be inadequate, EVA was introduced as a viable method to characterize these wind ramps. After validating that EVA in fact gave more reasonable results than traditional statistical methods, both the AMS and POT methods were used to model the wind ramps. Lastly, a more rigorous method presented in the previous section was used to find x_{min} within the POT method to validate the results.

Evidence of stationarity must exist for any model's statistics to be a reasonable predictor for future wind ramp behavior. It is well known that the non-stationarity of wind speeds presents a challenge to statistical wind speed modeling [74, 75]. However, many non-stationary processes are either trend-stationary or difference-stationary, and simply taking the derivative of the time series data will result in a stationary process [76, 77]. Common models such as the autoregressive integrated moving average (ARIMA) model commonly use the difference process of wind speed or power to forecast, as it has been seen that the difference in wind power output may be stationary [78]. To ensure that the assumption of stationarity is reasonable, resulting extreme value models will be examined for trends.

The National Renewable Energy Laboratory (NREL) provided wind generation data from one 300.5 MW wind farms in Colorado for both 2009 and 2010, with 10-minute resolution. This led to a total of 105120 data points for total wind farm output for statistical analysis of wind power ramps. The data was provided in an Excel worksheet, and was imported into the software MATLAB for analysis. A small amount of preprocessing was necessary, as some data values were incorrectly negative due to calibration errors. All negative power output data was changed to an output of 0 MW.

2.3.1. Traditional Gaussian Characterization of Wind Ramps

The present most popular distribution for wind ramps and forecasting errors is the Gaussian distribution, which stems from both the central limit theorem and the familiar properties that it presents [21].

First, the wind power ramps for a set interval were calculated by taking the differences of power output in MATLAB software. That is, sweeping over all $X(t)$, $X(t) - X(t - dt)$ was iteratively calculated, resulting in a vector of data stored as wind ramps with dt duration. With 10-minute resolution, there are 105119 wind ramps for analysis. Figure 2-4 shows the wind ramps, which are also the errors in persistence forecasts, as a distribution over the two years divided into 1000 bins between the absolute extreme values of the distribution. A small portion of the histogram is shown in Figure 2-5, showing the difference in the Gaussian fit versus the actual data. The maximum value was 232.009 MW (the greatest down ramp was 232.009 MW), while the minimum value was -213.062 MW (the greatest up ramp was 213.062 MW).

By simple inspection, it can be seen that the distribution is not well approximated by the Gaussian distribution. The Gaussian peak predicts the number of events to be around

1600, while the actual data presents a frequency of around 11,000 at its peak. Also, using the Gaussian curve to predict the number of events past + or - 90 MW suggests that no events should occur, while the empirical data shows events out to hundreds of megawatts.

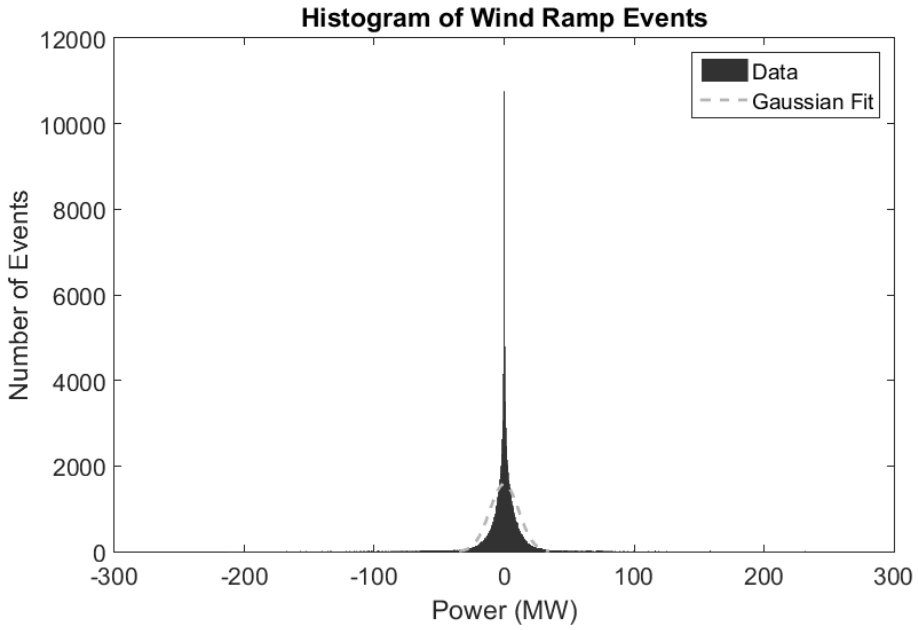


Figure 2-4 Frequency of Wind Ramps with 1000 Uniform Bins with Gaussian Fit Overlay

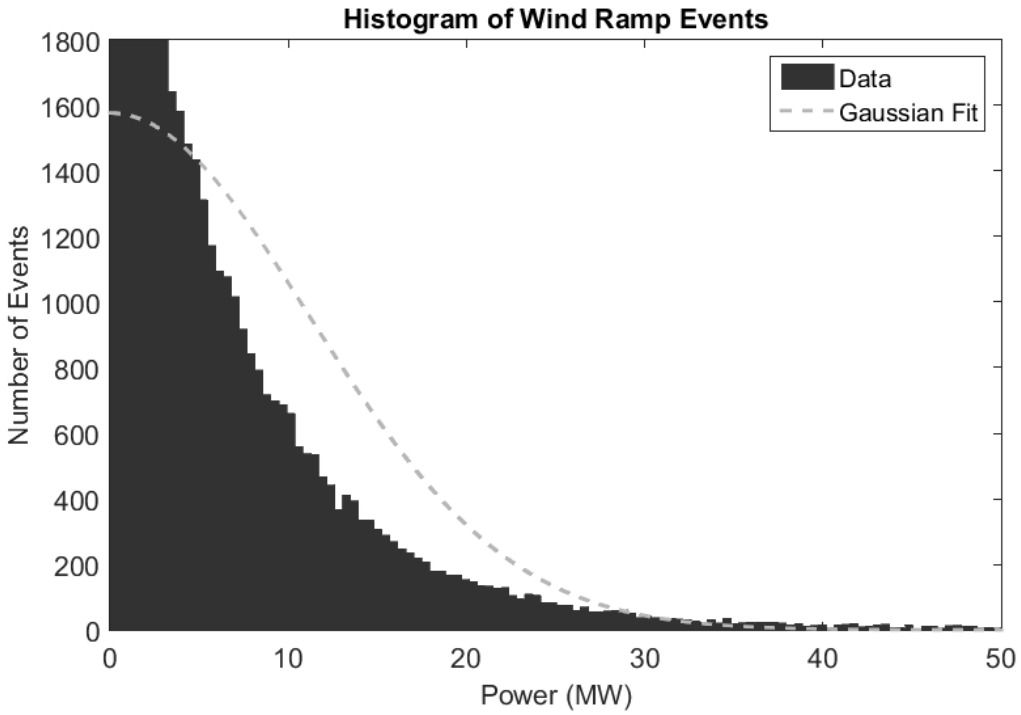


Figure 2-5 Close-up of One Section of the Gaussian Overlay

Using the Gaussian distribution to predict the errors will incorrectly place weight within a few standard deviations of the mean, falsely showing a light tail over the clearly heavy tail of the data.

2.3.2. Extreme Value Analysis Application to Wind Ramps

To characterize wind power ramps, two years of wind farm output data was used for analysis. Two separate methods for characterization of extreme probability events were used based on the first and second theorems in extreme value theory, the annual maxima series (AMS) method and the peaks over threshold (POT) method. Both methods utilized a user-defined parameter *dt* to establish the time over which a 'ramp' would be defined.

There was a total of 105120 data points, and 49372 (about 50%) were down-ramps using a 10-minute resolution. Of the 50%, about 730 were used for the AMS method while

10430 of those were used for extreme value analysis (EVA) with the POT method (about 20% of the down-ramps, as P ($X > s_n$) ~ 0.2). The standard deviation of the data was σ = 11.165 MW.

To calculate frequency of ramp events, constant bin sizes were used, and any bins with a zero value were erased from the final data set. Parameters for each distribution were created using maximum likelihood estimators by sweeping over other controlling parameters if necessary. Best fits between the empirical and theoretical CDFs are shown in Figure 2-6 and Figure 2-7. Then, correlation coefficients were calculated between the observed and model-estimated data, and the coefficient of determination R^2 values are shown in the Table 2-1 and Table 2-2 for both the AMS and POT methods, respectively.

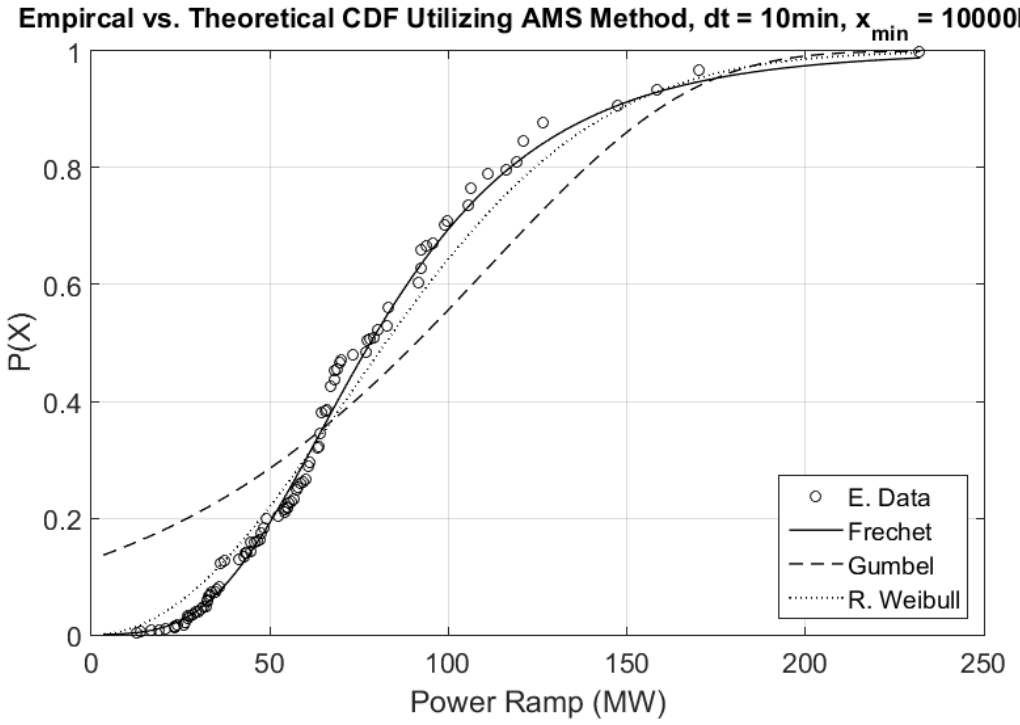


Figure 2-6 Fits of Gumbel, Fréchet, and Reverse Weibull CDF Distributions of the Empirical Data

Table 2-1 R^2 for Various Distributions (AMS Approach)

***dt* (min)**	**Type I**	**Type II**	**Type III**
10	0.92312417	0.99673653	0.98932896
20	0.93840394	0.98988342	0.98133986
30	0.95552943	0.99042254	0.98638262

For the AMS method, daily maxima were used, meaning that over the two years only 730 data points were used.

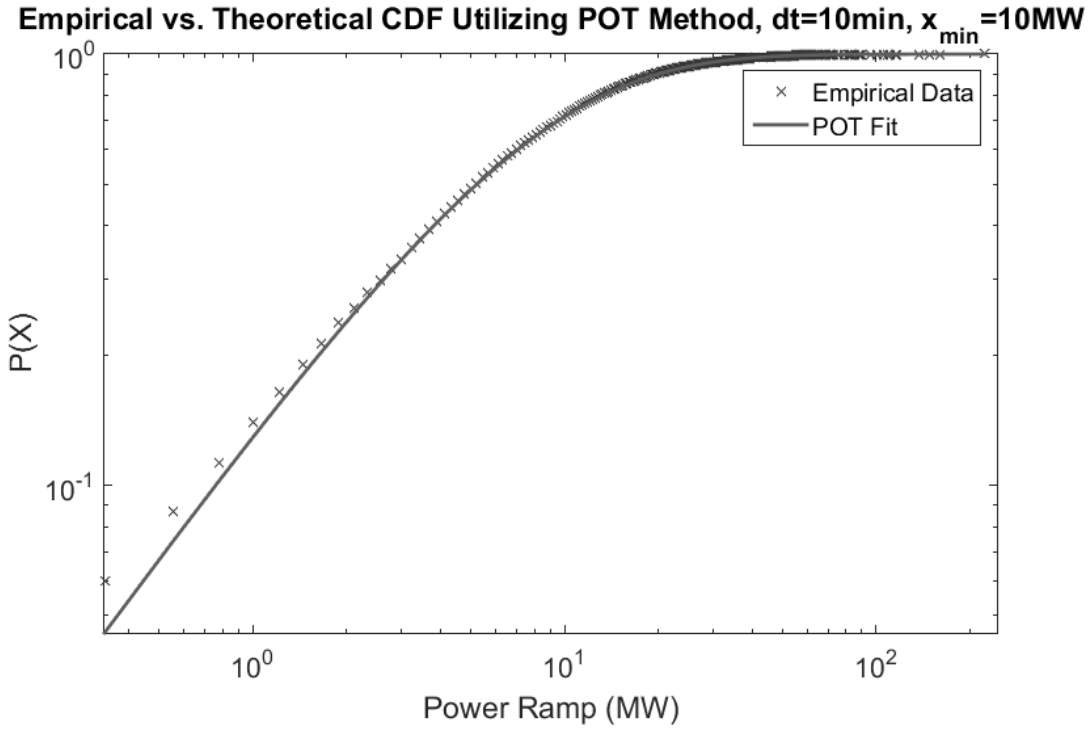


Figure 2-7 Frequency of Wind Ramps with 1000 Uniform Bins

Table 2-2 R^2 for Generalized Pareto Distribution (POT Approach)

	x_{min} **(MW)**				
***dt* (min)**	**10**	**15**	**20**	**25**	**30**
10	0.99963597	0.9980683	0.99831987	0.99743223	0.99281193
20	0.99989439	0.99983859	0.99930766	0.9986667	0.9991671
30	0.99995681	0.99991618	0.9997878	0.99953153	0.99894556

For the POT method, 10430 data points were used for a threshold of 10 MW ramps. There is around ten times more data available for the POT method versus the AMS method, displaying one of the strengths of the POT method. Note that if X is the dP down-ramp, $P(X) = P(X, X > s_n) = P(X > s_n) * P(X \mid X > s_n)$.

The POT method fit showed a shape parameter ξ of 0.170 and a scale parameter σ of 7126, with a x_{min} of 10 MW. The positive shape parameter suggests that the best representation comes from a Fréchet Type II distribution under the generalized Pareto distribution (GPD) family.

Determining x_{min} for characterization with the POT method involved the bootstrapping procedure discussed before, as well as verifying the choice through various other common methods. Usually, the first test is 'eyeballing' the log-log plot of frequency vs. value, shown in Figure 2-8, choosing a threshold value of x_{min} that seems to sit on the beginning of the linear tail of the plot. It can be seen that at 10^4 kW, or 10 MW, the distribution seems to transition into its tail. A second test was performed with the Hill plot, where a suitable threshold x_{min} can be found by finding a general region in which the Hill plot starts to converge. In Figure 2-9, the Hill plot shows a flattened response around 10 MW as well.

To verify these observations, a grid of n_1 was chosen to be 1000 to 5000 in steps of 500. A finer grid or a larger range could have been chosen to improve the consistency of the results. The lowest ratio $\mathrm{AMSE}(z_{n_1})^2/\mathrm{AMSE}(z_{n_2(n_1)})$ was seen at a threshold level of $x_{min} = 22.7$ MW. This was close enough to verify the choice of $x_{min} = 10$ MW, so this value was used in the POT method fitting of the data. Final POT fitted results for the values of the parameters are $\xi = 0.1707$, $\sigma = 7125.9$, and $x_{min} = \mu = 10$ MW.

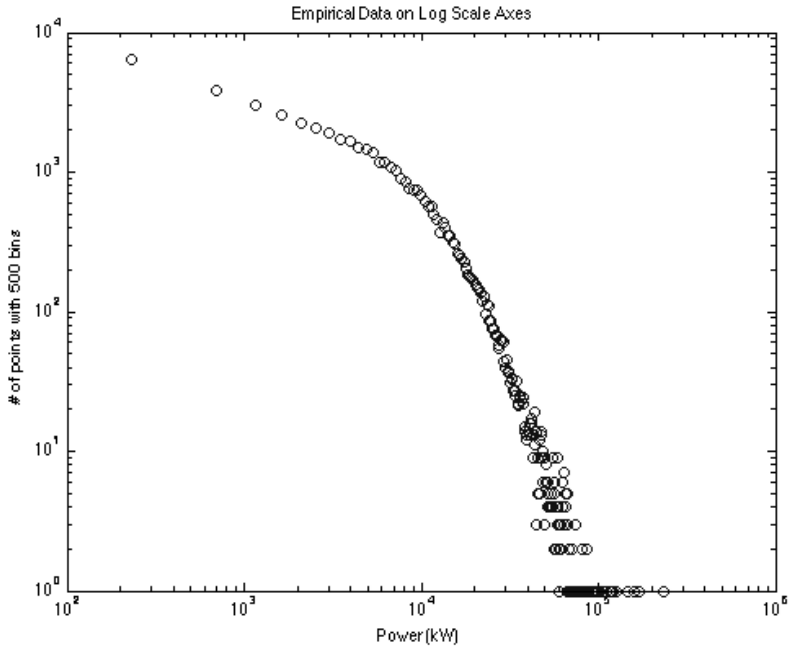


Figure 2-8 Empirical Data of Frequency vs. Power Ramp

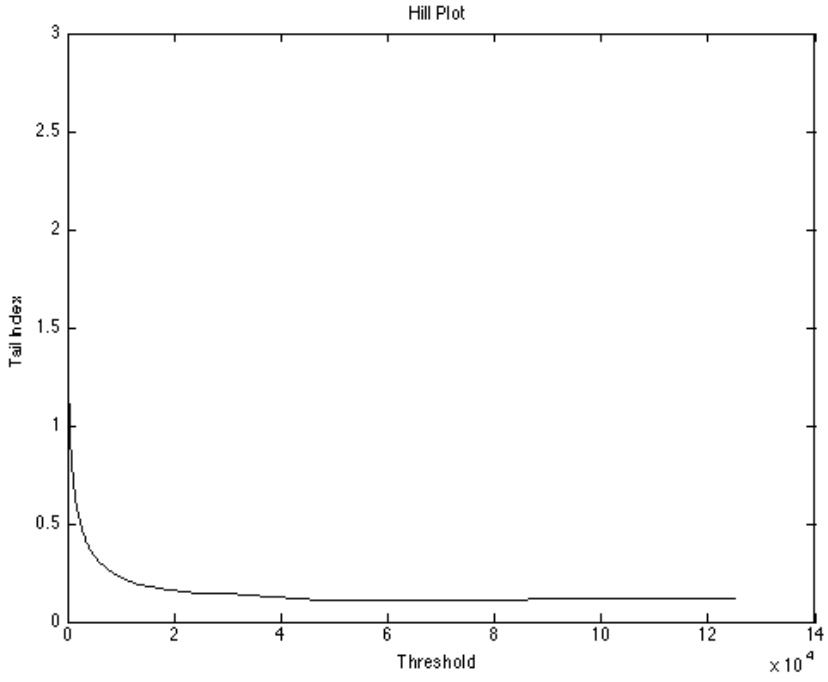


Figure 2-9 Hill plot of Empirical Data

For planning purposes, this distribution must apply to future years of interest, so there must be evidence of non-stationarity between years of data. Xcel Energy's 2009 data was fit to a Fréchet GPD curve and was compared to the fit of the second year. The two results were very similar, with the PDFs of the two shown in Figure 2-10 (R^2 of 0.9993). Figure 2-11 shows the 90% confidence bounds around the 2-year best GPD fit.

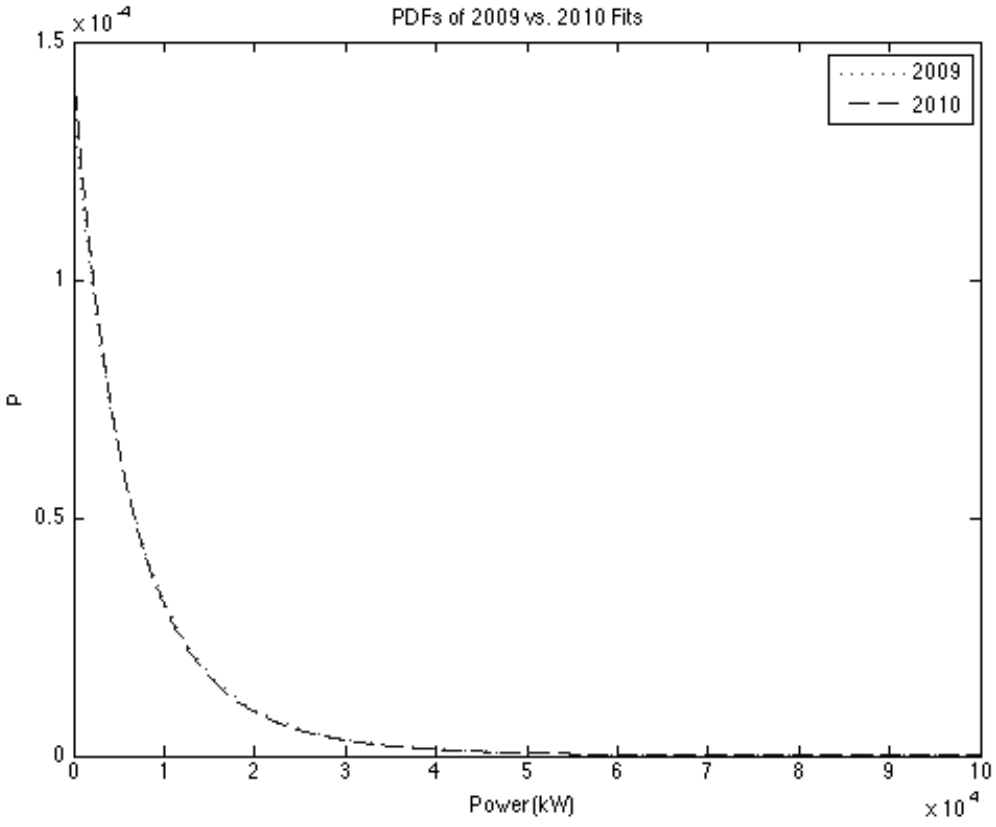


Figure 2-10 PDFs of Year 1 (2009) and Year 2 (2010)

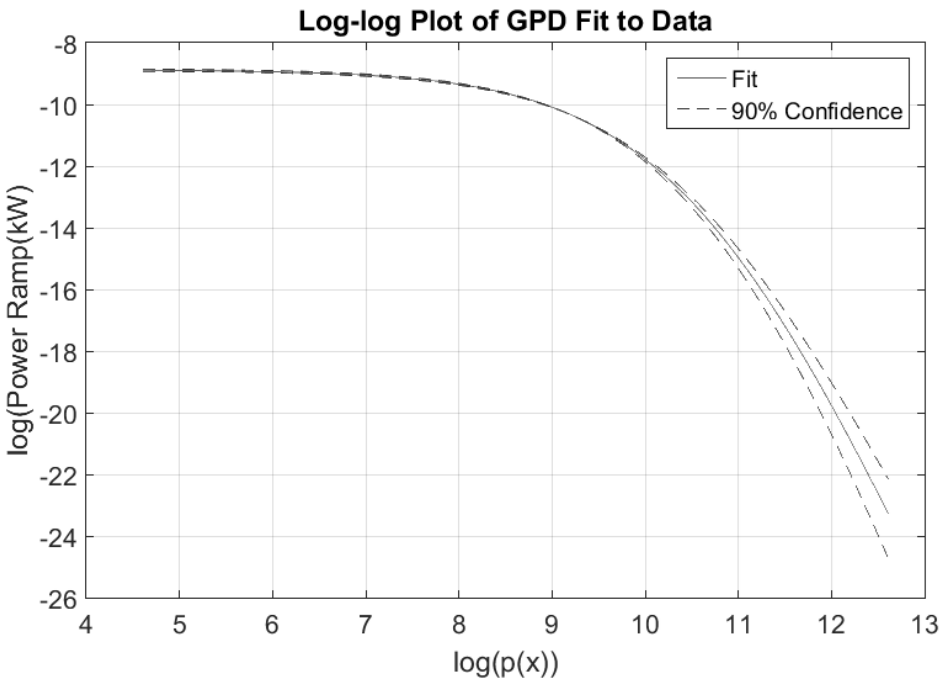


Figure 2-11 Log-log Plot of a 90% Confidence Interval about GPD Fit

To display the fit to the empirical data, Figure 2-12 shows the GPD curve fit to the empirical data in a log-log format for better visibility. In this figure, the *x*-axis shows the logarithm of the power above the x_{min} threshold cutoff of the POT method. It seems that the extremely high power ramp data does not fit the curve very well, over about 11.5 MW on the plot. However, the theoretical curve at this point dips below zero on the *y*-axis log scale, and the logarithm of frequencies of an event cannot be negative. Anything over 11.5 MW lies in a region where the probability is very small that an event would occur given the number of samples. The data is more sparse (spaced apart) in this area, which agrees with the fit. If much more data was used, the curve would be shifted up and the empirical data would fit well onto the curve.

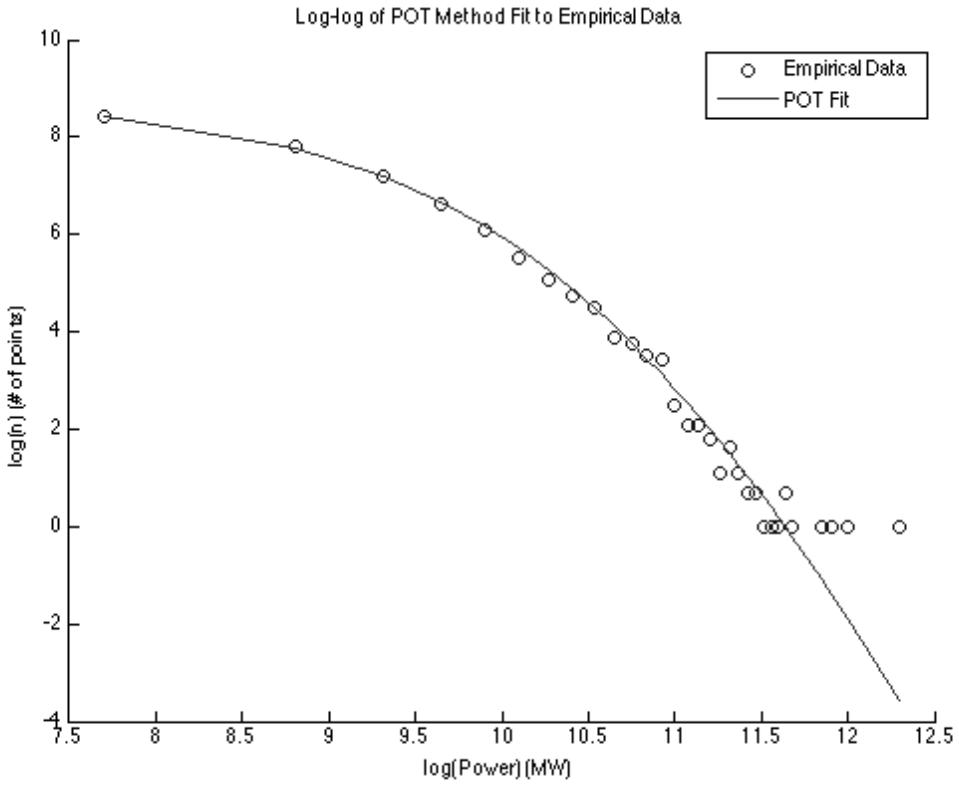


Figure 2-12 Log-log Plot of GPD Fit to Empirical Data

The POT fits shown are given for a shifted version of the data. The threshold cut-off x_{min} is simply subtracted from all the data such that the GPD would easier fit the data. Because this extreme value distribution only accounts for 20% of the down-ramps, the majority of the down-ramps (the lower 80%) must be modeled with another distribution. The uniform or normal distribution would suffice for this, though based on the sharp drop-off some might choose to use a beta distribution.

2.3.3. Wind Ramp Stationarity

Though some work has already shown that year-to-year stationarity in the wind ramp data exists, a more formal procedure testing wind ramp stationarity is needed.

Currently, there is no work that explicitly proves that wind ramps are stationary, but it is well known that wind speeds are non-stationary [74, 75]. This work has assumed that the ARIMA models applied to wind speeds hold, or that the wind speed is difference stationary. However, testing for a deterministic trend within the year as well as stochastic trends throughout the wind ramp processes would be necessary to ensure that the assumption holds.

2.4. Conclusions on EVA

The results show that traditional statistical methods do not appropriately model wind ramp events, corresponding to the errors in persistence wind power forecasting methods. EVA methods were applied to fit various distributions to the data, with both AMS and POT approaches. The POT approach was first used with traditional fitting methods, and then the parameters were verified with a newer approach using sub-sample bootstrap and control variate techniques. The wind ramp distribution can be used to take into account the time-varying nature of wind power, used in applications such as controller design or power system planning.

This chapter discussed wind power modeling which is useful in offline planning situations. For improvements in real-time forecasting, the next chapter presents a metric which will be helpful in determining the set of training data to be used, subject to the quantified degree of non-stationarity in the signal.

3. NON-STATIONARY DATA ANALYSIS FOR SHORT-TERM WIND POWER FORECASTS

3.1. Importance of Stationarity for Statistical Forecasting Methods

In short-term wind forecasts, statistical models are preferred to physical models as physical modeling approaches are generally computationally intensive or inaccurate for use in short-term wind forecasting [79]. Past data is used to produce point forecasts or probabilistic forecasts in statistical models, both for parametric and non-parametric models. These statistical models are frequently built with the assumption that the modeled data is, in some sense, stationary, even though real time series data seldom meet any criteria for stationarity. However, there is no common metric intended for use in a real-time forecasting algorithm that quantifies the level of stationarity of time series data.

This chapter will detail a novel approach to quantify the degree to which a signal may be non-stationary, based on existing work in Empirical Mode Decomposition (EMD). This work will 1) introduce a metric called the Ensemble Degree of Non-Stationarity (EDNS) to quantify the degree of non-stationarity present in a time series, 2) introduce a real-time algorithm to adaptively determine optimal training windows with the assistance of the EDNS, and 3) analyze the performance impact of determining optimal training windows using this real-time algorithm on two different sets of wind power output data. The work is organized as follows. Section 3.1 covers the importance of stationarity, also

exploring basic definitions and existing tests for stationarity. Section 3.2 reviews three popular models used for forecasting, the persistence model, the autoregressive model and the support vector regression model. Section 3.3 describes the novel EDNS metric, detailing the steps to arrive at the EDNS. Section 3.4 presents a real-time algorithm that uses the EDNS, also showing some simulation results using this real-time algorithm. The chapter concludes in Section 3.5.

3.1.1. Motivation for Quantifying Stationarity

The length of training data can affect the forecasting performance of statistical forecasting methods, which will be illustrated through an example in this section. Autoregressive (AR) models were used to perform one-step-ahead forecasts on two separate 2-month segments of 5-minute resolution wind power output data in a 160 MW wind farm in the Australian Energy Market Operator (AEMO) during 2012, one year of a two-year data set of wind farm power outputs in AEMO[1] [80]. A new AR model was fitted at each of the 34,560 time steps in each segment with a sliding window of the most recent set of data of length T. Four lags were used in the AR models as a result of applying the Bayesian Information Criterion (BIC) [81]. See Section 3.2.2 for details on AR models.

Training window lengths ranging from 2 days to 90 days were used, tested in increments of 1 day, which resulted in the forecasting errors shown in Figure 3-1 and Figure 3-2. The mean absolute error (MAE) is an important metric in this work, where for some time series x with n elements, and with a predicted time series $\hat{x}$, the mean absolute error

[1] The authors would like to acknowledge the original provider of the data (AEMO in Australia), as well as Jethro Dowell (University of Strathclyde), Stefanos Delikaraoglou, and Pierre Pinson (Technical University of Denmark), for preparing the dataset and making it available.

is defined as MAE $= \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_i - x_i|$. A roughly "convex" behavior is seen over each of the two datasets, where a minimum mean absolute error (MAE) exists for some training data length. The result shows a clear dependence of the forecasting performance on the length of the training dataset for AR. In addition, it is seen that a poor choice of the training window length may result in worse AR forecasting performance in comparison with persistence forecasting. Lastly, it is clear that the optimal training window length is different in the two subsets of data. The optimal training length in the 1st subset of data is approximately 17-25 days, whereas the optimal training length in the 2nd subset of data is approximately 30-40 days. Appendix A I contains results for a similar test with another forecasting method, SVR.

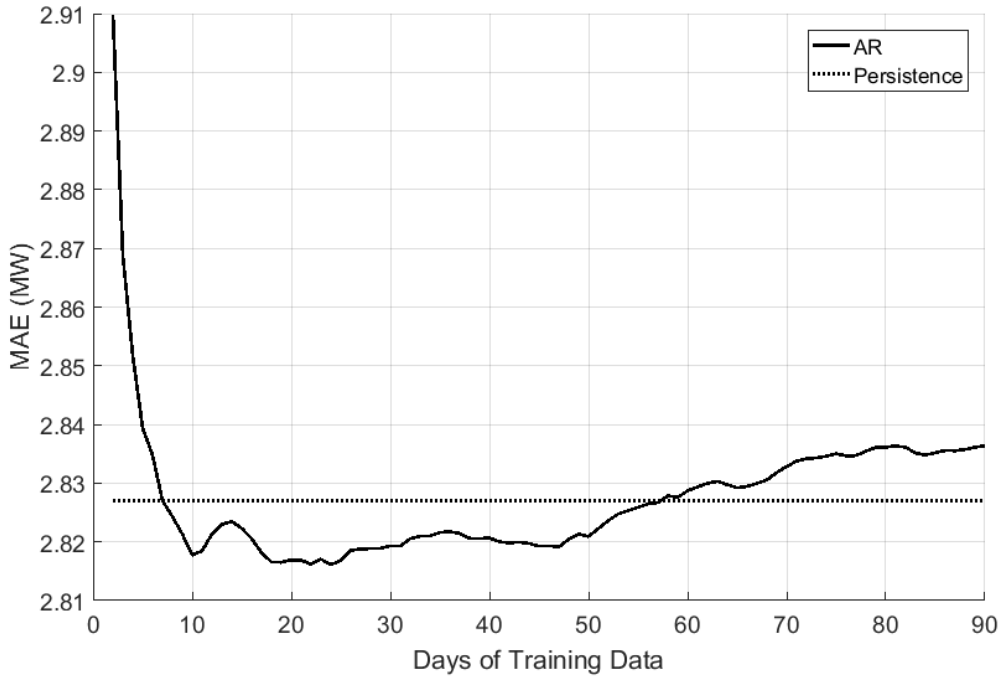


Figure 3-1 AR MAE Error Versus Training Data Length for the 1st Subset of AEMO Data

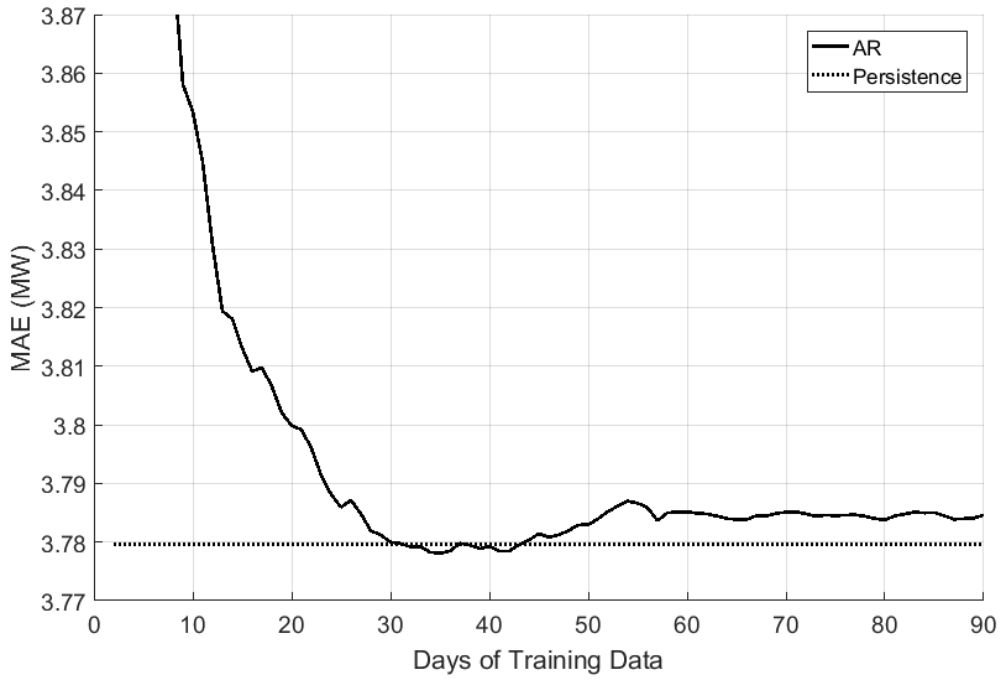


Figure 3-2 AR MAE Error Versus Training Data Length for the 2nd Subset of AEMO Data

Intuitively, it is well known that noise dominates parameter fitting with small amounts of training data, so it is common practice to use as much training data as possible to obtain a forecasting model. However, using excessively large amounts of training data may result in extra forecasting error because of the inclusion of some "irrelevant data" far away from the present time. Specifically, the irrelevance phenomenon can be attributed to non-stationarity, because the underlying model for the process may have changed over time. As a result, it is desirable to avoid including excessively large amounts of data in a non-stationary time series. However, a key question is: At what time scale does a training window contain an excessively large amount of data? Historically, this point of excess has been found by increasing the training data length until forecasting performance suffers, but this is not rigorous and grants only a qualitative insight into the non-stationarity of the

process. Further, one static training window (e.g. 60 days) is typically applied to an entire time series, though it is seen through the present example that the point of excess changes over time, and so the training window should change over time. This work seeks to quantify the non-stationarity of a time series process in a rigorous manner, which should also allow for a quick identification of the point of excess in training windows. This metric of non-stationarity would then allow for the dynamic adjustment of training windows in real-time.

To reiterate, it is presently common practice in wind power forecasting to select a single training window length that gives a minimum error for some amount of data (e.g. 1 year) as a result of a sweep over many training window lengths. It is unlikely that a drastically small training data set would be selected because of the large errors introduced by such a small training data set that would be seen (see Figure 3-1 and Figure 3-2). However, slight over-selection or under-selection is likely, especially as standard forecasting methods do not adjust training window size over time at all. Based on Figure 3-1 and Figure 3-2, expected improvements of correctly adjusting the training window size may be relatively small, when compared to a static, intelligently selected training window. For example, if a 90 day static training window was selected, there could be at maximum a 0.5% decrease in error by adjusting the training window size, as a 90 day static training window is close to optimal. Even smaller improvements may be seen if the static training window was chosen to be 40 days, with this data set. However, any performance gain in forecasting is desirable, however small.

This exercise was motivated by a work in which autoregressive and vector autoregressive forecasting models were used on this AEMO dataset, which resulted in errors greater than that of persistence forecasting [82]. A constant window size of 60 days

was used for AR forecasting in this study over an entire year of data, resulting in an AR MAE of 2.347% of nominal capacity versus the persistence MAE of 2.308%. Transformations were applied to the data in this work to ensure that the noise was Gaussian white noise. However, the assumption of stationarity was not addressed, which may account for the majority of the extra error in AR over persistence forecasting.

3.1.2. Definitions of Stationarity

To understand the importance of stationarity in forecasting and the construction of a stationarity metric, a brief review of stationarity is necessary. First, formal definitions of stationarity will be presented. Let X_t be a stochastic process as function of t, the index of time, and let $F_X\left(x_{t_1+\tau} \dots x_{t_k+\tau}\right)$ be the cumulative distribution function (CDF) of the joint distribution of X_t. A time series X_t is defined to be strictly stationary if, for all τ, for all k, and for all $t_1, \dots, t_k$,

$$F_X\left(x_{t_1}, \dots, x_{t_k}\right) = F_X\left(x_{t_1+\tau}, \dots, x_{t_k+\tau}\right). \quad (3\text{-}1)$$

That is, a time series is strictly stationary if the joint distribution for some contiguous range of values in X_t is the same for any other contiguous range of values in X_t. A weaker definition of stationarity is wide-sense stationarity (WSS), which applies to a time series X_t when the conditions below are met:

$$E(|X_t|^2) < \infty \quad (3\text{-}2)$$

$$E(X_t) = m \quad (3\text{-}3)$$

$$C\left(X_{t_1}, X_{t_2}\right) = C\left(X_{t_1+\tau}, X_{t_2+\tau}\right) = C(t_1 - t_2) \quad (3\text{-}4)$$

where $C(\cdot)$ is the covariance function and $E(\cdot)$ is the expected value. The first WSS condition (3-2) states that the variance $E\left(\left(X_t - E(X_t)\right)^2\right)$ is finite, while the second

condition (3-3) states that there exists a mean m for the time series. The third condition (3-4) states that the covariance of the time series with a shifted version of itself is only a function of the shift τ. Other definitions of stationarity exist, such as asymptotic stationarity, which deals with restricting the definitions of WSS to the regime in which τ approaches infinity in (3-2) through (3-4) [83]. However, WSS remains the most widely used notion of stationarity, and so stationarity in this work will now refer to the WSS definition of stationarity. Further details of stationarity can be seen in many existing references or textbooks [84].

The stationarity of a signal relates to the fact that the statistical properties of the signal remain constant over time. If a time series meets the criterion of stationarity, the time series would be able to be forecasted to a high degree of accuracy. In other words, if a dataset were to be described by an unchanging, underlying model, one model could be fit to the existing data, and all forecasts would be accurately represented by this model. However, existing definitions of stationarity are extremely restrictive, and there exist very few datasets (even in controlled, ideal simulations) that meet the criterion of being stationary. There is also no agreed-upon metric to show *how* non-stationary a time series is. Thus, time series analysis techniques are typically applied to non-stationary datasets. It is worth noting that, although measured data is seldom stationary, there has been much success in the application of statistical forecasting methods to existing data, even to wind speed or power output data, which is recognized to be highly variable and non-stationary [85]. An underlying metric which can show the degree to which a time series is non-stationary would be useful to verify the stationarity of a training set, or to even select a suitably stationary training set for a forecasting model.

3.1.3. Existing Methods to Test for Stationarity

Although there are no common methods to quantify the degree of stationarity in a signal for real-time forecasting use, there are existing methods to test for stationarity within a dataset. These methods typically test for a specific type of non-stationarity within the dataset as well, which include sensing the existence of a deterministic trend or a stochastic trend (unit root). This includes ad hoc methods, unit root tests such as the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test, and spectrum analysis methods such as the Priestley-Subba Rao (PSR) test.

Ad hoc methods refer to the simple 'eyeball' analysis of the autocorrelation function of a signal, performed over many lags at various points in a signal. To be WSS, a signal must have a constant autocovariance curve regardless of the subset of data analyzed, so plotting the autocovariance at various points gives a user some notion of the non-stationarity in the signal. If the autocovariance function changes rapidly in one section, but stays relatively constant over another section, a relatively non-stationary and stationary segment has been identified. Although the method is simple and intuitive, this approach is very qualitative and subjective, as the procedure itself is not rigorously defined. The user cannot calculate the autocovariance function to infinite lags in a finite dataset, so some subjective lag length must be used. In addition, it is difficult to tell exactly how different one autocovariance function is from another, which often makes it difficult to tell even if one segment is more non-stationary than another.

The KPSS test was introduced in 1992 as a popular test to determine the existence of a stochastic trend, commonly called a unit root [86, 87]. The term 'unit root' comes from the fact that a coefficient of one is applied to the $t-1$ lag (when determining the point at

time t), which creates a random walk. This test takes as its null hypothesis the absence of a unit root, with an alternative hypothesis of the existence of a unit root. The test will signal that there is a unit root present if enough evidence exists; otherwise, the test will accept the null hypothesis of stationarity. There are other tests that have the same general goal, such as the Dickey-Fuller test, the Phillips-Perron test, and the ADF-GLS test, but these tests are all designed to tackle the exact problem of determining a unit root in a time series [88, 89, 90]. They are only designed to provide a binary decision and not a quantitative degree of non-stationarity. In addition, unit-root tests were created with the assumption that the analysis would be based on an autoregressive model, limiting the utility of the method in non-linear forecasting.

The PSR test was introduced in 1969 as one of the first spectrum analysis methods, which involves calculating the frequency information in the signal and determining the signal's stationarity from this [91, 92]. The PSR test uses the fact that a stationary time series has a constant frequency spectrum over time, and so investigates the Fourier spectrum $f_t(w)$ for its variations over time in order to reject or accept its null hypothesis of stationarity. Specifically, it analyzes the variance in the Fourier spectrum estimate by applying the standard rules of the analysis of variance (ANOVA). If there are significant differences in the Fourier spectrum over different sections of time, the PSR test rejects its null hypothesis of stationarity. These spectrum analysis techniques are meant to provide a binary decision on stationarity, but provides the notion of non-stationarity in a more general sense than unit-root tests as it is not meant to apply to simply an autoregressive framework. These methods are the closest works related to the novel EDNS metric that is provided in this chapter, as they are built on the same intuition of relating non-stationarity in the time

series to variations in the frequency spectrum. However, one critical disadvantage of spectrum analysis methods based on Fourier transforms or wavelet analysis lies in the lack of resolution and accuracy in the spectral estimate itself. A detailed review of the drawbacks of Fourier or wavelet analysis is presented in N. Huang's work [93].

In any case, a stationarity test is useful in forecasting only if it can influence the decision of the forecasting method in its selection of training data. The next section will introduce popular forecasting methods and the training data used in each method.

3.2. Popular Statistical Forecasting Approaches for Wind Power Forecasts

Existing statistical forecasting methods for short-term wind power forecasting include autoregressive (AR) forecasting and all of its variants as well as machine learning methods [94]. For instance, AR models and vector AR models can linearly relate past inputs of one or more time series to determine the output of a wind farm [82]. Machine learning methods such as artificial neural networks, Markov chains, and support vector machines (SVMs) can generally fit a non-linear model, and provide a forecast from a smaller set of training data [31, 85]. This section provides some background on three popular time series forecasting models that are used in this work, the persistence model, the AR model and the SVR model.

3.2.1. Persistence Forecasts

The persistence model is the simplest statistical model that can be used for short-term forecasting, but the model tends to perform very well in practice. Persistence forecasting assumes no change in output, such that

$$X_t = X_{t-1} \qquad (3\text{-}5)$$

where X_t is the wind power process at time t, which is assumed to be the same value as in the previous period $t-1$. Persistence forecasting remains the main forecasting method by which all new short-term forecasting methods are benchmarked against in the wind power forecasting community, partially because of its good performance but also because of its simple implementation. No training or tuning is needed with this method, and the method can provide a forecast instantly. However, because of its lack of training data and model complexity, it does not have the ability to model time-varying phenomena accurately.

3.2.2. Autoregressive (AR) Forecasts

Autoregressive (AR) models define a process in which an output variable depends linearly on previous values in the process, a stochastic noise term, and a constant. This AR model can be written in discrete form as

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t + c \qquad (3\text{-}6)$$

where the process X_t at time t is defined by p previous data points weighted by the AR coefficients α_i, a Gaussian white noise term ε_t, and some constant c. To fit an AR model, some set of training data with a vector X_t and a set of vectors X_{t-i} is used to estimate the α_i parameters.

Given a model with $p = 1$ lags, fitting coefficients α_i to a set of training data can be as simple as the use of a least-square estimator, solving for α_1 in the over-determined system $X_t = \alpha_1 X_{t-1}$. However, for a larger set of lags, the computation and inversions involved in directly calculating α values can be excessive. The Yule-Walker equations can then be used to solve for the AR coefficients, to save on computation time [95].

To calculate the lag length p, typically the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are used [81, 96]. Both the AIC and BIC present a metric that weight the complexity of the model with the likelihood or accuracy (goodness of fit) of a model, shown here for completeness:

$$\mathrm{AIC} = 2k - 2\ln(L) \quad (3\text{-}7)$$

$$\mathrm{BIC} = k\ln(n) - 2\ln(L) \quad (3\text{-}8)$$

where k is the number of free parameters, n is the sample size, and L and is the optimized likelihood value of the model (or, the probability that the outcomes came from the parameter fit of the model). Given a set of candidate models, the optimal model is the one with the lowest AIC or BIC. Note that there is a trade-off captured between model complexity and accuracy in both the AIC and BIC, which is meant to prevent overfitting.

After a model is selected, the training data is selected, which determines the best fit α coefficients in the AR model. To train the model, m previous data points are selected, where m is the number of training samples. The computational complexity of the model training (parameter fitting) in AR is $O(m^2)$, when assisted by the use of the Yule-Walker equations in conjunction with the Levinson-Durbin algorithm. The choice of m is important and ultimately affects the performance of the forecasting model, as shown in Figure 3-1 and Figure 3-2 in Section 3.1.1.

3.2.3. Support Vector Regression Forecasts

Support Vector Regression (SVR) is a supervised learning method which can provide non-linear fitting to data. SVR is built on Support Vector Machines (SVM), which is applied in many classification and function fitting problems. Only a brief explanation of

SVR will be presented here, and more details on SVR and other tools using support vectors can be found in many reviews and papers [97, 98, 99]. This method can be thought of as a simple linear regression similar to AR, but in a transformed, higher-dimensional space.

Let empirical data be provided in the form $\{(x_1, y_1), \ldots, (x_m, y_m)\} \subset \mathcal{X} \times \mathbb{R}$ for some training data set length m. The goal of SVR is to map some set of features $x_1, \ldots, x_m$ to corresponding labels $y_1, \ldots, y_m$ by first mapping the features into a higher-dimension feature space $\mathcal{F}$ via some nonlinear mapping Φ and perform linear regression in this higher dimensional space. For clarity, the linear case will be first introduced and the nonlinear mapping will be introduced later. Also, for its relevance to time series forecasting, the features space $\mathcal{X}$ will be limited to the d-dimensional real space $\mathbb{R}^d$. Consider the following function

$$f(x) = \langle w,x \rangle + b \qquad (3\text{-}9)$$

where $\langle w,x \rangle$ is the dot product of w and x in $\mathbb{R}^d$. An accurate model with a low degree of complexity is desired for this fit. The complexity of the model is related to its flatness, where a greater degree of flatness corresponds to a smaller w. Minimization of the norm of w would ensure a smaller w and thus a less complex model, so the fitting problem can then be seen as an optimization problem

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|w\|^2 \\ &\text{subject to} && y_i - \langle w_i,x_i \rangle - b < \varepsilon \\ & && -y_i + \langle w_i,x_i \rangle + b < \varepsilon \end{aligned} \qquad (3\text{-}10)$$

where $\|w\|^2$ is the inner product of w with itself [100]. This optimization will perform a linear fit on the provided data such that all pairs (x_i, y_i) will be within ε of the fitted function f. However, this optimization problem has limited utility in that a large ε would

have to be used to ensure feasibility, and so the fit may not be very accurate or precise. Thus, slack variables ξ_i and ξ^* can be introduced as shown

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*) \\ &\text{subject to} && y_i - \langle w_i, x_i\rangle - b \;< \varepsilon + \xi_i \\ & && -y_i + \langle w_i, x_i\rangle + b < \varepsilon + \xi_i^* \\ & && \xi_i, \xi_i^* \;\geq 0 \end{aligned} \qquad (3\text{-}11)$$

such that a constant $C > 0$ penalizes the deviations of points beyond ε of the fitted function. This constant C determines a trade-off between the flatness of the function and the tolerance of deviations greater than ε from the fit, so larger values of C will encourage a less flat, or more complex, fit. This tolerance of values larger than ε introduces what is called an ε-insensitive loss function $|\xi|_\varepsilon$ described as

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise.} \end{cases} \qquad (3\text{-}12)$$

Figure 3-3 shows the graphical representation of a linear SVR fit, taken from [98], which shows the ε-insensitive loss function.

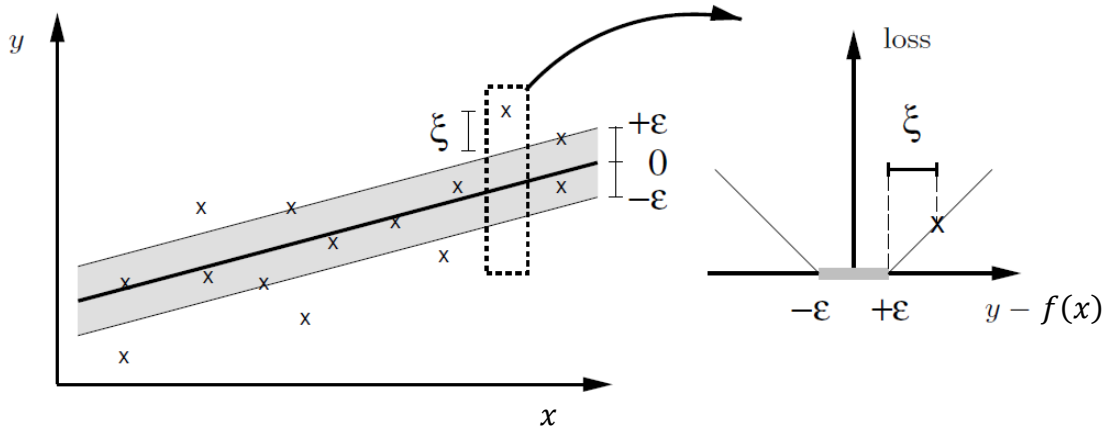


Figure 3-3 Graphical Representation of SVR, with the ε-insensitive Loss Function [98]

Now, the Lagrange function of the formulation in (3-11) is shown:

$$L := \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*) - \sum_{i=1}^{\ell}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$
$$-\sum_{i=1}^{\ell}\alpha_i\,(\varepsilon + \xi_i - y_i + \langle w_i,x_i\rangle + b)$$
$$-\sum_{i=1}^{\ell}\alpha_i^*\,(\varepsilon + \xi_i^* + y_i - \langle w_i,x_i\rangle - b) \qquad (3\text{-}13)$$

where η_i, η_i^*, α_i, and α_i^* are the Lagrange multipliers and thus all non-negative. The partial derivatives of L with respect to the primal variables w, b, ξ_i, and ξ_i^* must be zero at an optimal solution, so

$$\delta_b L = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) = 0 \qquad (3\text{-}14)$$

$$\delta_w L = w - \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)x_i = 0 \qquad (3\text{-}15)$$

$$\delta_{\xi_i} L = C - \alpha_i - \eta_i = 0 \qquad (3\text{-}16)$$

$$\delta_{\xi_i^*} L = C - \alpha_i^* - \eta_i^* = 0. \qquad (3\text{-}17)$$

The dual optimization problem is then

$$\text{maximize} \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\ell}(\alpha_i - \alpha_i^*)\big(\alpha_j - \alpha_j^*\big)\langle x_i,x_j\rangle \\ -\varepsilon\sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{\ell} y_i(\alpha_i - \alpha_i^*) \end{cases}$$
$$\text{subject to} \quad \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) \quad = 0 \qquad (3\text{-}18)$$
$$\alpha_i, \alpha_i^* \qquad \in [0, C]$$

as η_i and η_i^* have been substituted in as $\eta_i = C - \alpha_i$ and $\eta_i^* = C - \alpha_i^*$. Also, (3-15) can be written as $w = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)x_i$, so

$$f(x) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\langle x_i,x\rangle + b. \qquad (3\text{-}19)$$

which provides the insight that w is described as a linear combination of the training data x_i. Also, note that when finding the fit $f(x)$, w does not even need to be explicitly computed, as the dot products between the data is all that is needed. The parameter b is

computed by using the Karush-Kuhn-Tucker (KKT) conditions, and details can be seen in the references [97].

Next, the non-linearity will be introduced. As the dot product between training data determines the fit of the function, kernel methods can be used to provide a computationally efficient method to map the existing feature set into a higher dimension [101]. The features in this higher dimensional space then can be better fit by a linear function, which provides a non-linear fit in the original space $\mathbb{R}^d$ of the training data. The kernel function $k(x_i, x)$ is now introduced, where

$$f(x) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)k(x_i, x) + b \qquad (3\text{-}20)$$

and the SVR algorithm applies nonlinear fits by transforming the training data through the kernel by a map $\Phi: \mathbb{R}^d \to \mathcal{F}$ where $\mathcal{F}$ is some feature space. This work will use a popular kernel function called the Gaussian radial basis function (RBF), which is typically used for datasets that require a non-linear fit. The Gaussian RBF is written as

$$k(x_i, x) = \exp(-\gamma\|x_i - x\|^2) \qquad (3\text{-}21)$$

where γ is a free parameter that the user of the SVR algorithm must choose.

It is important to point out that the optimization problem is a quadratic programming problem, as the dual objective function is quadratic in α_i and has linear constraints. In addition, the problem itself is convex, and thus has a unique global optimum. Thus, a variety of convex quadratic programming methods can be used to solve this problem, including interior point methods.

(i) Considerations for SVR Implementation with Gaussian RBF Kernel

There are three parameters that are user-defined in the SVR algorithm with the use of the Gaussian RBF kernel: the parameter C, the parameter ε, and the Gaussian RBF kernel parameter γ. The typical approach to fitting an SVR model is to sweep all parameters over a large parameter space and select the parameter set that results in the highest accuracy, or lowest fitting error (so for example, each parameter can be swept over some set of values $\{2^{-15}, 2^{-13}, \dots, 2^{15}\}$) [85, 102]. The model complexity is controlled by both C and ε, so C can be fixed while ε remains a free parameter, to save on computation time [103].

In addition, cross validation within the training data set should be used to ensure a robust set of parameters. In a k-fold cross validation approach, where k is usually 5 or 10, the training data is split into k subsets. One subset is selected as the test data and $k-1$ subsets are selected to be the training data, and a parameter grid search is performed to find the optimal parameter values. This search is run k times such that each subset is the test dataset exactly once, and parameter values with the highest average accuracy are selected to represent the entire training dataset. Lastly, before using SVR, it is important to normalize the time series to $[0,1]$ or $[-1,1]$ to avoid numerical difficulties.

Existing software packages such as R and MATLAB have well-documented toolboxes that can handle SVR and SVM. However, an external, open-source library *LIBSVM* was used in this study for its flexibility (with source code in C++, Java, and functions for use in MATLAB), prevalence in existing literature, and available extensions [104].

The complexity of applying SVR to forecasting problems widely varies with the use of different kernel functions. The worse-case complexity for training an SVR model

once is $O(m^3)$ where m is the number of features in the training data, though in practice, the complexity is seen to be approximately $O(m^2)$ [105]. The complexity for running an SVR model (for fitting or prediction with test data) is linear with respect to the number of support vectors, or $O(mS)$, where S is the number of support vectors in the SVR model. Added onto the core complexity of SVR model training is the computation time added by running parameter sweeps as well as the k-fold cross validation. Specific to this work, only two parameters are examined as free parameters in a grid search, introducing approximately 60 iterations of the original SVR training problem, as well as a 5-fold cross validation, resulting in 300 total iterations of the SVR training problem. Formally, this would be as a worst-case complexity of $O(k\varepsilon_{sw}\gamma_{sw}m^3)$, where k is the number of subsets of data in the k-fold cross validation and ε_{sw} and γ_{sw} are the number of parameters values to sweep over for ε and γ, respectively.

The selection of the training data length m is as important in SVR forecasting as it is in AR forecasting, as seen in Appendix A I. The length of training data will affect the SVR model fit, but there is presently no common way to select the training data set other than with a brute force approach that iteratively tests some range of training data lengths, where the training data length with the lowest error is selected. A metric that quantifies the non-stationarity in the data would show to what degree the underlying model has changed over time, and it would provide insight into the selection of an appropriate training window.

3.3. The Ensemble Degree of Non-Stationarity (EDNS) Metric

To quantify the non-stationarity of a time series, the EDNS metric is introduced. The EDNS metric is based upon the use of the Hilbert-Huang Transform and the Hilbert

amplitude spectrum, all of which will be reviewed in this section. Results will be presented for each step of the process with a simple example.

3.3.1. The Hilbert-Huang Transform

The Hilbert-Huang transform (HHT) was introduced as a tool to analyze non-linear, non-stationary time series data by decomposing a signal into instantaneous frequency components [93]. In obtaining the instantaneous frequency components, this method improves upon other methods based on the Fourier transform or wavelet transform by providing much higher resolution and more accurate instantaneous frequencies over time. The core of the HHT is in empirical mode decomposition (EMD), which breaks down a signal into a relatively small number of components which are called intrinsic mode functions (IMFs). Each IMF exhibits a well-behaved Hilbert transform, which accurately represents the frequencies present throughout time. The flow diagram representing the steps to arrive at the Hilbert amplitude spectrum, which is simply a matrix of frequency amplitudes over time, is shown in Figure 3-4.

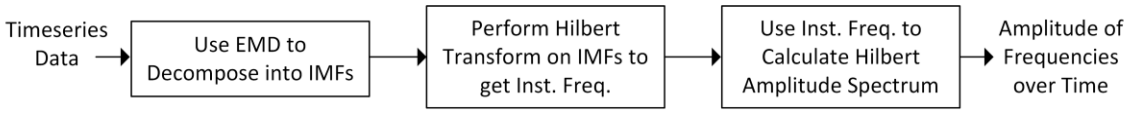


Figure 3-4 Diagram Detailing the Steps to Arrive at the the Hilbert Amplitude Spectrum

To illustrate this process with a reference example, an example discrete time series process will be introduced. The input time series data in Figure 3-4 will be defined as the sum of a stationary signal $X_{st}(t) = \sin\left(2\pi\left(\frac{1}{20}\right)t\right) + 5$, and a popular non-stationary signal, the random walk process $X_{rw}(t)$ defined by $X_{rw}(\mathrm{t}+1) = X_{rw}(t) + N(0,1)$, where $N(0,1)$ is a zero-mean, unit variance Gaussian white-noise process. The random walk

process is initialized with an initial value of zero, and the sampling frequency is 1 Hz. The final resulting signal $X(t)$ is

$$X(t) = \begin{cases} X_{\text{st}}(t) & \text{for } t \in [0,200) \\ X_{\text{st}}(t) + X_{rw}(t) & \text{for } t \in [200,400]. \end{cases} \quad (3\text{-}22)$$

which results in the signal shown in Figure 3-5.

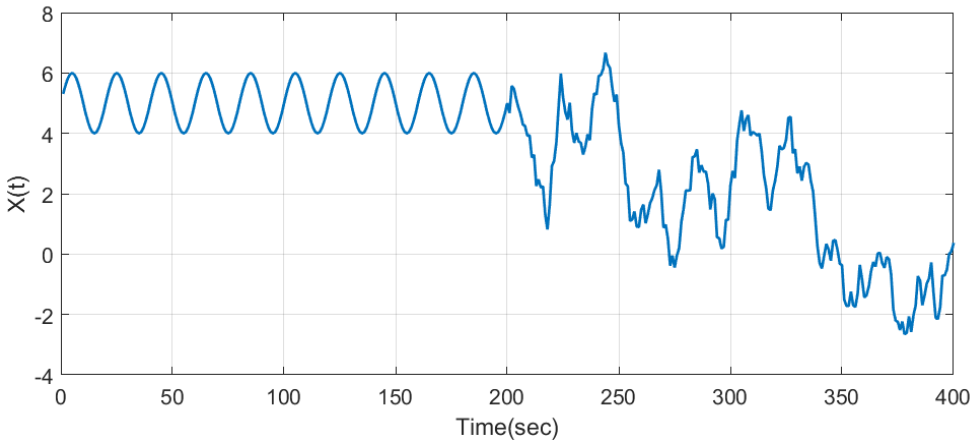


Figure 3-5 Example Time Series Signal

(i) Empirical Mode Decomposition (EMD)

EMD is an iterative, empirical method used to decompose a single signal into a set of oscillatory signals, each with meaningful amplitudes and phases, with the assumption that the process is composed of a group of oscillatory modes [106, 107]. One result of this decomposition is the ability to arrive at a set of instantaneous frequencies over time. Given some signal $f(t)$, the decomposition results in a set of modes

$$f(t) = \sum_{j=1}^{M} \varphi_j(t) \quad (3\text{-}23)$$

where each mode should have the form

$$\varphi(t) = r(t)\sin(\theta(t)). \quad (3\text{-}24)$$

Note that there are infinitely many ways to satisfy (3-23) with a set of modes. Now, however, each mode is restricted to satisfy two conditions: 1) the number of extrema and zero crossings must equal or differ at most by one and 2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. Modes satisfying these criteria are termed intrinsic mode functions (IMFs).

To arrive at a set of IMFs from a signal, an iterative procedure called sifting is applied. First, the signal envelopes must be calculated as a spline fit of the local maxima and local minima, as shown in Figure 3-6 and Figure 3-7. The mean of the envelopes, m_1, is then subtracted from the original signal $f(t)$ to arrive at the signal $h_{1,1}$, that is, $h_{1,1} = f(t) - m_1$, shown in Figure 3-8. The resulting $h_{1,1}$ signal does not yet satisfy the criteria of an IMF, so the process of subtracting the mean of the upper and lower envelopes is repeated to arrive at $h_{1,2} = h_{1,1} - m_2$. The process is repeated k times until the signal meets the criteria of the IMF, and the signal $c_1 = h_{1,k}$ is generated. See Appendix A II for details on the practical implementation on enforcing the IMF criteria.

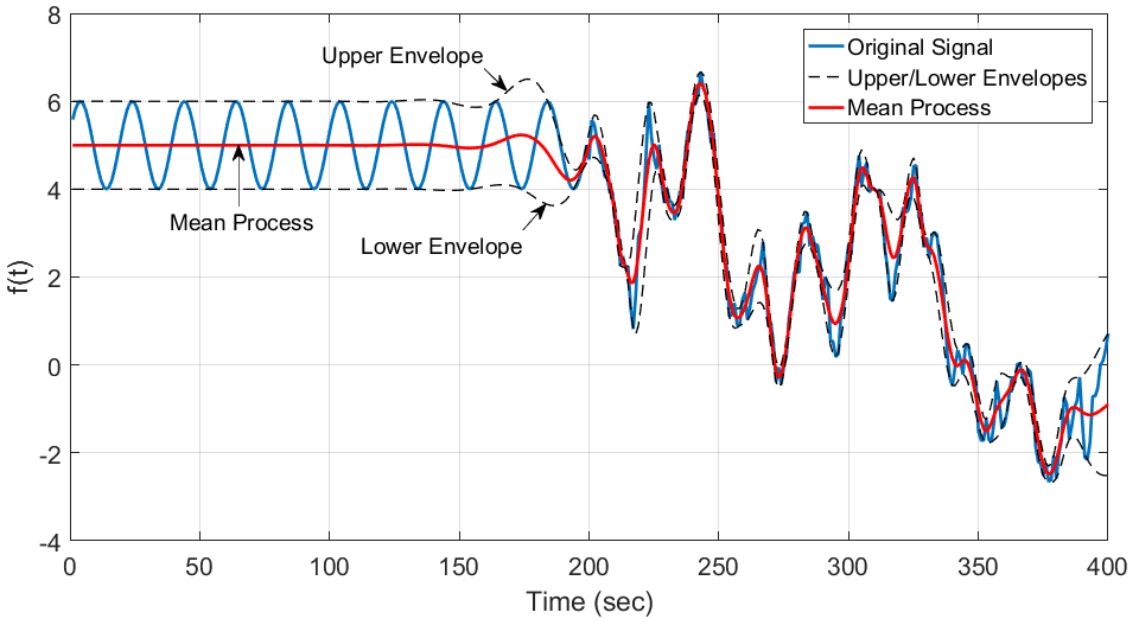


Figure 3-6 Envelopes and Mean Values of Original Signal

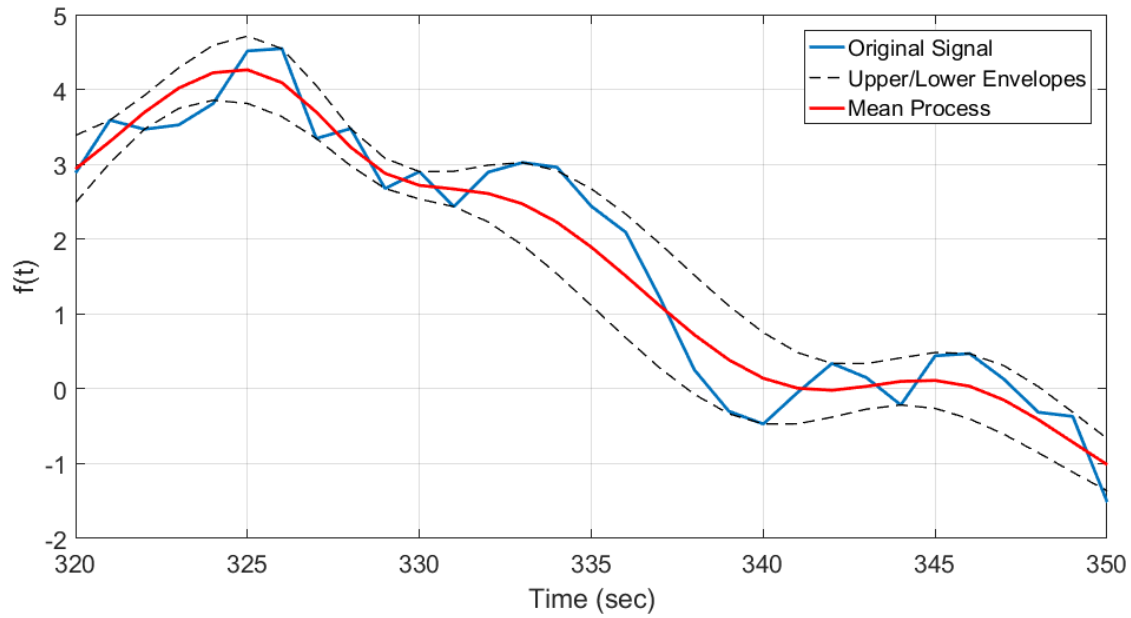


Figure 3-7 Close-up of Envelopes and Mean Values of Original Signal

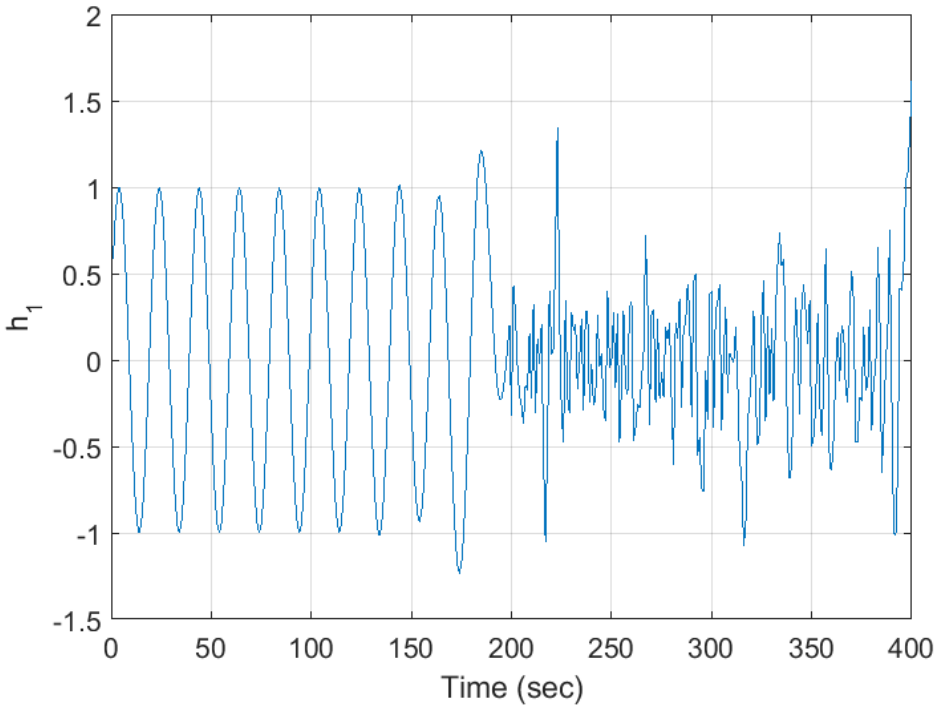


Figure 3-8 Resulting Signal of h_1, after Subtracting m_1

After the first IMF, c_1, is calculated, the new residual signal $r_1 = f(t) - c_1$ is now considered the 'original' data signal on which the sifting process is applied to, such that c_2 is calculated and $r_2 = r_1 - c_2$. The sifting process is repeated M times until the residual r_{M-1} itself is constant or monotonic, such that $f(t) = \sum_{j=1}^{M} c_j$ and the last signal $r_{M-1} = c_M$. Note that the time series c_M is usually not an IMF, and can be considered a trend in the data. The result is a set of modes that are considered complete, in that the sum of the modes recreates the original signal $f(t)$. All of the modes calculated through this process from the example signal are shown in Figure 3-9. The time complexity of EMD is $O(m \log m)$ where m is the number of points, which is equivalent to the complexity of the Fourier Transform [108].

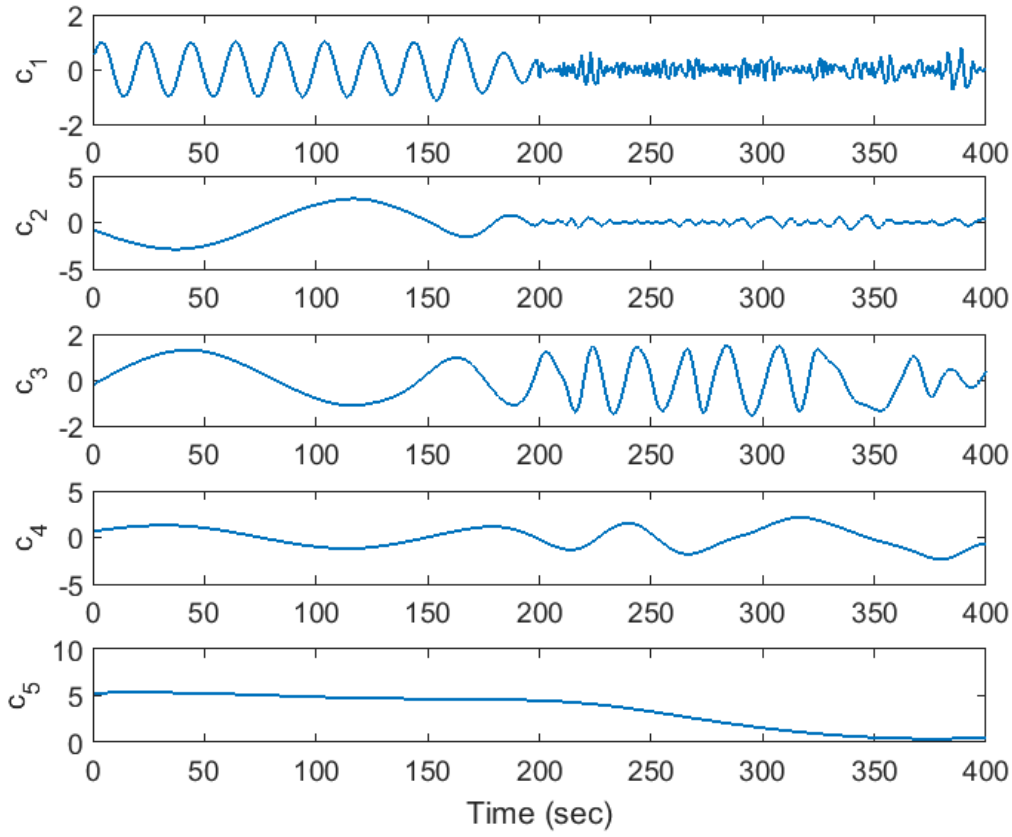


Figure 3-9 All IMFs in Original Signal

(ii) Perform Hilbert Transform on IMFs for Instantaneous Frequency Information

To calculate instantaneous frequencies, the Hilbert transform is applied to each IMF. Let $x(t)$ be a real-valued signal. The Hilbert transform $y(t) = H\{x(t)\}$ is

$$y(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{\tau - t}d\tau \qquad (3\text{-}25)$$

where the Cauchy principal value of the integral is used. This can be thought of as a convolution of the original signal $x(t)$ with the function $\frac{1}{\pi t}$, which heavily weights the local values of $x(t)$ at some time t. A complex analytic signal $z(t)$ is then defined as

$$z(t) = x(t) + jy(t) = a(t)e^{j\theta(t)} \qquad (3\text{-}26)$$

where the amplitude is $a(t) = \sqrt{x(t)^2 + y(t)^2}$ and the phase is $\theta(t) = \tan^{-1}\frac{y(t)}{x(t)}$. The instantaneous frequency can then be defined as

$$\omega(t) = \frac{d\theta(t)}{dt}. \qquad (3\text{-}27)$$

Applying this definition of instantaneous frequency to the analytic signal of each IMF results in a set of instantaneous frequency time series, as shown in Figure 3-10.

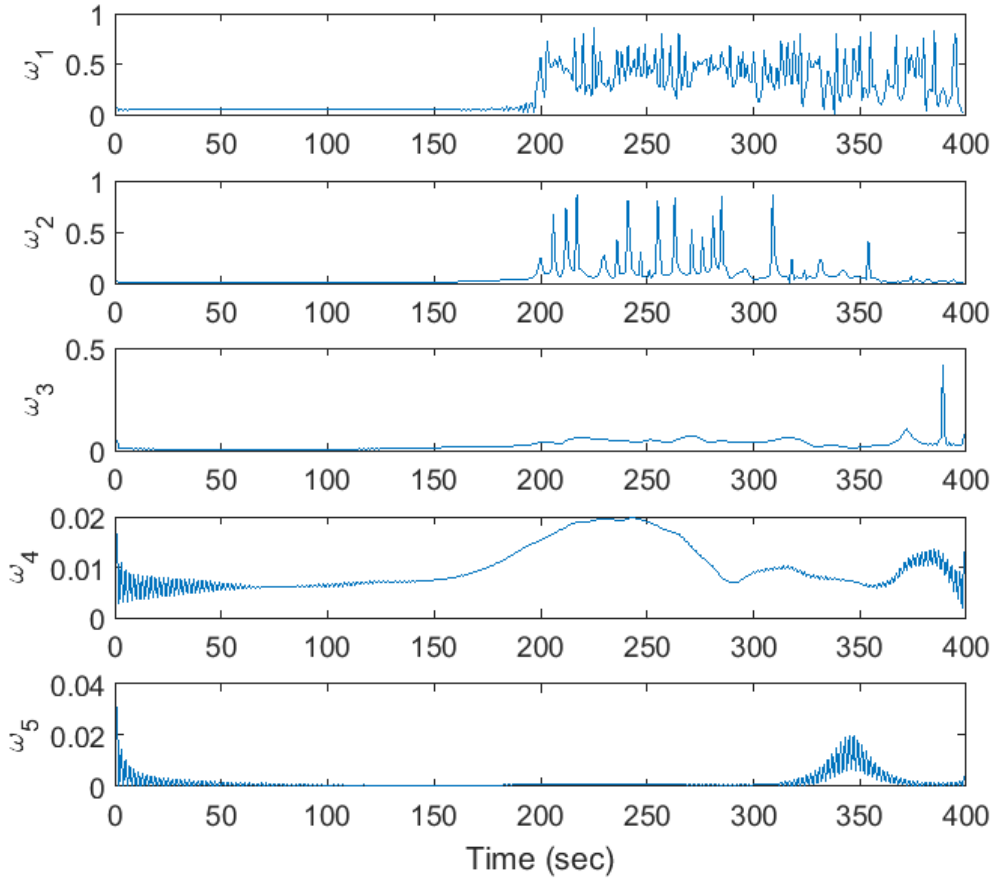


Figure 3-10 Instantaneous Frequencies Present for Each IMF

Before creating the Hilbert amplitude spectrum $H(\omega, t)$, discretization of the frequencies is necessary. If the total data length is T and the sampling rate is Δt, then the lowest frequency that can be extracted is $f_{min} = \frac{1}{T}$ while the highest frequency is $f_{max} =$

$\frac{1}{n\Delta t}$, where n is the minimum number of samples that are needed to accurately represent the frequency. Note that although five points are needed to completely describe a sine wave oscillation, fewer points are needed to define a stable derivative, so this parameter is kept as a user-defined variable. The maximum number of frequency cells that can be used in the discretization is defined as

$$N = \frac{f_{max}}{f_{min}} = \frac{T}{n\Delta t}. \qquad (3\text{-}28)$$

To compose $H(\omega, t)$, which is a matrix of size N x T, the amplitude $a(t)$ is added to the frequency bin that contains its instantaneous frequency $\omega(t)$ for each IMF, effectively summing the amplitudes of frequencies across all IMFs. The Hilbert amplitude spectrum for the example signal is shown in Figure 3-11, with adjacent cells averaged for presentation purposes. Notes on the Hilbert amplitude spectrum are presented in Appendix A III.

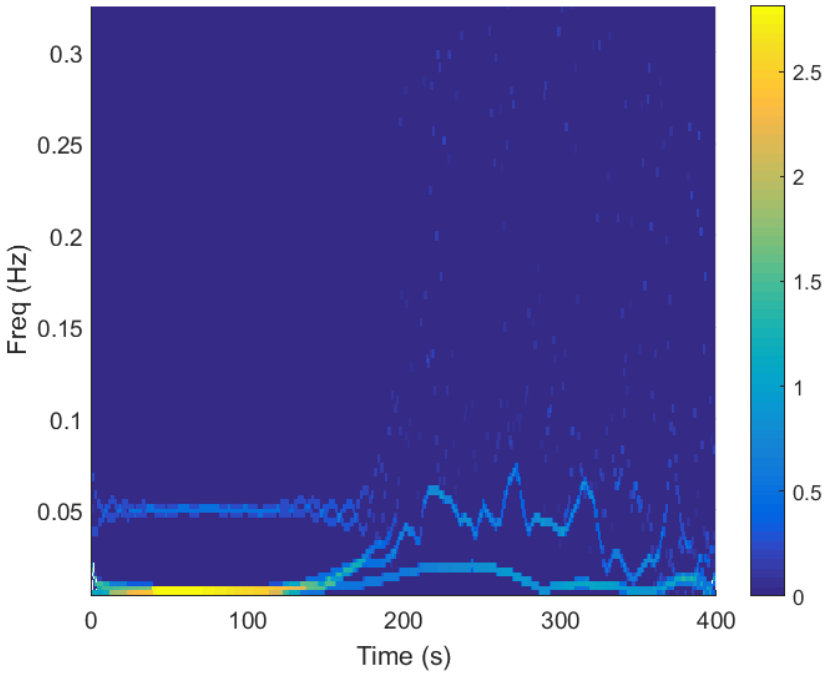


Figure 3-11 Hilbert Amplitude Spectrum for Entire Example Signal

3.3.2. The EDNS Metric

The Ensemble Degree of Non-Stationarity (EDNS) metric is based upon the Degree of Stationarity (DS) definition presented by N. Huang, so the definition of the DS will first be presented [93]. The definition of the DS is

$$\mathrm{DS}(\omega) = \frac{1}{T}\int_0^T \left(1 - \frac{H(\omega,t)}{n(\omega)}\right)^2 dt \qquad (3\text{-}29)$$

where $n(\omega) = \frac{1}{T}\int_0^T H(\omega,t)\,dt$, the mean amplitude for a given frequency, and T is the total length of time series data. This work will consider discrete time series and frequency bins, so the definition of the DS can be rewritten as

$$\mathrm{DS}(\omega) = \frac{1}{T}\sum_{t=0}^{T} \left(1 - \frac{H(\omega,t)}{n(\omega)}\right)^2 \qquad (3\text{-}30)$$

and $n(\omega) = \frac{1}{T}\sum_{t=0}^{T} H(\omega,t)$.

The DS captures the variation of a signal in a certain frequency bin over time. If the DS is large (i.e. the ratio $\frac{H(\omega,t)}{n(\omega)}$ is far from one), this indicates large variations in the bin. If the DS is zero (i.e. the ratio $\frac{H(\omega,t)}{n(\omega)}$ is 1), there is no variation in the frequency bin. Thus, a larger DS is indicative of non-stationary behavior, while a smaller DS indicates more stationary behavior. The DS values for the example signal are shown in Figure 3-12. For a given frequency bin, if there are only very few non-zero instances of the frequency in time, the value of the DS is approximately $1/N$, as seen in multiple frequencies in Figure 3-12. In this work, a DS of 0 is assigned to frequency bins with no energy.

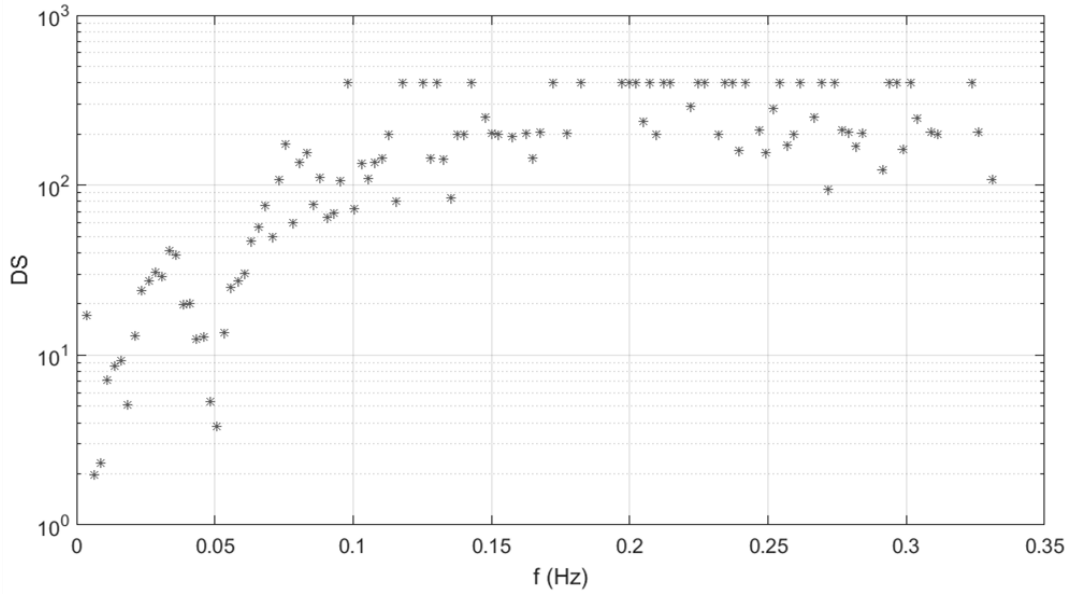


Figure 3-12 DS Values of the Example Signal

The DS is only defined for a given frequency bin, which does not capture the stationarity of the entire time series. To provide a measure for the stationarity for the signal as a whole, the EDNS is introduced as the sum of the DS values across all frequency bins, weighted by the average amplitude in each frequency bin, written as

$$\mathrm{EDNS}_T = \frac{\sum_\omega \mathrm{DS}(\omega) \cdot n(\omega)}{n_{sum}} \qquad (3\text{-}31)$$

where $n_{sum} = \sum_\omega n(\omega)$, and where the subscript in EDNS_T simply denotes that the EDNS was calculated for some length of data T. The EDNS sums the DS values, but weights them by the amplitude of the corresponding oscillation frequency to ensure that small, relatively insignificant oscillations do not dominate the metric. A graph of $\mathrm{DS}(\omega) \cdot n(\omega)$ is shown in Figure 3-13 to contrast with Figure 3-12. The normalization by n_{sum} allows the EDNS to be compared to other EDNS values, regardless of the signal amplitudes. The resulting value

for the EDNS is 15.19 for the example signal, which is the sum of the values in Figure 3-13 divided by n_{sum}.

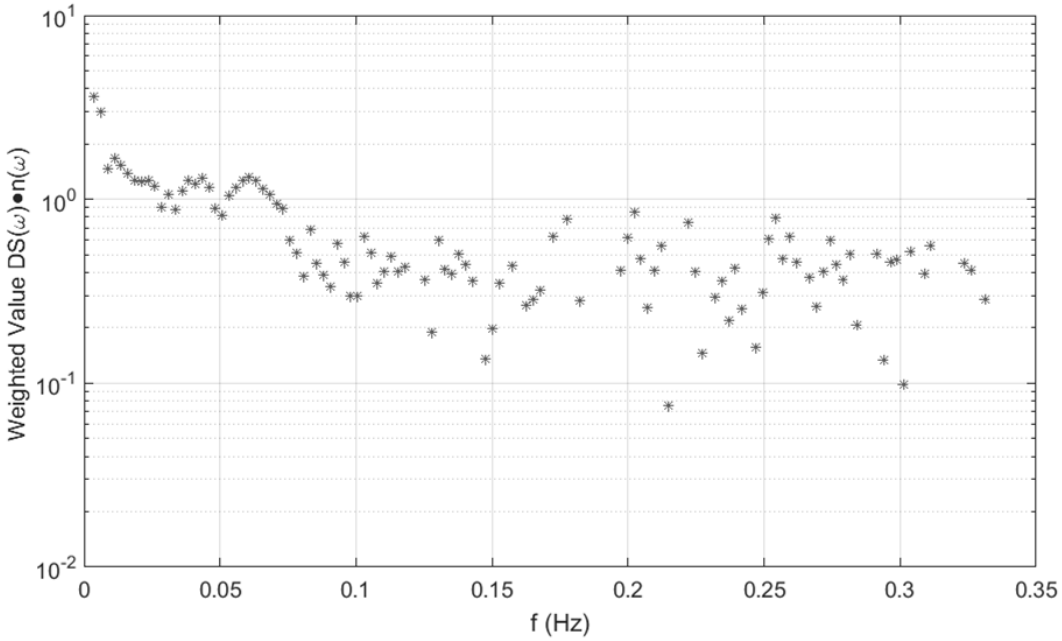


Figure 3-13 Weighted DS Values of the Example Signal

For use in forecasting, the EDNS would be used to determine the non-stationarity of training data subsets in the time series signal. To establish the degree of non-stationarity for different subsets of training data, the EDNS can be calculated on each subset and compared. Thus, for applications in time series forecasting, the EDNS values from different past time series data lengths $T_{min}, \ldots, T_{max}$ can be compared. An EDNS curve is created as a function of the time series data length, defined as

$$\mathrm{EDNS}(T_\ell) = \left.\frac{\sum_\omega \mathrm{DS}(\omega)\cdot n(\omega)}{n_{sum}}\right|_{T_\ell} \qquad (3\text{-}32)$$

where $T_\ell \in T_{min}, \ldots, T_{max}$.

An example of this comparison is shown in Figure 3-14, where EDNS values for data lengths from $T = 20$ seconds to $T = 400$ seconds were calculated in steps of 20 seconds, where the length of data T was taken to start from $t = 0$. As expected, the EDNS is very low for the stationary period from 0 to 200 seconds, but rises rapidly as the non-stationary random walk is included in the data length. This is the result of variations in frequency from the underlying, non-stationary model. It is important to note that the EDNS only shows the degree of non-stationarity of different windows of data and does not determine the optimal training windows alone. Discussion on the use of the EDNS in forecasting is presented in Section 3.4.1.

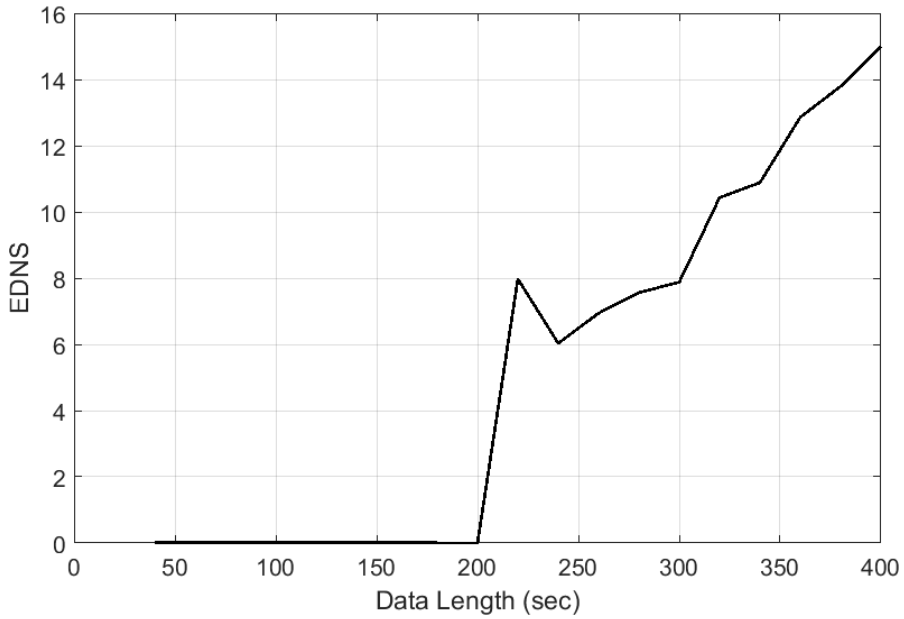


Figure 3-14 EDNS Values for a Range of Dataset Lengths

Note that the computational complexity is now $O\left(\sum_{i=T_{min}}^{T_{max}} T_i \log T_i\right)$, as the EMD method is now repeated over lengths $T_{min}, \ldots, T_{max}$. This can be repeated at every time step (e.g. every 5 minutes) to calculate the most accurate EDNS curve for the most recent data.

To establish this EDNS curve, an iterative calculation over different dataset lengths is summarized in Figure 3-15. This method allows a user to arrive at a curve similar to that shown in Figure 3-14 that relates different lengths of training data to their EDNS values, which can be parallelized to significantly speed computation.

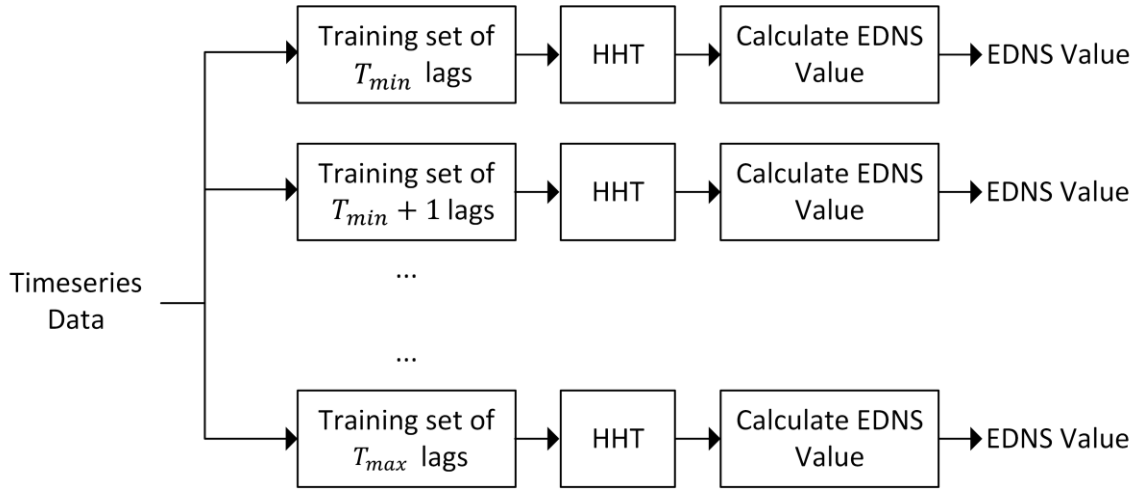


Figure 3-15 Diagram Detailing the Steps to Arrive at a Range of EDNS Values

3.4. Application of the EDNS in Wind Power Forecasting

The EDNS only provides a notion of the non-stationarity within a signal. However, the EDNS can be used in an algorithm to adaptively provide optimal training window sizes in real-time, with the goal of establishing some optimal training window T^* that may change over time. Use of the EDNS in this manner will be termed the dynamic training window method, which will be contrasted with the use of optimal static training windows in wind power forecasting simulations.

3.4.1. Dynamic Training Window Method for Wind Power Forecasting

The strength of the dynamic training window method lies in its ability to adaptively adjust the training window size subject to the stationarity of the past data, with the use of

the EDNS. An outline of the algorithm is shown in Figure 3-16. Each step will be detailed in the following subsections.

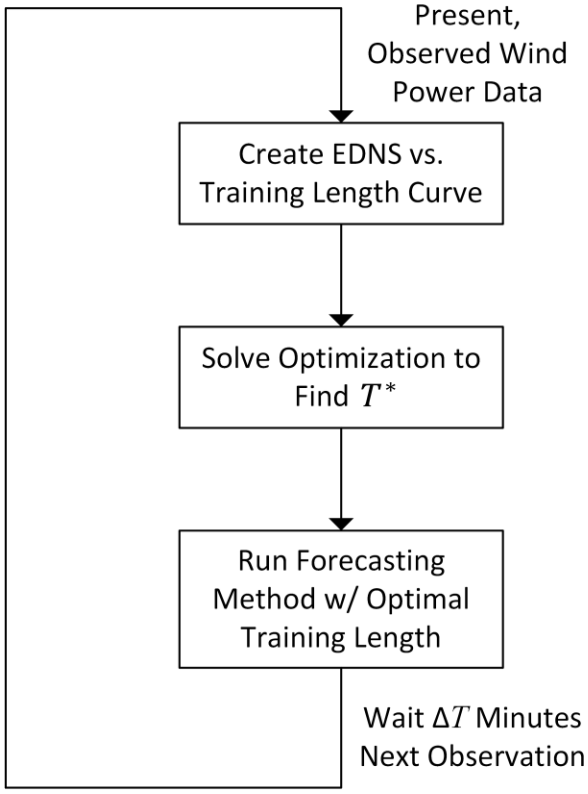


Figure 3-16 Dynamic Training Window Method for Use in Short-term Forecasting

(i) Create EDNS vs. Training Length Curve

After obtaining the present power output of the wind farm time series, the EDNS curve must be calculated for a range of past data $T_{min}, T_{min} + \Delta T_\ell \ldots, T_{max}$ with some step size ΔT_ℓ. The process shown in Figure 3-15 is used to arrive at a curve of the EDNS vs. training length. One consideration is that this curve must include T^*, such that $T_{min} \leq T^* \leq T_{max}$. Though the optimal training windows change over time, in practice, T^* has some finite variance over time specific to a time series (as shown in simulations later in this work), so it is possible to define some static $T_{min}, \ldots, T_{max}$ range such that it contains T^*

for over all time. As a rule of thumb, sweeping from $T_{min} = \Delta t$ to $T_{max} = 2\overline{T}^*$, where Δt is the sampling rate and $\overline{T}^*$ is the mean optimal training window over some representative set of data (e.g. 1 month or over 8000 data points) in a time series works well.

(ii) Determining Optimal Training Length

Forecasting models are built on the assumption that the analyzed length of data is stationary, and so it may seem that the optimal set of training data should be the most stationary one (for example, a set with the lowest EDNS). Intuitively, small sets of data exhibit a low degree of non-stationarity, and taken to the extreme, a dataset consisting of a single point would grant the most stationary process (and indeed, the EDNS of a single point is always 0). However, using a single point as training data is usually undesirable and not implemented in practice for most forecasting methods, as there would be little to no information to build a forecasting model. Figure 3-17 shows the EDNS curves at four randomly selected points in the year for a varying length of potential training windows, which shows that wind power output data generally increases in its EDNS value (and therefore, non-stationarity) over time. The legend shows the day that the training window is calculated backwards from, such that a training window of 48 hours on February 22nd shows the EDNS for a set of data from February 20th to February 22nd.

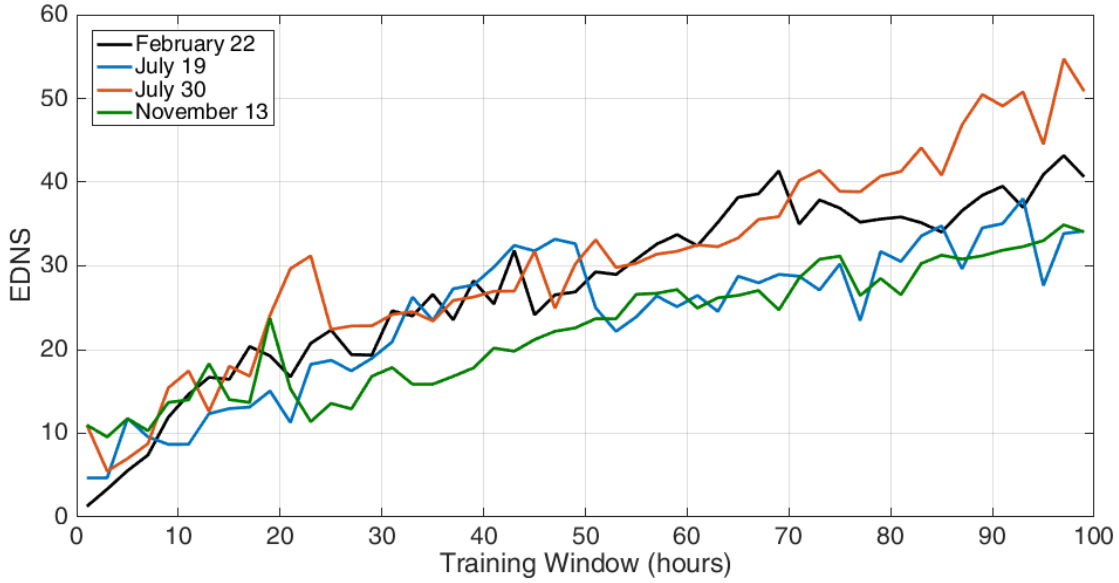


Figure 3-17 EDNS Curves at Four Randomly Selected Points in AEMO Data

There is a fundamental trade-off in time series forecasting between stationarity and the need for a forecasting method to use as much training data as possible. Most forecasting methods tolerate some degree of non-stationarity in producing a useful forecast, but previously, this tolerance was never quantified. This degree of tolerance will be captured by ε, which will treated as a threshold for the EDNS value of the optimal training window. The EDNS vs. training length curve will be used to select an optimal training length T^* at every time step according to the following optimization problem

$$T^* = \quad \text{maximize} \quad T \tag{3-33}$$

$$\text{subject to} \quad \frac{\sum_\omega \mathrm{DS}(\omega)\cdot n(\omega)}{n_{sum}}\Big|_T < \varepsilon \tag{3-34}$$

$$T \in [T_{min}, T_{min} + \Delta T_\ell \dots, T_{max}] \tag{3-35}$$

in which (3-34) provides an ε upper limit for the non-stationarity allowed in the EDNS value for all lengths T, and (3-35) provides the bounded range of training windows. If the problem is infeasible, it is recommended to use the training window $T = T_{min}$ to ensure

that the threshold of ε non-stationarity is met in the training data. This optimization will allow the forecasting method to use as much training data as possible, subject only to a constraint on the EDNS of the training window.

To solve the optimization problem efficiently, the following algorithm was used. Note that the algorithm requires a sweep over the ordered set $\{T_{min}, \dots, T_{max}\}$ such that larger values of T overwrite T^* if the EDNS value is under the ε threshold.

Algorithm 1 Algorithm for determining optimal training window length
Input: $T_{min}, \Delta T_\ell, T_{max}, \mathrm{EDNS}(T_\ell)$
Output: T^*

Initialization: $T^* = 0$
1. **for** $T_i = T_{min}$ to T_{max} in steps of ΔT_ℓ **do**
2. **if** $\mathrm{EDNS}(T_i) < \varepsilon$
3. $T^* = T_i$
4. **end if**
5. **end for**
6. **return** T^*

The selected optimal training lengths points are visually shown in Figure 3-18 given an ε.

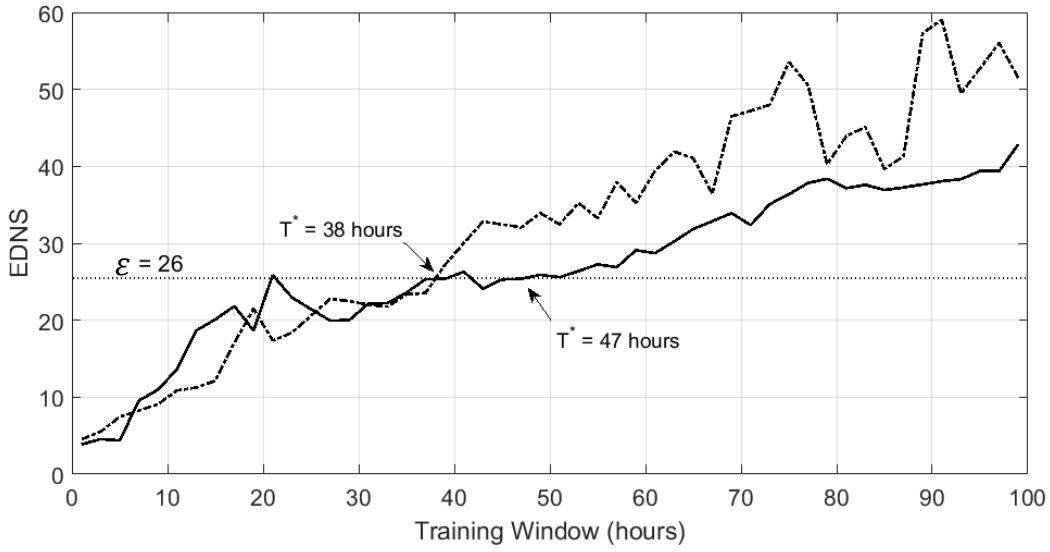


Figure 3-18 Optimal Windows Given Some ε at Two Points in AEMO Data

The value of ε is determined in this study by two methods: 1) creation from as much prior data as possible, termed the rigid epsilon approach and 2) creation from some set of most recent data, termed the adaptive epsilon approach. In the rigid epsilon approach, the value of epsilon may be found simply by sweeping epsilon values over lengths of training data to find the best epsilon value for a forecasting method (which in this study is one year of data). The epsilon value is only calculated once in the rigid epsilon approach and does not change. However, the adaptive epsilon approach acknowledges that the EDNS is based off of a noisy Hilbert amplitude spectrum, which does not always represent the frequencies present in the signal accurately due to the empirical nature of the EMD method. The resulting EDNS values then do not perfectly capture the non-stationarity in the time series, and so the optimal ε threshold may change in practice over time for a single forecasting method. In the adaptive epsilon approach, a smaller, more recent set of data is used to determine ε by sweeping over a range of $[\varepsilon - \varepsilon_{sw}, \varepsilon + \varepsilon_{sw}]$ thresholds and evaluating performance, where ε_{sw} defines the swept range around the original ε. This is performed iteratively throughout time, and it has been found that changing the ε threshold every month, based on the best ε of the past month of data, has shown good results. This monthly update of ε is used in this study for the adaptive epsilon approach.

(iii) Run Forecasting Method w/ Training Length

After selecting the optimal set of training data, the forecasting method is simply run for one time step with this set of training data. The process of selecting a training window T^* can then be repeated at the next time step. If the data has relatively low EDNS values in general, then it is possible to run the forecasting method for multiple time steps with the

same T^* training window to limit computational effort in arriving at the EDNS values for different training window lengths.

3.4.2. Simulations using the EDNS

This real-time algorithm was tested with SVR models because better forecasting performance was seen with SVR than with AR on the wind power output time series, and SVR also used less training data to achieve the results. This resulted in drastically faster computation for both the creation of the EDNS curves and the forecasting itself. The features of the SVR model were defined to be the $p = 4$ past power output measurements, which corresponds to the results of the AIC and BIC methods, and the label was defined to be the difference between the past and present power output. As a result, the matrix of training data was of dimension $(T^* - 3)$ x 4, while the label vector was of dimension $(T^* - 3)$ x 1 for the creation of an SVR model. A new SVR model was created at every 5-minute time step based on the most recent data of length T^*, which was contrasted with an SVR created at every 5-minute time step based on a static training window T_{st}^*, the optimal static training window. MATLAB 9.0 was used for these simulations in conjunction with the *LIBSVM* library [104]. In this work, $C = \max y - \min y = 1$ after data normalization, and a grid search is performed over $\varepsilon = 0{,}0.01, \ldots 0.05$ and $\gamma = 2^{-2}, 2^{-1}, \ldots, 2^{7}$. A 5-fold cross validation was performed at every time step to optimize the SVR parameters.

(i) Simulations using the EDNS with a Single Wind Farm in AEMO

The AEMO dataset was used for the first set of simulations, as it is a highly non-stationary dataset with statistical forecasting methods that performed worse than persistence in published literature [82]. Five forecasting approaches were applied to the

dataset: 1) persistence, 2) static training window SVR, in which the optimal training window from one year was used for the next, 3) rigid ε dynamic training window SVR, in which the optimal ε from one year was used for the next, 4) monthly static training window SVR, in which the most recent month was iteratively used to establish the optimal static training window, and 5) adaptive ε dynamic training window SVR, in which the optimal ε was computed iteratively at the beginning of each month, based upon results from the last month.

Two years (2012-2013) of a power output time series with a 5-minute resolution from a 160 MW wind farm in AEMO were used in this forecasting study. One-step-ahead forecasting tests were performed on the complete year of 2013. The training window of 67 hours was identified as the optimal static training window in 2012 by sweeping over values of training windows and identifying the one that yielded the lowest forecasting errors. For the rigid epsilon method, the EDNS value of $\varepsilon = 27$ was chosen as the best performing epsilon value for the year of 2012. The adaptive epsilon method used a range of $\varepsilon \in [19{,}32]$ at a resolution of 0.25, where each month the best ε value from the prior month was chosen for forecasting. Training windows and chosen epsilon values over each month of the test year are shown in Figure 3-19.

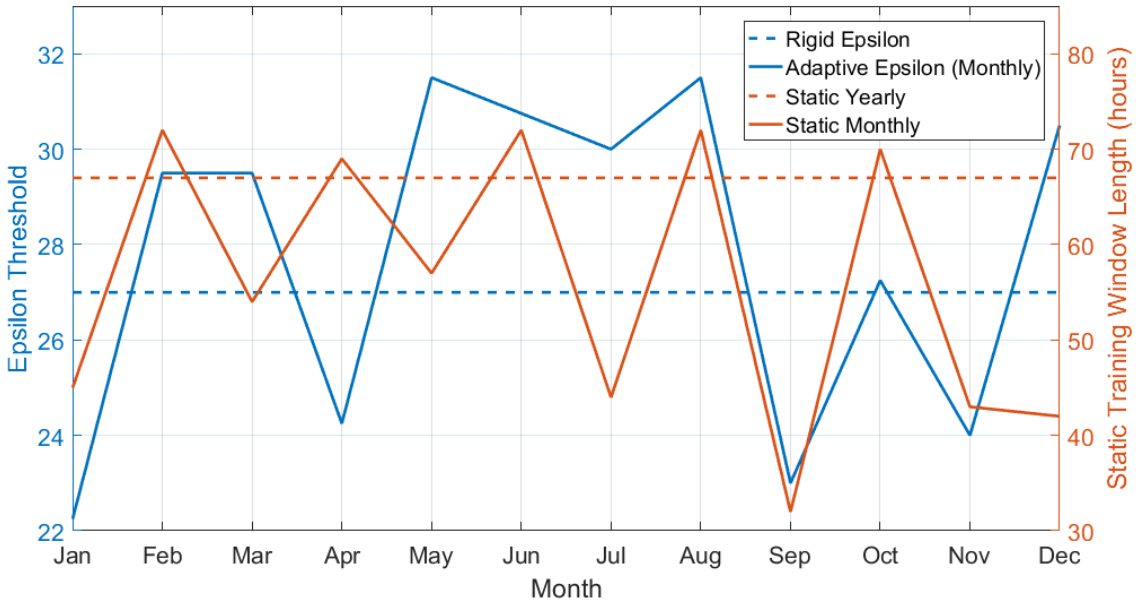


Figure 3-19 Training Windows or Epsilon Values over the Test Year of 2013

Table 3-1 Forecasting Results for AEMO Wind Farm, 2012

Forecasting Method	MAE (%)	MAE (MW)	RMSE (%)	RMSE (MW)	Ramp MAE (%)	Ramp MAE (MW)
Persistence	2.015	3.224	3.566	5.706	7.830	12.53
Static 67-hour Training Window SVR	2.003	3.205	3.513	5.621	7.513	12.02
Rigid $\varepsilon = 27$ Dynamic Training Window SVR	1.998	3.197	3.507	5.611	7.490	11.98
Monthly Static Training Window SVR	1.993	3.189	3.507	5.611	7.520	12.03
Adaptive Epsilon, Dynamic Training Window SVR	1.990	3.184	3.506	5.610	7.501	12.00

Forecasting results on the entire year of 2013 data are presented in Table 3-1, showing that the use of a dynamic training window results in improved results for both the complete year and just in the presence of ramps, where a ramp is defined to be a change in power of greater than 7.5 MW in 5 minutes (which is 5% of the wind farm capacity). When compared with the static training window methods, the forecasting improvement using the dynamic training window decreases the MAE by 0.2%-0.3%, which is close to expected as discussed in Section 3.1.1. In this dataset, recalculating the optimal training windows (both

for epsilon and for the static training window size) results in lower error for the entire year, compared with the case in which an entire year was used to calculate a single epsilon or training window size. Notice that no special attention has been paid to ramp identification or prediction, and that only the overall MAE was minimized. This is seen in the monthly training cases, where the ramp MAE is higher in the monthly re-calculated cases when compared with the yearly cases. Ramp prediction can be improved with the implementation of a ramp detection algorithm, such as the one in the work by L. Yang et al. [85].

(ii) Simulations using the EDNS with a Wind Farm in Colorado

The dynamic window algorithm may show improved results over use of a static training window when a highly non-stationary process is forecasted, but if the time series itself is highly stationary, the dynamic window algorithm may result in the same performance as when a static training window is used. Forecasting performance was compared using a single 10-minute resolution time series from a 300.5 MW wind farm in Colorado, with two full years of data from 2011-2012, sourced from NREL. The Colorado data has a more stationary time series than in the AEMO data, as the EDNS curves for the Colorado data have lower EDNS values than the AEMO data, as shown in Figure 3-20. In addition, AR forecasts were performed on two months of Colorado data with the same model and procedure as shown Section 3.1.1, with a 4-lag AR, one-step-ahead forecasting model. The MAE vs. AR model training length plot is seen in Figure 3-21, which shows that for extremely long training window lengths (up to half of a year for the AR model), the performance of the AR forecast does not degrade. This can compared with Figure 3-1 and Figure 3-2, where the forecast performance does degrade with long training window lengths.

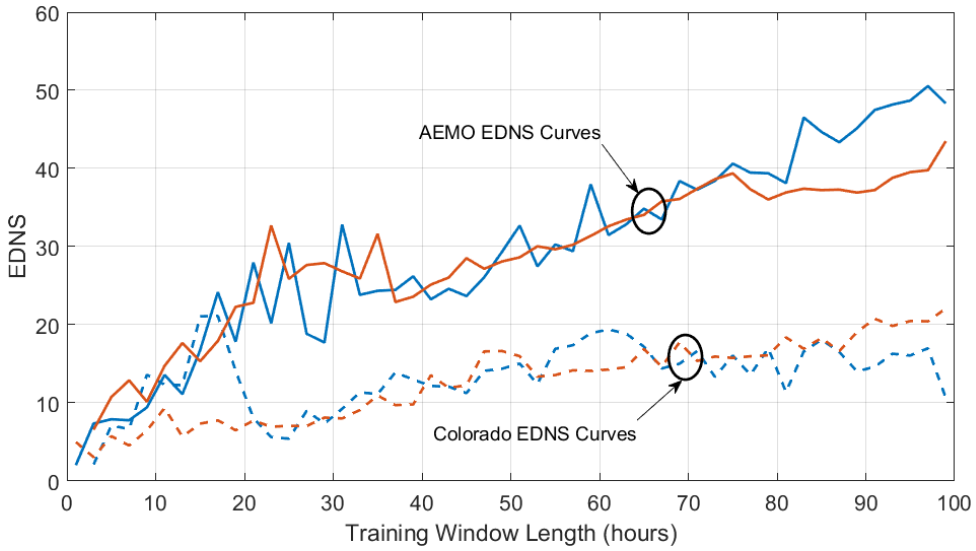


Figure 3-20 Comparison of Two Colorado EDNS Curves and Two AEMO EDNS Curves

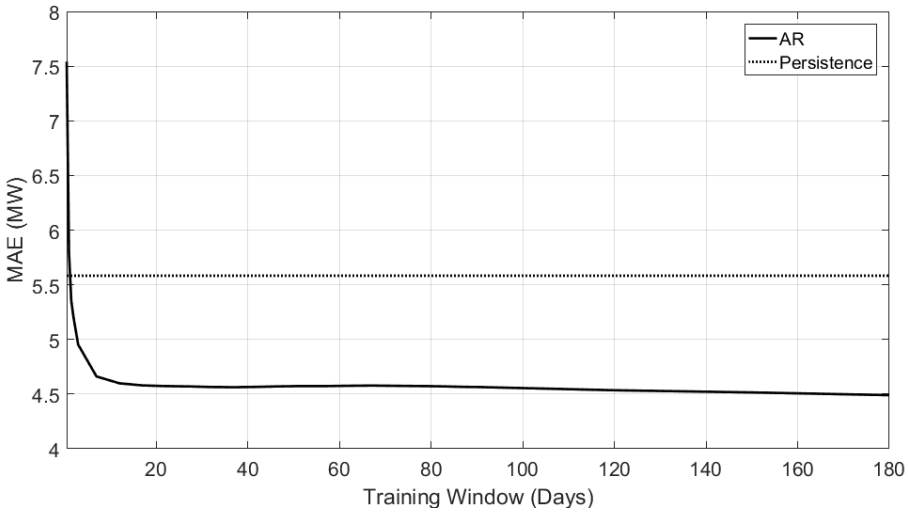


Figure 3-21 AR MAE Error Versus Training Data Length for 2 Months of Colorado Data

SVR forecasting models with a feature length of 4 were also created using the same procedure as with the AEMO data, and the SVR forecasting error versus epsilon for the year of data is shown in Figure 3-22, with the use of the dynamic training windows. The curve shows that SVR performance does not degrade with long training lengths. An epsilon of 23 was selected for the year of data, and results for the rigid epsilon approach of the

dynamic training window algorithm are shown in Table 3-2. Similar results between the use of a dynamic training window and static training window are seen in the table, where a ramp was defined as 5% (15 MW) of the wind farm capacity.

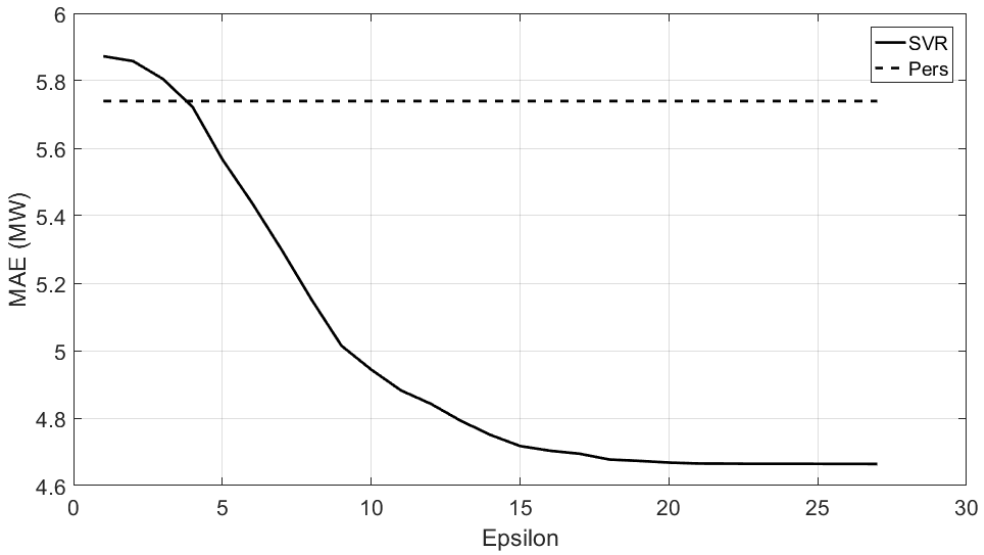


Figure 3-22 SVR MAE Error versus Epsilon for a Year of Colorado Data, 2011

Table 3-2 Forecasting Results for Colorado Wind Farm, 2012

Forecasting Method	MAE (%)	MAE (MW)	RMSE (%)	RMSE (MW)	Ramp MAE (%)	Ramp MAE (MW)
Persistence	1.809	5.436	3.351	10.07	8.593	25.82
Static 121-hour Training Window SVR	1.500	4.506	2.840	8.533	5.711	17.16
Rigid $\varepsilon = 27$ Dynamic Training Window SVR	1.500	4.506	2.839	8.530	5.711	17.16

To examine the ability of the adaptive epsilon algorithm to correct for poor initial choices of epsilon, an epsilon value of 10 was initially selected for the month-to-month adaptive epsilon and contrasted with an rigid epsilon approach with $\varepsilon = 10$ for the year of Colorado data. For this simulation, the range over which surrounding epsilons were tested each month in the range $[\varepsilon - \varepsilon_{sw}, \varepsilon + \varepsilon_{sw}]$ was set to $\varepsilon_{sw} = 2.5$, with a sweep resolution of 0.5. This allowed for a maximum change of 2.5 in the epsilon value each month. The

movement of epsilon over time is shown in Figure 3-23, which shows movement to the optimal epsilon value of 27. Table 3-3 shows that the adaptive epsilon approach improves upon the static epsilon case with a poor initialization of epsilon.

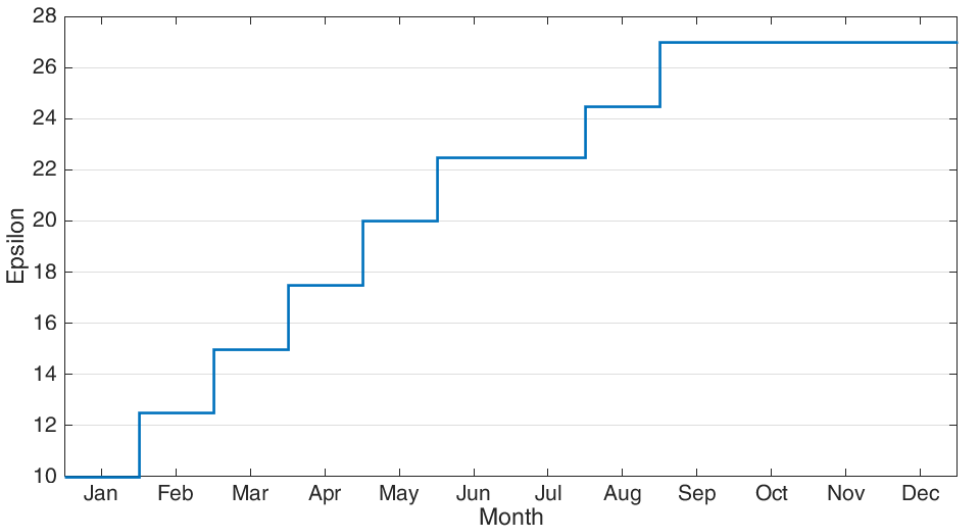


Figure 3-23 Epsilon Movement with the Adaptive Epsilon Approach, Over Each Month

Table 3-3 Rigid vs. Adaptive Epsilon Approaches with Poor Initialization of Epsilon

Forecasting Method	MAE (%)	MAE (MW)	RMSE (%)	RMSE (MW)	Ramp MAE (%)	Ramp MAE (MW)
Rigid $\varepsilon = 10$ Dynamic Training Window SVR	1.627	4.890	3.055	9.180	6.557	19.70
Adaptive Epsilon, Dynamic Training Window SVR	1.515	4.552	2.908	8.738	5.966	17.93

(iii) Simulations using the EDNS Utilizing Spatio-Temporal Information for Wind Farm Forecast in AEMO

Performance of the dynamic window algorithm was also tested in a multi-wind farm forecast case using SVR, as spatio-temporal forecasts are becoming increasingly common in published literature regarding wind power forecasting [31, 82]. Forecasting tests were performed on the same single power output time series from Subsection (i), a 5-minute resolution time series from a 160 MW wind farm in AEMO (referred to as the main

wind farm), also using the power output a nearby wind farm (referred to as the auxiliary wind farm) less than 10 km away to assist in the forecast of the main wind farm. There were twice the number of features used versus the SVR forecasting case with only one wind farm, which were the $p = 4$ past power output measurements of both the main and auxiliary wind farms. The label definition remained as the difference between the past and present power output of the main wind farm. As a result, the matrix of training data was of dimension $(T^* - 3)$ x 8, while the label vector was of dimension $(T^* - 3)$ x 1 for the creation of an SVR model.

The same epsilon was used for both wind farms, as the EDNS curves were fairly similar as shown in Figure 3-24. The dates in the legend are the dates for which the training length was calculated backward from, such that a training window length of 72 hours from September 30th corresponds to a training window of the first hour of September 27th to the first hour of September 30th. The optimal training window T^* was found for both the auxiliary wind farm and the main wind farm at every time step. This resulted in two T^* values, T^*_{aux} for the auxiliary wind farm and T^*_{main} for the main wind farm, so the simple minimum among the two T^* values was used as the global T^* to ensure that both training sets were suitably stationary, such that

$$T^* = \min(T^*_{main}, T^*_{aux}). \quad (3\text{-}36)$$

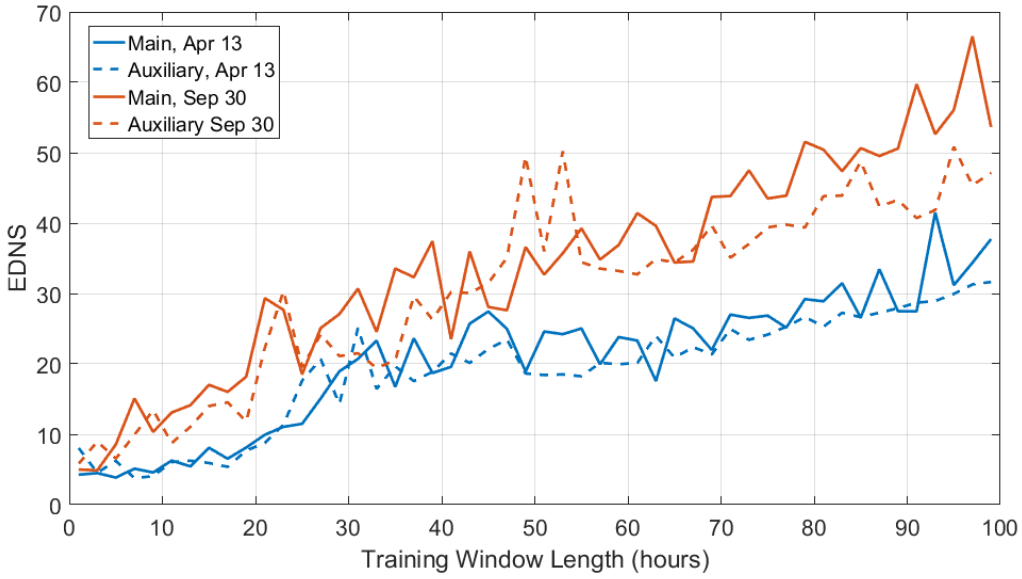


Figure 3-24 EDNS Curves at Two Times for Both the Main and Auxiliary Wind Farms

Forecasting results are shown in Table 3-4, showing that the use of a dynamic training window results in slightly improved results for overall performance. The small improvements in performance in this dual wind farm approach (less than 0.1%) as opposed to the single wind farm approaches may be due to the fact that any changes in correlation between the wind farms are ignored, as the present algorithm only focuses on the stationarity of an individual time series.

Table 3-4 Forecasting Results Using Spatio-temporal Information in AEMO, 2013

Forecasting Method	MAE (%)	MAE (MW)	RMSE (%)	RMSE (MW)	Ramp MAE (%)	Ramp MAE (MW)
Persistence	2.015	3.224	3.566	5.706	7.830	12.53
Static 59-hour Training Window SVR	1.970	3.151	3.460	5.537	7.226	11.56
Rigid $\varepsilon = 26$ Dynamic Training Window SVR	1.968	3.148	3.459	5.534	7.210	11.54
Monthly Static Training Window SVR	1.968	3.149	3.461	5.538	7.224	11.56
Adaptive Epsilon, Dynamic Training Window SVR	1.967	3.147	3.458	5.532	7.207	11.53

3.5. Conclusions on the EDNS

This work introduced the EDNS, which is a metric that quantifies the degree of non-stationarity present in a time series. In addition, a real-time dynamic training window algorithm was introduced that used the EDNS, which was able to determine an appropriately stationary set of training data for SVR short-term wind power forecasting. The dynamic training window algorithm, with the assistance of the EDNS, was able to provide an estimate of the optimal training window length based on the quantified non-stationarity in the past data. When using the real-time dynamic training window algorithm, improvements of 0.2% to 0.3% were seen when compared with the optimal static training window selection through its application to a highly non-stationary dataset, AEMO data. The adaptive epsilon approach in the dynamic training window algorithm saw further performance improvements, especially if the training window or epsilon were initialized far from the optimal epsilon. In a more stationary dataset from a wind farm in Colorado, it was seen that the use of the dynamic training windows did not improve performance, as very long training windows did not degrade performance and did not need to change over time. Further research directions with the EDNS are identified in the final chapter of the dissertation.

Quantifying stationarity may increase the accuracy of short-term forecasts in highly non-stationary datasets, which can then be used in various power system operation tools to improve operations performance in the presence of high penetrations of renewable energy. One such application could be in the use of security constrained economic dispatch, in which dispatch occurs at 5-minute intervals in some modern power systems such as those in controlled by PJM Interconnection [109]. However, these applications exclude the

proposed anticipatory control improvement to secondary control presented in the next chapter, as secondary control operates on a sub-minute time scale and thus would likely use persistence forecasts for useful short-term forecasts (e.g. 30-second-ahead forecasts) because of the small errors expected from persistence forecasts over such small time intervals. The next chapter presents an improvement to secondary control with the use of a disturbance forecast provided approximately 30 seconds in advance of the disturbance, which is separate from the work performed in this chapter.

4. ANTICIPATORY SECONDARY CONTROL

4.1. Introduction to Anticipatory Secondary Control

In power systems, secondary control is used to continuously balance supply with demand to maintain a stable and reliable service to the consumers. Automatic Generation Control (AGC) provides secondary control, sending commands to a set of responsive generators to change generation levels and minimize load and generation imbalances, which ultimately correct for deviations in system frequency or tie-line power flow. The fundamental operation of AGC as a feedback controller responding to deviations in measured frequency or tie-line flow has changed little over its industry adoption [110, 111, 112, 113]. However, as significant amounts of renewable energy continues to be integrated into modern power systems, extra uncertainty in the generation and load balance has been introduced which can dwarf the short-term load forecasting errors as a percentage error of generation [31, 114, 115]. Also, with increased integration of renewable generation, ramp rates in generation can be much larger than previously experienced in a power system, so spinning or non-spinning reserve on-hand may not be able to follow these ramps as tightly as desired [39, 116]. Prediction accuracy of renewable generation is improving, as much research is focused on short-term renewable energy forecasts, including both point and distributional forecasts, which can predict future disturbances in renewable energy with

increasing accuracy [31, 85]. In addition, predictable generation or load events, including large social events and some transmission outages, can also result in sudden and relatively large generation and load imbalances that can be anticipated to a high degree of accuracy.

When a disturbance is anticipated, it may be reasonable to manually raise or lower the frequency area reference before the disturbance, in anticipation of the sudden generation and load imbalance. However, manual frequency set-point changes require human intervention and may have limited or no feedback response to allow for control corrections to an imperfect forecast of an anticipated event. An automated, anticipatory control solution would save time for system operators and improve system operational performance by responding to forecasts of large disturbances before the disturbance, and by correcting for imperfections in the control signals with feedback control. In this work, prior knowledge of an event in the power system will be used in an anticipatory controller to improve operational performance in a simulated power system by adjusting the system frequency reference as an external controller to the existing AGC system. This work will only consider anticipated generation or load imbalances that can be predicted to some degree of accuracy in the very short term (< 5 minutes), though a short study will be performed using a 5-minute-ahead forecast. The main contributions of this work are:

1) *Use of a (possibly imperfect) disturbance forecast in secondary control*: The integration of anticipated events or generation forecasts into the secondary control loop greatly improves the ability of the power system to respond to forecasted changes in generation or load as compared to conventional AGC.

2) *Design of the controller as an add-on module*: Much existing literature on novel secondary control methods involves the partial or complete replacement of traditional

AGC, which may not be practical. Designing the MPC controller as an add-on module preserves the existing AGC system architecture, allowing for a plug-and-play operation. In addition, this allows the operator to easily weight the decision of the controller with traditional AGC action or preplanned actions if desired.

3) *Highlighting trade-offs between inter-area communication of forecasted disturbances*: Communication of data over great distances is becoming more common, so system operators between different control areas may be linked by communication channels in order to increase operational performance for all areas. However, forecasts are frequently incorrect, and depending on the magnitude of the errors, broadcasting an incorrect forecast may result in worse system performance than in the case of no broadcasting.

The work is organized as follows. Section 4.1 includes the motivation for the work as well as some background on secondary control in the power system. Section 4.2 introduces the general MPC model that will be used for anticipatory secondary control. Section 4.3 introduces the power system model for testing and the selection of the parameters for the controller. Section 4.4 presents the detailed MPC formulation used in this study along with simulations and the comparisons of the anticipatory secondary controller with other control methods. Conclusions are presented in Section 4.5.

4.1.1. Secondary Control in the Power System

Secondary control refers to generation and load balancing services that control area typically provides within a few minutes, restoring frequency and tie-line flows to its nominal values. The core of secondary control is automatic generation control (AGC), which is a control scheme that uses measurements from the Supervisory Control and Data

Acquisition (SCADA) system in an electric system to respond to changes in measured frequency and tie-line flows. AGC has two main objectives: 1) to hold the system frequency close to a specified nominal value and 2) to maintain the correct interchanges between control areas subject to contracts and transmission line constraints. These objectives are captured by minimizing the ACE, defined as

$$\mathrm{ACE} = \Delta P_{tie} - 10\mathrm{B}\Delta f = \left(P_{tie} - P_{tie\_ref}\right) - 10\mathrm{B}\left(f_{sys} - f_{ref}\right) \qquad (4\text{-}1)$$

where ΔP_{tie} is the difference between the net actual interchange P_{tie} and the net scheduled interchange P_{tie_ref} (excess power flow out of the area is defined to be positive), B is the frequency bias factor in units of MW/0.1 Hz (the value of B is negative), and Δf is the deviation of actual frequency f_{sys} from the area frequency reference f_{ref} in Hz (the value is negative when the area is below its reference frequency). A negative value of ACE means that the control area, or balancing area (BA), should be generating more power to return the ACE value to zero. AGC minimizes the ACE value with a proportional-integral (PI) controller that sends signals to responsive generation, used for its simplicity and ability to eliminate steady-state errors [117].

In North American power systems, the frequency bias factor B is typically close to the frequency response of the system, which is roughly the change in power (in MW) over ten times the change in frequency (in Hz) given some event. For example, a loss of around 1000 MW of generation in the Electric Reliability Council of Texas (ERCOT) resulted in a drop of around 0.17 Hz in frequency, and a loss of 150-300 MW in Bonneville Power Administration (BPA) resulted in a drop of 0.1 Hz. Based on these two events, the frequency responses of the ERCOT and BPA systems are approximately 588 MW/0.1Hz and 150-300 MW/0.1 Hz, respectively [118, 119]. The precise calculation of the power

system frequency response involves multiple measurements to compose an appropriate value, and more details are available in the references. Further details of secondary control, especially details regarding its implementation within the North American interconnected power system, is available in a North American Electric Reliability Corporation (NERC) training document [120]. In this work, B is calculated using the droop and load damping factors of the system, discussed in Section 4.3.2.

A high-level block diagram of the AGC system and its relationship to the control area is shown in Figure 4-1, where ω_{set} is the vector of governor frequency set points for responsive generation in the power system. The measurements P_{tie} and f_{sys} are sampled and sent through the SCADA system to the central controller for the control area, AGC. ACE is calculated within the AGC block, and the vector of governor frequency set points is then adjusted accordingly to minimize ACE. The scheduled power flow P_{tie_ref} and the reference frequency f_{ref} is not represented in this diagram, but note that the reference value of P_{tie_ref} and f_{ref} is contained within the AGC block as a constant reference. Detailed implementation of the blocks in Figure 4-1 is discussed in the next section.

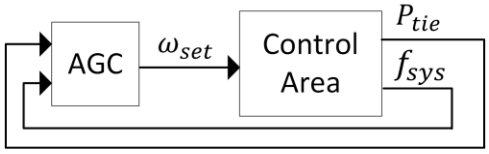


Figure 4-1 Information Flow Between the AGC System and the Power System

Traditional AGC does not take advantage of prior knowledge of disturbances because it only reacts to deviations in frequency or scheduled tie-line flows. Especially with the enforcement of the newer CPS1 and CPS2 standards, interconnections must strive to minimize large ACE values when disturbances occur in the power system, so it may be

desirable to act before a disturbance occurs to minimize frequency or tie-line deviations [111, 120, 121]. Also, in cases of extreme changes in the generation and load balance, traditional AGC may not be able to react as quickly as desired, so a modification of traditional AGC is of interest [39, 116]. Anticipatory secondary control is introduced in this work, which is a novel secondary control scheme that will react prior to a disturbance, using MPC to optimally calculate control actions for the area.

4.2. Anticipatory Secondary Control Design

A new method of secondary control, termed anticipatory secondary control can be used to minimize frequency and tie-line deviations with feed-forward control in the presence of anticipated disturbances. In this section, model predictive control (MPC) is introduced as the optimal control framework by which anticipatory secondary control is implemented. The integration of MPC with the power system will be discussed in a general sense as well, with details of the formulation presented in Section 4.4.1, after the power system model is introduced.

To implement anticipatory secondary control, an external predictive controller is used with a disturbance forecast time series input and various measured inputs from the power system to change the area frequency reference, which modifies the traditional ACE value to that of an effective ACE (eACE). This eACE is not the true ACE of the system, but is nevertheless minimized by the existing conventional integral controller in the AGC system. This anticipatory approach to secondary control utilizes externally-provided event forecasts in the power system's secondary control loop, thus significantly improving the power system's ability to respond to large changes in generation or load fluctuations. The

proposed approach also allows for seamless integration into the existing power system control architecture, as the existing control infrastructure can be efficiently augmented with the envisioned controller, as shown in Figure 4-2. The MPC controller uses a forecasted disturbance estimate $\hat{d}$ within its own area as well as estimates of the future disturbances in other areas $\{\hat{d}^a\}$ (if available), measured system variables such as system frequency f_{sys}, the net inter-area tie-line flow P_{tie}, the effective area control error eACE, and the set of area frequency references $\{f_{set}^a\}$ from the other a areas in the power system interconnection to provide a frequency area reference f_{set} for AGC. The existing AGC system then uses the modified f_{set} to send out a vector of governor speed set-points ω_{set} to the responsive generators in the system. Note that the actual disturbance d to the power system differ from the forecasted disturbance $\hat{d}$ by some forecast error ε_d.

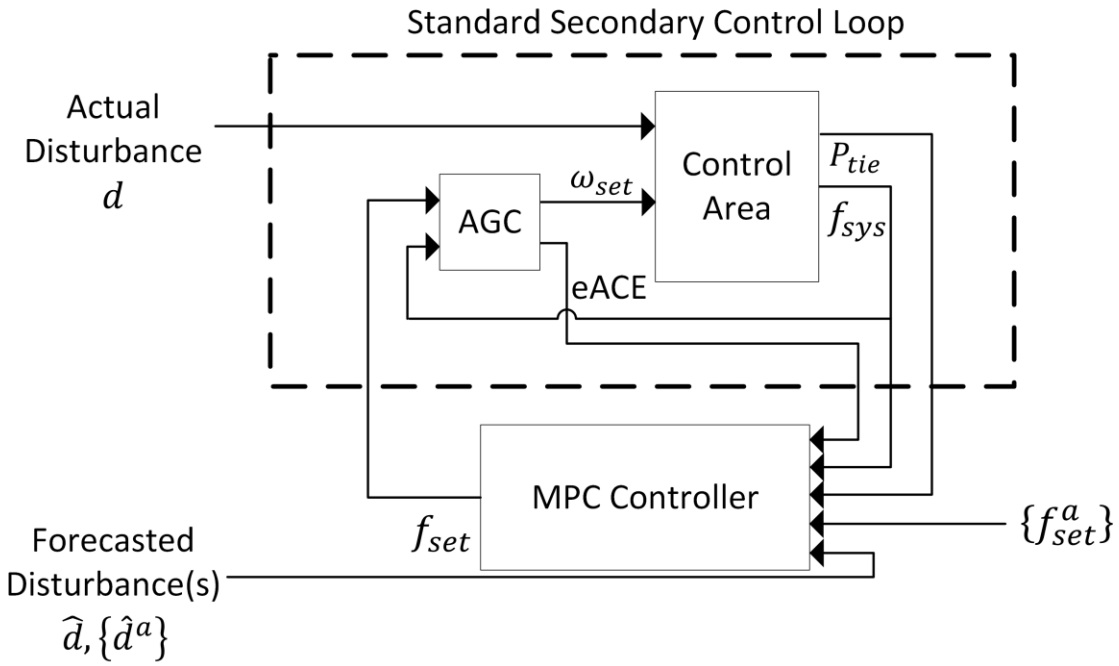


Figure 4-2 Anticipatory Secondary Control, with an Augmented Controller

Traditional AGC in a power system minimizes deviations from a system area frequency reference f_{set} using a tuned integral controller. Specifically, it minimizes ACE, repeated here for convenience:

$$\text{ACE} = \Delta P_{tie} - 10\text{B}\left(f_{sys} - f_{ref}\right) \quad (4\text{-}2)$$

which is a combination of the net tie-line power flow deviation ΔP_{tie} and the system frequency deviation from a seldom-adjusted frequency reference f_{ref} weighted by the frequency bias B. To adjust the area frequency set point and leave the existing AGC system with its integral control intact, the effective system area frequency reference will be changed by an external anticipatory MPC controller by adjusting f_{set} in eACE, defined as:

$$\text{eACE} = \Delta P_{tie} - 10\text{B}\left(f_{sys} - \left(f_{ref} + f_{set}\right)\right) \quad (4\text{-}3)$$

where eACE is still minimized using the same integral controller that exists within AGC. The only difference between eACE and the traditional ACE calculation is the change in the effective area frequency reference term. With traditional ACE, f_{ref} seldom changes, usually only offset in normal operation by a maximum of 0.02 Hz for time error corrections [120]. Using eACE, however, the external controller will manipulate f_{set} to shift the effective area frequency reference $f_{ref} + f_{set}$, which changes the behavior of AGC. This will enable the external controller to ultimately control the governor speed reference values to more quickly correct frequency or tie-line flow deviations.

4.2.1. Background on MPC Controller Design

MPC is an optimal control method used to solve control problems, with advantages over traditional linear quadratic regulation in being able to compensate for uncertainties in future inputs or errors in the plant model. MPC can offer these benefits by solving a

quadratic programming problem at every time step and applying a control action after every solution [122].

To describe the MPC controller that performs anticipatory secondary control (in Figure 4-2), some background material on MPC controllers must first be presented. Let a system be represented by the linear time-invariant (LTI) discrete state-space process model

$$X_{i+1} = AX_i + BU_i \quad (4\text{-}4)$$

where A is the N x N matrix describing the system, B is the N x M matrix for the effect of the input, X_i is the N x 1 vector of states at time i, and U_i is the M x 1 vector of manipulated variables at time i. A measured, uncontrollable input (such as a step change in generation) is modeled as a disturbance d_i, which is contained within the vector U_i. In this model, d_i differs from the actual disturbance by some forecast error ε_d.

MPC is an iterative control method, solving an optimization problem at every time step that minimizes a cost function subject to a set of constraints over a finite, rolling horizon. Resulting control actions are discrete with a resolution of T_s seconds, also called the control interval. MPC minimizes a weighted sum of p future predicted states values where $i \in [0,1 \ldots p-1]$ and the states X_i and inputs U_i at some present time t are defined as:

$$X_i = X(t + T_s i) \quad (4\text{-}5)$$
$$U_i = U(t + T_s i) \quad (4\text{-}6)$$

Let there be N states with M inputs to a plant to be controlled. MPC minimizes the cost function J as follows [122]:

$$\min_{U_1 \ldots U_p} J = \sum_{i=1}^{p} X_i^T Q X_i + U_i^T R U_i \quad (4\text{-}7)$$

where X_i is the N x 1 state vector on the ith step in the prediction horizon, U_i is the M x 1 input vector on the ith step, Q is the N x N penalty matrix for state deviations, R is the M x M penalty matrix for control actions, and p is the length of the prediction horizon. Note that Q must be positive semi-definite and R must be positive definite for a solution to be obtained. This formulation minimizes movements from both desired states and desired control levels. To both control the shape of the control action and reduce computation time, an extra cost term may be applied to smooth the controller output and a control horizon m can be defined for control actions:

$$\min_{U_1 \dots U_m} J = \left(\sum_{i=1}^{p} X_i^T Q X_i\right) + \left(\sum_{j=1}^{m} U_j^T R U_j + \dot{U}_j^T S \dot{U}_j\right) \quad (4\text{-}8)$$

where $\dot{U}$ is the 1st derivative of the control action, and S is the weighting matrix associated with $\dot{U}$. Note that $p \geq m$, and

$$\dot{U}_j = U_j - U_{j-1} \quad \text{for } t > 0 \quad (4\text{-}9)$$
$$\dot{U}_j = 0 \quad \text{for } t = 0. \quad (4\text{-}10)$$

Now the formulation allows for some smaller set of actions $U_1 \dots U_m$ to be calculated, rather than the full set of $U_1 \dots U_p$, which saves on computation time if $m < p$. Because the states X_i must be calculated out to the prediction horizon to solve the optimization problem, the control inputs after the control horizon are defined as

$$U_j = U_m \text{ for } m < j \leq p. \quad (4\text{-}11)$$

Let p denote the length of the prediction horizon in the optimization problem, where all system states and m control actions during this horizon are accounted for in the minimization. After the optimization problem is solved at some time t, only the first action in $[U_1 \dots U_m]$ is taken, U_1. In the next time step, all states are sampled, and the future states are estimated again while MPC again calculates a new optimal series of U_i, then again

executing the new first action U_1. Note that the size of $[U_1 \dots U_m]$ does not change, always maintaining m vectors of size M x 1. In this way, MPC implements its receding horizon, and it is in this sense that it is termed a real-time or online method of optimization. Determining a suitable length of the horizon p is non-trivial, though it has been shown that there exists a horizon p for which MPC is stable and feasible, for any given controlled system [123].

The choice of T_s, p, and m can drastically affect the function of the controller, balancing performance with computational effort [124]. Smaller T_s results in better performance at the cost of computational effort, as the greater temporal resolution in state computation allows the controller more overall control actions and a better estimate of the impact of these control actions over a given time. Typically, T_s is preferred to be less than 10% to 25% of the desired closed-loop response time. The size of p controls how far in the future the controller will calculate system states, defining the horizon of the process model. A needlessly large p introduces a large computational burden, but may also exacerbate any model inaccuracies as any errors build up over the prediction horizon. Too small of a p will result in poor performance due to the inability of the controller to compensate for future effects of a disturbance. Typically, p should be chosen to be greater than both the desired closed-loop response time and the plant delay. The choice of m determines the number of control actions calculated, defining the horizon of the controller model. After m steps, the system inputs remain constant for the rest of the process model horizon p. A small m reduces the computational burden with a tradeoff in performance from the limited number of control actions the controller can take, and $p \geq m$.

4.2.2. Implementing MPC for Anticipatory Secondary Control

The MPC controller and its plant model for a single controller in a multi-area system is shown in Figure 4-3, where the plant represents the standard secondary control loop group shown in Figure 4-2. The power system plant model shown in Figure 4-3 is the linear state-space process model that the controller uses to describe the dynamic behavior of the actual plant, which is the multi-area power system. The full plant model will be described in Section 4.4.1, after the construction of the power system model. For this section, the generic state-space plant definition shown in (4-4) will suffice.

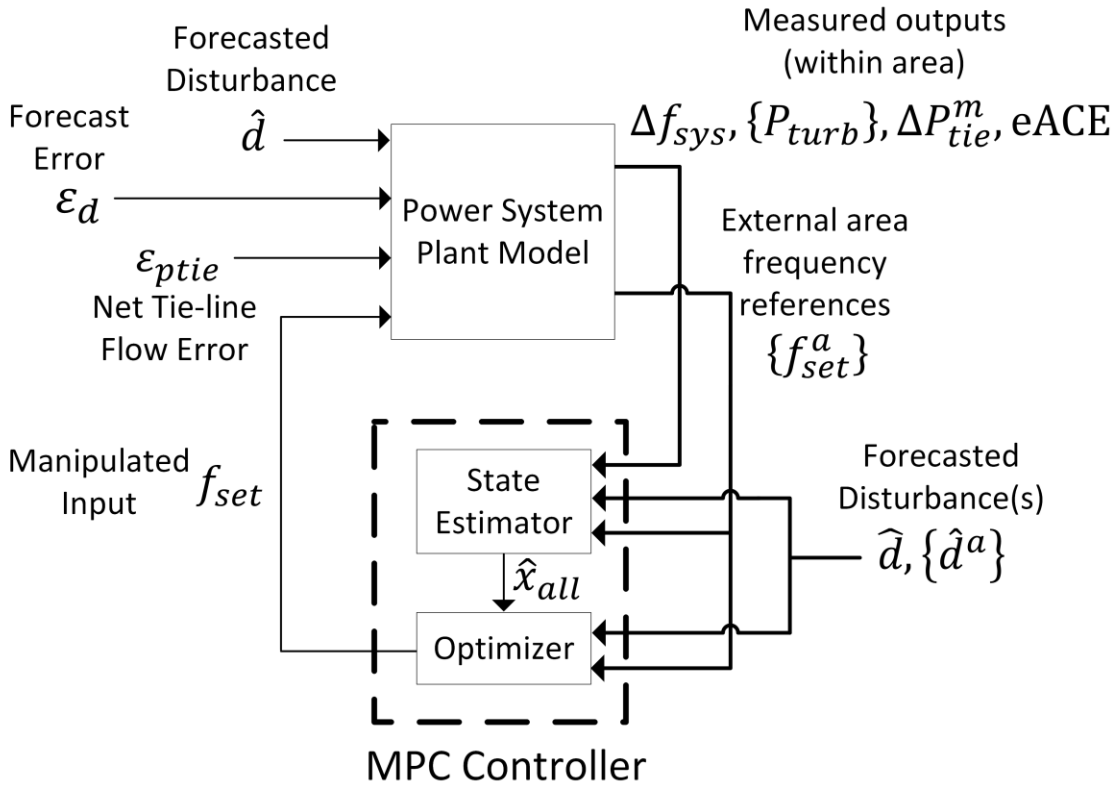


Figure 4-3 The MPC Controller and Plant Model Relationship

The plant is modeled to have inputs of a manipulated variable input f_{set} from the controller, a forecasted disturbance $\hat{d}$, a forecast error ε_d, and the estimation error in net

tie-line interchange ε_{ptie}. The plant also has a set of measured outputs consisting of the measured system frequency at a bus f_{sys}, a set of measured turbine total outputs $\{P_{turb}\}$, measured net tie-line flow (through a low-pass filter) out of the area P_{tie}^{m}, and the eACE. Note that in the state-space model, the calculated tie-line flow can differ from P_{tie}^{m} because of a disturbance in another area. Instead of attempting to estimate this disturbance, the disturbance is estimated as a power injection into the area, which is why ε_{ptie} is an input. The actual disturbance d to the plant is equivalent to the sum of the forecasted disturbance $\hat{d}$ and the forecast error ε_d, though the two are separated for the convenience of modeling the forecast error of the disturbance, which is unmeasured and unknown. Thus, from the controller's perspective, input variables into the plant are all measured and known variables, with the exception of ε_d and ε_{ptie} as unknown inputs. The MPC controller is composed of an optimizer and a state estimator, as most of the states in the system are not measured and must be estimated with the possibility of measurement errors, imperfect modeling, or a non-zero forecast error ε_d. The state estimator uses all measured plant outputs, the provided time series of forecasted disturbance in its area $\hat{d}$ as well as other areas $\{\hat{d}^a\}$, and sampled external area frequency references $\{f_{set}^a\}$. The optimizer uses the state estimator's estimated states $\hat{x}_{all}$ as well as the provided data for the forecasted disturbance and external area frequency references. At every time step, the state estimator estimates the state vector $\hat{x}$ based on the measured outputs in the power system, then passes this estimated state vector into the optimizer. The optimizer predicts the sequence of frequency set point movements that minimizes the frequency error, using a linear system model to predict the effect of its frequency set point movements some number of steps in

advance. Details of the optimizer and state estimator used in this work will be provided in the next two subsections, though further details are available in MATLAB's documentation [125].

With the structure shown in Figure 4-3, the state-space plant model can be further refined. Recall that the general state space formulation of the plant is written as

$$X_{i+1} = AX_i + BU_i. \quad (4\text{-}12)$$

Now, a vector of outputs is defined as $Y_i = \left[f_{sys}, \{P_{turb}\}, P_{tie}^m, \text{eACE}\right]^{\text{T}}$ and a vector of inputs is defined as $U_i = \left[\hat{d}, \varepsilon_d, P_{tie}, f_{set}\right]^{\text{T}}$ such that a more detailed state-space representation can be written:

$$X_{i+1} = AX_i + BU_i \quad (4\text{-}13)$$
$$Y_i = CX_i + DU_i \quad (4\text{-}14)$$

Details of the actual states used are presented in Section 4.4.1 after the power system is formulated in detail. Now, the MPC controller formulation is written as

$$\min_{f_{set_1} \cdots f_{set_m}} J = \left(\sum_{i=1}^{p} Y_i^T Q Y_i\right) + \left(\sum_{j=1}^{m} U_j^T R U_j + \dot{U}_j^T S \dot{U}_j\right) \quad (4\text{-}15)$$

s.t.
$$X_{i+1} = AX_i + BU_i \quad (4\text{-}16)$$
$$Y_i = CX_i + DU_i \quad (4\text{-}17)$$
$$u_{min} \leq f_{set} \leq u_{max} \quad (4\text{-}18)$$

where u_{min} and u_{max} are the bounds for the manipulated frequency set-point movement f_{set}, which are set to -0.1 Hz and 0.1 Hz, respectively. Notice that the objective function minimizes some combination of outputs Y_i, inputs U_i, and the change in inputs $\dot{U}_i$. In particular, the controller will only be concerned with minimizing the system frequency deviations Δf_{sys} and the tie-line deviations ΔP_{tie}^m in the output vector Y_i, its controlled input f_{set} in the input vector U_i, and the rate of change of its controlled input $\dot{f}_{set}$ in $\dot{U}_i$. Thus,

the weighting matrix Q has only two non-zero elements on its diagonal, and is otherwise empty. In addition, R and S have only one non-zero element. This structure, where weighting matrices are used with only a couple of non-zero elements, is directly used for the ease of implementation of the MPC controller in MATLAB. However, the problem can be equivalently stated as follows:

$$\min_{f_{set_1} \cdots f_{set_m}} J = \left(\sum_{i=1}^{p} Q_f \left(\Delta f_{sys}\right)^2 + Q_{ptie} (\Delta P_{tie}^m)^2 \right) + \left(\sum_{j=1}^{m} R_f (f_{set})^2 + S_f \left(\dot{f}_{set}\right)^2\right) \quad (4\text{-}19)$$

s.t.
$$X_{i+1} = AX_i + BU_i \quad (4\text{-}20)$$
$$Y_i = CX_i + DU_i \quad (4\text{-}21)$$
$$-0.1 \leq f_{set} \leq 0.1 \quad (4\text{-}22)$$

where the objective function is now explicitly defined with scalar penalties Q_f for the frequency deviation, Q_{ptie} for the net tie-line flow deviation, R_f for non-zero f_{set}, and S_f for non-zero changes in the frequency set point $\dot{f}_{set}$. These weights are time-invariant in this work, and will be defined prior to the actual simulations, as they can drastically change the performance of the controller. However, the weights Q_f and Q_{ptie} are greater than or equal to those of the frequency set point weights, which are set to $R_f = 0.15$ and $S_f = 0.75$ for these simulations.

With the formulation now described, the MPC state estimator and optimizer will be discussed in the following two subsections.

(i) MPC State Estimator

The controller does not have measurements for all states inside of the plant, and so many of the states must be estimated based upon measured outputs. The state estimator is based on a steady state Kalman filter, where the state-space matrices *A*, *B*, *C*, and *D* are

time-invariant and the states are iteratively estimated at each time step. The state estimator is based on the state space observer, which is simply the same formulation as shown in (4-13) and (4-14). For convenience of discussing the Kalman filter, the equations are rewritten as

$$x(i+1) = Ax(i) + Bu(i) + Gw_k \quad (4\text{-}23)$$
$$y(i) = Cx(i) + Du(i) + Hw_k + v_k \quad (4\text{-}24)$$

where w_k and v_k represent white noise vectors that is adds white noise to all states or measurements, respectively, following the normal distributions $w_k \sim N(0, \text{diag}(Q_{kal}))$ and $v_k \sim N\big(0, \text{diag}(R_{kal})\big)$, where $\text{diag}(\cdot)$ is the diagonal of a matrix. In this work, G and H are identity matrices to create a one-to-one correspondence from a given value in Q_{kal} or R_{kal} to a state or measurement. The importance of these variance values in the state estimation will be presented later in this section. The Kalman filter uses observations to provide a corrected estimate of the states and measurements in the system. Assume the following are given:

- $x(i|i-1)$, the controller state estimate from the previous control interval $i-1$
- $u^{act}(i-1)$, the manipulated variable used in the plant from $i-1$ to i
- $d(i)$, measured disturbances
- $y_m(i)$, measured plant outputs
- B_u, B_v, columns of observer parameter B corresponding to measured plant outputs
- D_{mv}, rows and columns of observer parameter D corresponding to measured plant outputs and measured disturbance inputs
- L, M, constant Kalman gain matrices

First, the innovation is computed as

$$e(i) = y_m(i) - [C_m x(i|i-1) + D_{mv} v_k(i)] \quad (4\text{-}25)$$

which is the difference between the measured and previously calculated output. Then, the states are updated to take into account the latest measurements

$$x(i|i) = x(i|i-1) + Me(i) \quad (4\text{-}26)$$

This corrected state $x_c(i|i)$ is then passed to the optimizer to solve the quadratic program at interval i, and the solution is the control action $u^{act}(i)$. The state estimator then calculates the next set of estimates for the next interval as

$$x(i+1|i) = Ax(i|i-1) + B_u u^{act}(i) + B_v v_k(i) + Le(i) \quad (4\text{-}27)$$

This iterative process repeats at every time step to arrive at estimates of the states, subject to user-defined noise levels in the signal. The Kalman gains L and M are directly proportional to Q_{kal} and R_{kal}, such that larger Kalman gain values result in a heavier weight on measurements.

A brief discussion on the importance of the process and noise covariances will be covered here, though more details are available in the references. Define the process noise covariance matrix as Q_{kal} and the measurement noise covariance matrix as R_{kal}. These matrices must be defined by the user, and may be time-varying, though in this work the matrices remain constant. If Q_{kal} is larger than R_{kal}, then the Kalman gains L and M grow which weights the error more heavily, and the Kalman filter places more weight on measurements in determining state estimates. If R_{kal} is larger than Q_{kal}, then the state estimates place more weight on the process model by weighting the residual less heavily. Thus, the choice of Q_{kal} and R_{kal} can be seen as a sort of trust measure, as if a particular state calculation or measurement is to be trusted more, its corresponding row in the

covariance matrix Q_{kal} or R_{kal} should be smaller. In this work, Q_{kal} and R_{kal} are simply diagonal matrices, and the values in Q_{kal} and R_{kal} are described in Section 4.4.1. More details on the Kalman filter can be seen in literature [126].

(ii) MPC Optimizer

The MPC optimizer solves a quadratic program (QP) at each control interval, which determines the value $u^{act}(i)$ that is used as the controller output to the plant until the next interval. The KWIK algorithm was used to solve the QP problem, as a part of the MATLAB Model Predictive Control Toolbox, which has performance gains over other popular QP solvers in that it can solve the problem in $O(n^2)$ time versus $O(n^3)$ time, where n is the number of degrees of freedom in the optimization problem [127]. More details are available in literature and in the MATLAB documentation [128].

4.3. Power System Models to Test Secondary Control Methods

A power system model was required to analyze the performance of secondary control schemes in a realistic environment, so a representation of a power system was constructed, based on publicly available data of the Electric Reliability Council of Texas (ERCOT) [129]. First, a single-area system was modeled and various simulations were run to validate the primary response and secondary response of the system under traditional AGC control, and to introduce the models used in this work. A 2-area system, based on the two validated single-area power systems, was tested under a load disturbance with tie-line bias control. Finally, the 3-area system was modeled and tested for use as a testbed for the anticipatory secondary control method.

4.3.1. Single-Area Model Construction & Validation

First, a 6-bus, single-area system was created as shown in Figure 4-4, with generation and load data as shown in Table 4-1 and a system base of 25,000 MVA. This system has 7,500 MW (0.3 pu on a system base) of primary response capability, all of which is also responsive to AGC, at buses 1, 4, 5, and 6. There is a total capacity of 17,500 MW of constant, non-responsive generation at buses 1, 2, 4, 5, and 6, operating at 100% of their capacity. In this study, there is 24,250 MW of load in the system, which results in responsive generation operating at 90% of their capacity.

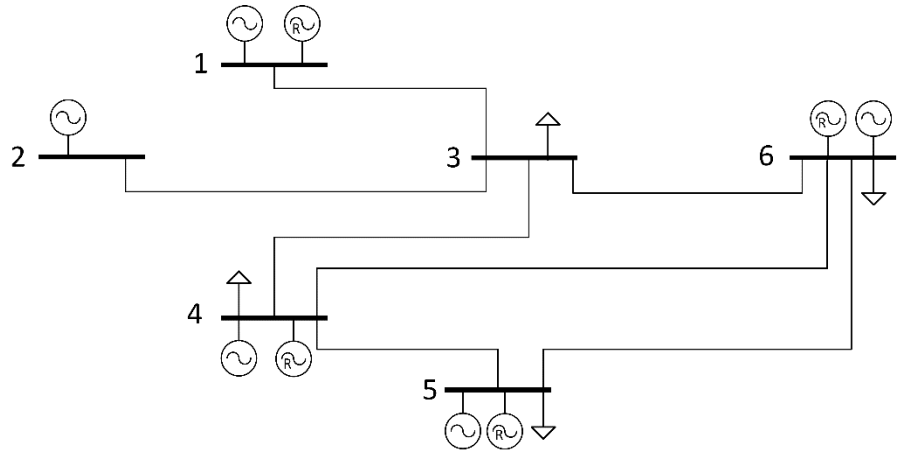


Figure 4-4 Power System One-line Diagram, Marking Responsive Generation with an 'R'

Table 4-1 Generation and Load in Test Power System

Bus	Total Generation Unit Size (MW)	Responsive Generation Capacity (MW)	Non-responsive Generation (MW)	Load (MW)
1	2700	900	1800	-
2	2500	0	2500	-
3	-	-	-	8000
4	4500	1500	3000	6000
5	6600	2200	4400	7000
6	8700	2900	5800	3250

Responsive generation, present at buses 1, 4, 5, and 6, are modeled with classical steam-electric models at each bus, while non-responsive generation, present at all buses except for bus 3, are modeled as electrical power injections at each bus. All buses were

assumed to be at 1 pu voltage, so electrical models of generation, such as exciter models, were not used. Figure 4-4 shows responsive generation marked with an 'R' in the diagram.

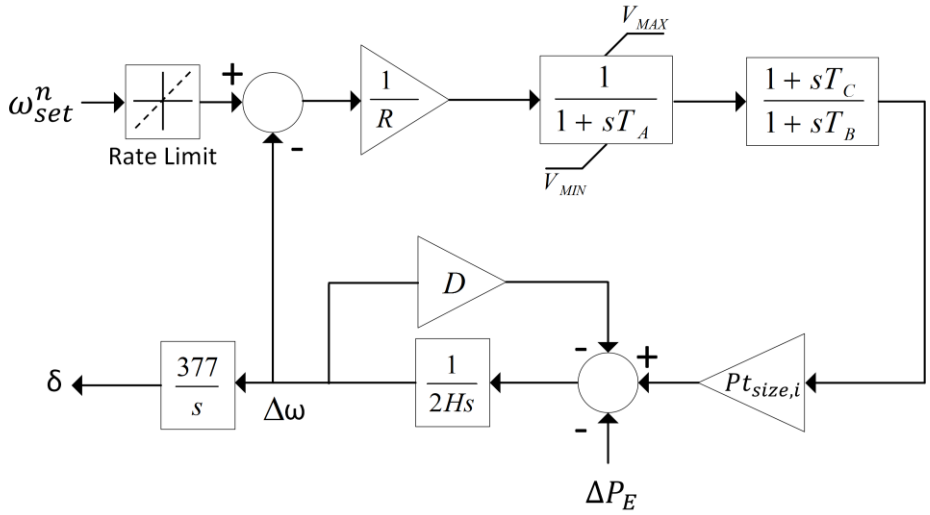


Figure 4-5 The Steam-electric Representation of a Generation Unit at a Responsive Bus

Table 4-2 Parameters of the Steam-electric Model

Parameter	Value
R	4%
T_A	0.5
T_B	10
T_C	3
$Pt_{size,i}$	Gen. size / 25,000 MW
V_{MAX}	1
V_{MIN}	0
D	1
H	4

The steam-electric model used in these simulations is shown in Figure 4-5 with parameter data shown in Table 4-2. The governor droop setting for all responsive generation was chosen to be 4%, resulting in an equivalent system droop value of 13.3%, and the time constant T_A of the governor was set to 0.5 seconds. The valve limits V_{MIN} and V_{MAX} were set to the full range of 0 and 1 pu on the machine base, respectively. The

turbine's T_B and T_C values were chosen to be 10 seconds and 3 seconds, respectively, to reflect that 30% of the power appears immediately, while the rest of the power appears over time. This lead-lag representation emulates a simple model of a steam plant's boiler and turbine, which has a high pressure turbine whose power follows the steam valve position with a negligible delay, as well as low pressure turbine whose power is lagged from the steam valve position as a result of the large amounts of volume in the reheater piping [130].

The values of $Pt_{size,i}$ for each generator at a bus i was the pu size of the generator in the system base, based on the values in Table 4-1. The load damping value D was set to 1/6 pu at every bus, resulting in a system per-unit value of 1 which reflects a 1% load loss for a 1% frequency decrease. The inertia constant H was set to 4/6 pu at every bus, resulting in a system per-unit value of 4.

The governor speed reference ω_{set}^{n} for responsive generation at bus n is provided directly by AGC. To avoid excessive strain on responsive generation, a rate limit is imposed such that the turbine power will only ramp at some value per minute. Unless stated otherwise, the ramp limit is set such that the responsive generation is ramp-limited at 0.02 pu/minute on the responsive turbine base. Note that the primary response in the system is not rate-constrained, and that only the secondary control commands are constrained.

The transmission line connections between each bus in the power system model are not rigid. The power flow between connected buses i and j follows a DC representation

$$P_{tie-ij} = K\left(\delta_i - \delta_j\right) \quad (4\text{-}28)$$

where *K*= 0.2 in these simulations, which corresponds to a 5 pu impedance on a 25000 MVA base or 0.02 pu on a 100 MVA base. Actual implementation in the model follows the equation

$$\Delta P_{tie-ij} = 377 * \frac{K}{s}\left(\Delta\omega_i - \Delta\omega_j\right) \quad (4\text{-}29)$$

where s^{-1} is the integration operator, as the per unit speed deviation $\Delta\omega_i$ is equivalent to the change in the per unit rotor angle deviation $\frac{d}{dt}(\Delta\delta_i) = s(\Delta\delta_i)$. Note that the factor of 377 is in rad/sec to convert $\Delta\delta_i$ (radians) to $\Delta\omega_i$ (pu speed change). Modeling the power flow in this manner lowers the number of states that must be calculated in the system by ignoring actual rotor angles. Details of the selection of transmission line impedances are presented in the Appendix B I.

(iii) System Response to a Unit Step Load Increase: No AGC

For this simulation, a disturbance was modeled as a unit step increase in load at bus 2, which can represent either a step increase in load or a step decrease in generation. Turbine speed deviations from nominal were used to represent system frequency at the buses. The frequency at bus 6 was chosen to represent the system frequency, as the frequencies at every bus through the system were tightly connected as shown in Appendix B, Section II .The frequency at the monitored bus as a result of a 250 MW load increase (0.01 pu system base) at bus 2 at *t* = 25 seconds is shown in Figure 4-6. The theoretical drop in frequency is

$$\Delta\omega = \frac{\Delta P_L}{D+\frac{Gen_{res}}{R}} = \frac{0.01}{1+\frac{0.3}{0.04}} = 0.0012\ pu = 0.0706\ \text{Hz} \quad (4\text{-}30)$$

where Gen_{res} is the fraction of responsive generation in the system, R is the governor droop for responsive generation, D is the load damping value, and ΔP_L is the increase in load. The system settles at 59.9294 Hz, which agrees with the theoretical results.

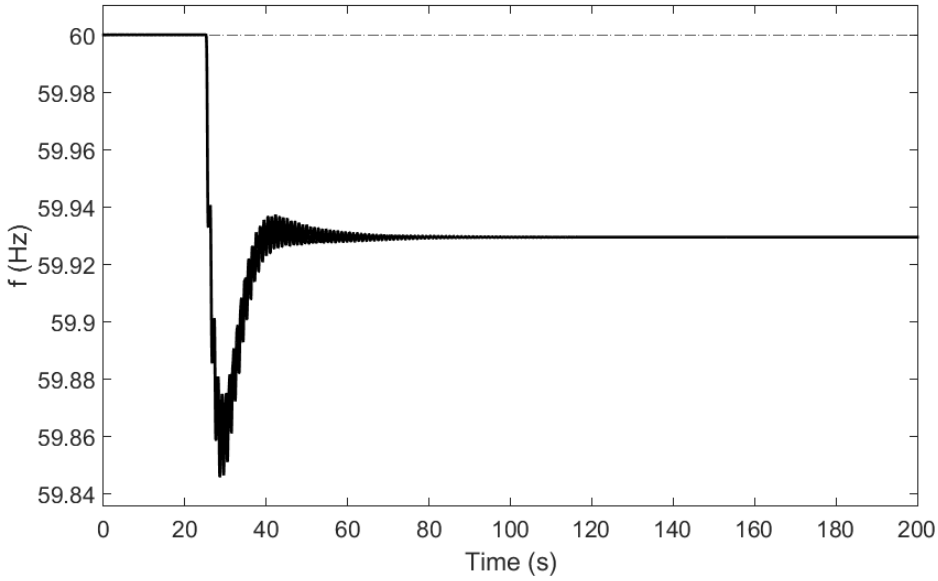


Figure 4-6 The Frequency Time Series of the System without AGC after a 0.01 pu Drop

Droop control in responsive generation is responsible for arresting the frequency. A 4% droop control corresponds to a full-range 1 pu turbine output change for 2.4 Hz of change in a 60 Hz system, which means that 0.0706 Hz change should result in 0.0294 pu change in responsive generation, which is seen in Figure 4-7.

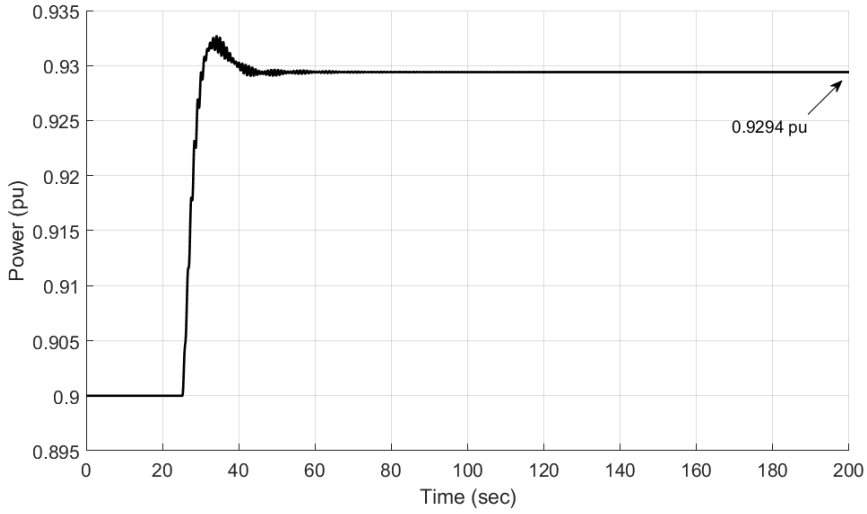


Figure 4-7 The Responsive Turbine Output of the System without AGC after a 0.01 pu Drop

In addition, the swing equation shows that the initial rate of decline of frequency should be

$$\frac{\mathrm{d}\omega}{\mathrm{dt}} = \frac{\Delta \mathrm{P}}{2H} = \frac{0.01}{2*4} = 0.00125 \text{ pu/sec} = 0.075 \text{ Hz/sec} \qquad (4\text{-}31)$$

which is shown in Figure 4-8. Low-frequency oscillations are seen at around 1 Hz.

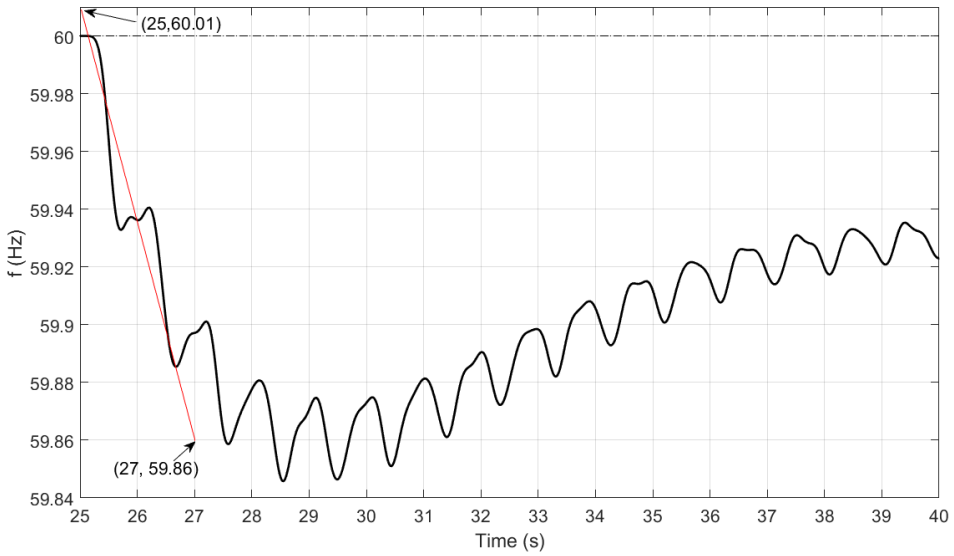


Figure 4-8 A Close-up of the System Frequency after a 250MW Load Change

(iv) Automatic Generation Control Model Description

Automatic Generation Control (AGC) functioned as pure frequency control in this single area system and was modeled as shown in Figure 4-9 and Table 4-3. The frequency of the system at bus 6 was chosen to represent the system frequency f_{sys} as the bus frequencies across the area were relatively tightly connected (see the Appendix B, Section II for details). The difference between the system frequency and the nominal reference frequency $f_{set} = 60$ Hz was then minimized by an integral controller. The gain of the integral controller was tuned to return the measured frequency to within 0.01 Hz of nominal in 200 seconds with no overshoot, given a 0.01 pu load increase at bus 2. This yielded $K_i = 0.015$. Participation factors pf_i for a unit at bus i were set equally to 0.25 for all four controllable generation units. The limiter imposes limits $\text{eACE}_{max} = 325$ MW $= -\text{eACE}_{min}$ on the eACE value passed into the integral controller such that the turbines will only ramp at 2% of the responsive generation capacity per minute, also staying under the ramp limit of the speed reference ω_{set}^{i} at a generator i. However, in this section, eACE is equal to ACE as the area frequency reference $f_{set} = 60$ Hz.

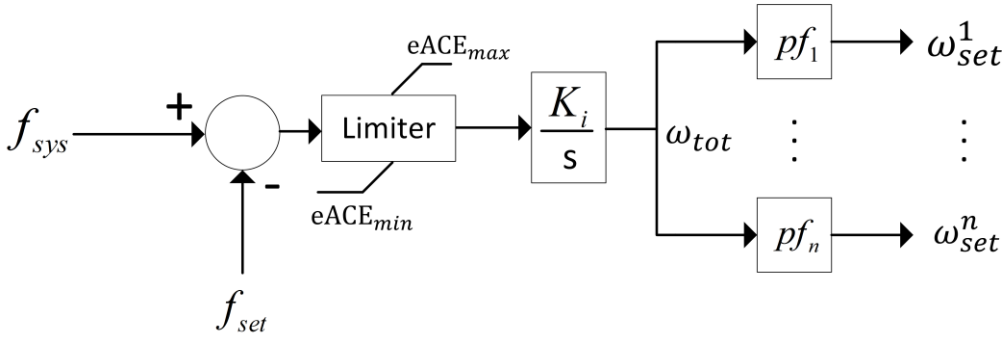


Figure 4-9 Automatic Generation Control Model

Table 4-3 Parameters of the AGC Model

Parameter	Description	Nominal Value
f_{set}	Frequency reference of AGC	60 Hz
f_{sys}	Frequency of system at bus 6	-
K_i	Integral gain of AGC	0.015
pf_i	Participation factor for unit at bus *i*	25% for each AGC-responsive generator
eACE_{max}	Maximum eACE value	325 MW
eACE_{min}	Minimum eACE value	-325 MW
ω_{tot}	AGC output set point	-
ω_{set}^{i}	Governor speed reference at bus *i*	-

(v) System Response to a Unit Step Load Increase: With AGC

For this simulation, a load disturbance was modeled as a unit step increase in load at bus 2. The result of this 250 MW load increase (0.01 pu system base) at *t* = 25 seconds is shown in Figure 4-10. The frequency settles to within 0.01 Hz of 60 Hz before 200 seconds.

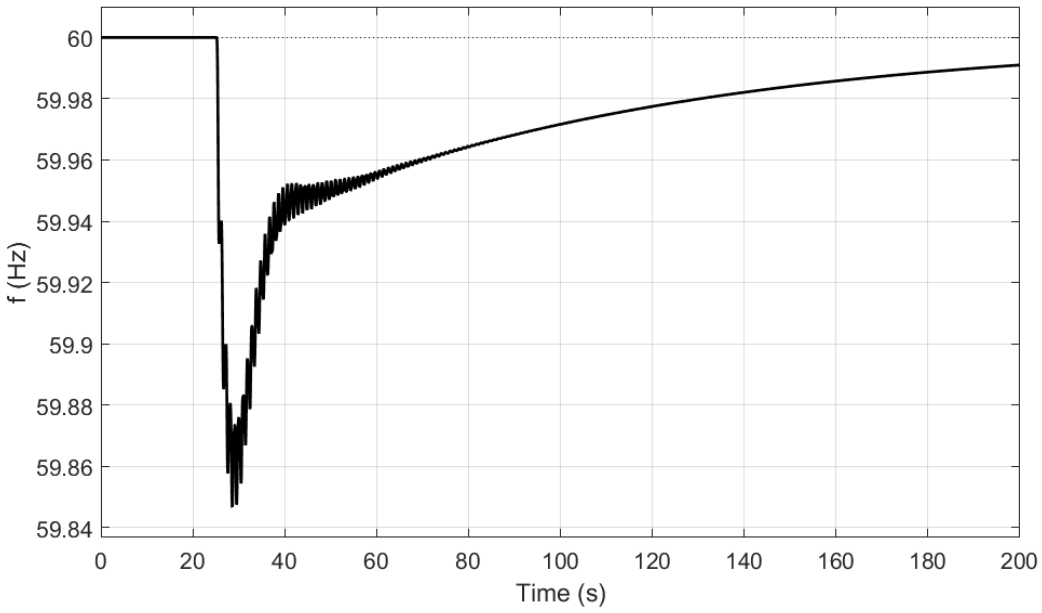


Figure 4-10 The Frequency Time Series of the AGC Controlled System after a 0.01 pu Drop at *t* = 25 seconds

Responsive turbine output is seen in Figure 4-11, which approaches the final value of 0.933 pu as it picks up all 250 MW of extra load.

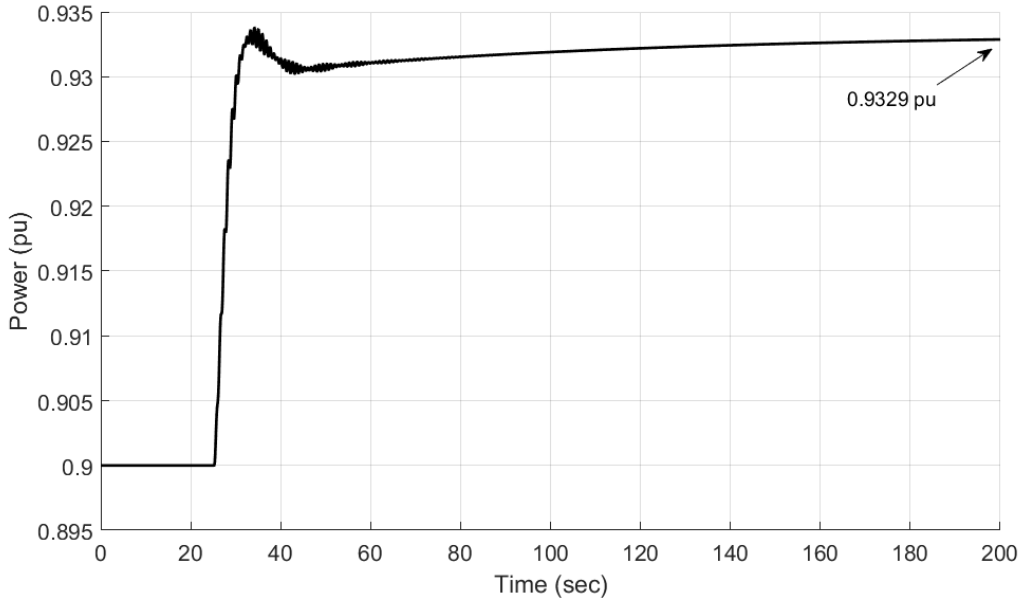


Figure 4-11 The Responsive Turbine Output with AGC in Turbine pu Base

4.3.2. Two-Area System Model Construction & Validation

The construction of a 2-area system is described in this section, to test and validate AGC. The 2-area system, shown in Figure 4-12, is simply two of the single-area systems described in Section 4.3.1, connected by a single tie-line. The system base is set to 25000 MVA, which is the generation capacity of either area. As in the single-area system, the DC power flow approximation is used to represent power flow across transmission lines, and the tie-line's inverse-impedance value K_{tie} is set to 0.2 pu. Note that this corresponds to a tie-line capacity of 20 percent of either area's capacity, as a tie-line flow of 0.2 pu results in one radian angular displacement between the two systems. The steam-electric model parameters were unchanged from the single-area system.

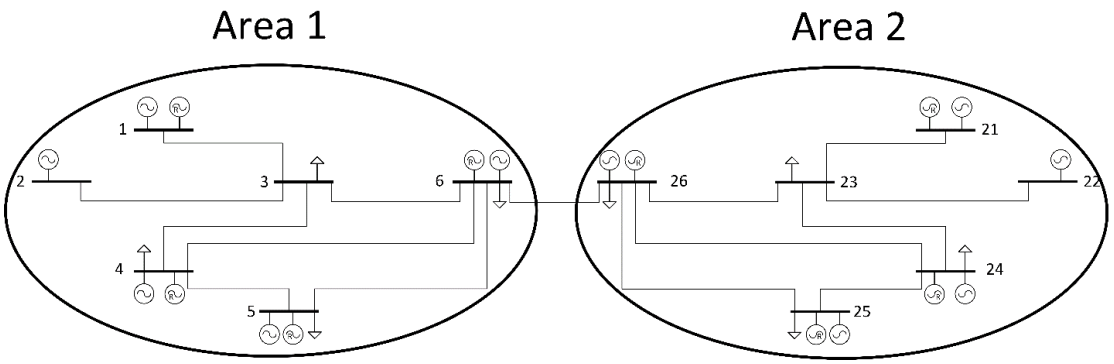


Figure 4-12 Two-area Power System with a Single Tie-line between Buses 6 and 26

(vi) Response of the System to a Unit Step Load Increase: No AGC

A 250 MW sudden load increase was simulated at bus 2 in Area 1 at $t = 25$ seconds. Shown in Figure 4-13 is the rotational speed of the turbines at bus 6, which was used to represent the system frequency, compared with the response in the single area system (Figure 4-6). Section III in the Appendix B shows that the speeds at all buses in the system were tightly connected following a disturbance. Note that per-unit effective system droop and load damping values for the two-area system are doubled when compared with the single-area case, as twice the response generation and load damping is in the system. Following a 0.01 pu sudden increase in load, the theoretical drop in frequency is

$$\Delta\omega = \frac{\Delta P_L}{D+\frac{Gen_{res}}{R}} = \frac{0.01}{2+\frac{0.6}{0.04}} = 5.8824 \text{ x } 10^{-4}\ pu = 0.0353 \text{ Hz} \qquad (4\text{-}32)$$

which results in a frequency of 59.9647 Hz, agreeing with simulated results.

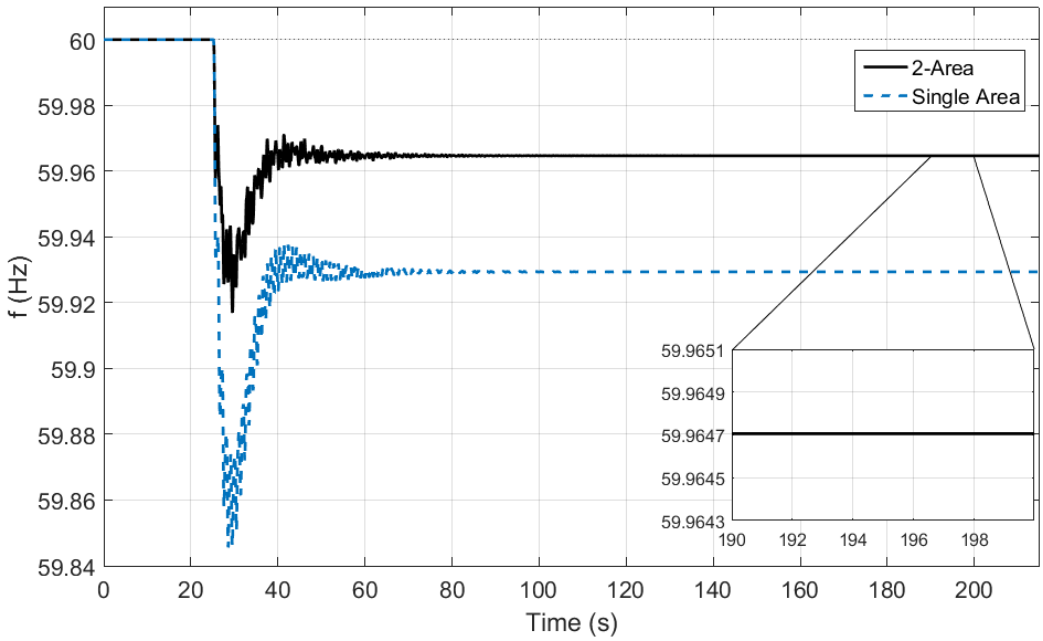


Figure 4-13 Frequency of the 2-area System Compared with the Frequency of the Single-Area System

The two-area system has twice as much load and turbine inertia as the single-area system, so the initial rate of decline of frequency is expected to be half of that in the single-area system. The swing equation shows that the initial rate of decline of frequency should be

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{\Delta P}{2H} = \frac{0.01}{2*8} = 6.25 \text{ x } 10^{-4} \text{ pu/sec} = 0.0375 \text{ Hz/sec} \qquad (4\text{-}33)$$

which is shown in Figure 4-14. Low-frequency oscillations are seen at approximately 1 Hz. As the two systems were identical there was exactly 125 MW of assistance from Area 2 to Area 1 with no secondary control, as seen in the tie-line flow in Figure 4-15.

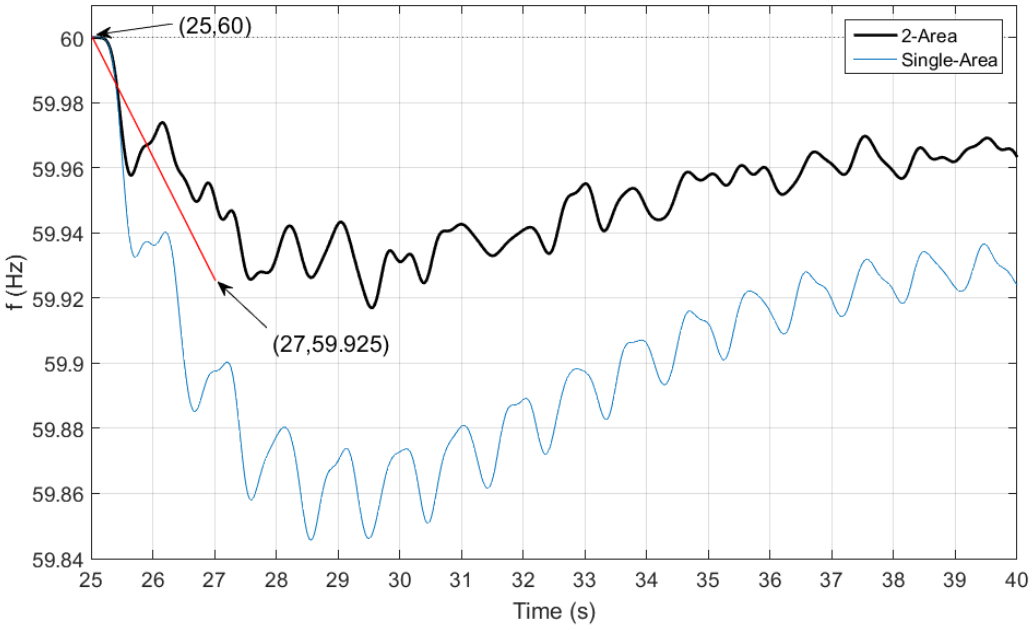


Figure 4-14 Initial Rate of Frequency Decline Comparison between the Single-area and 2-area System

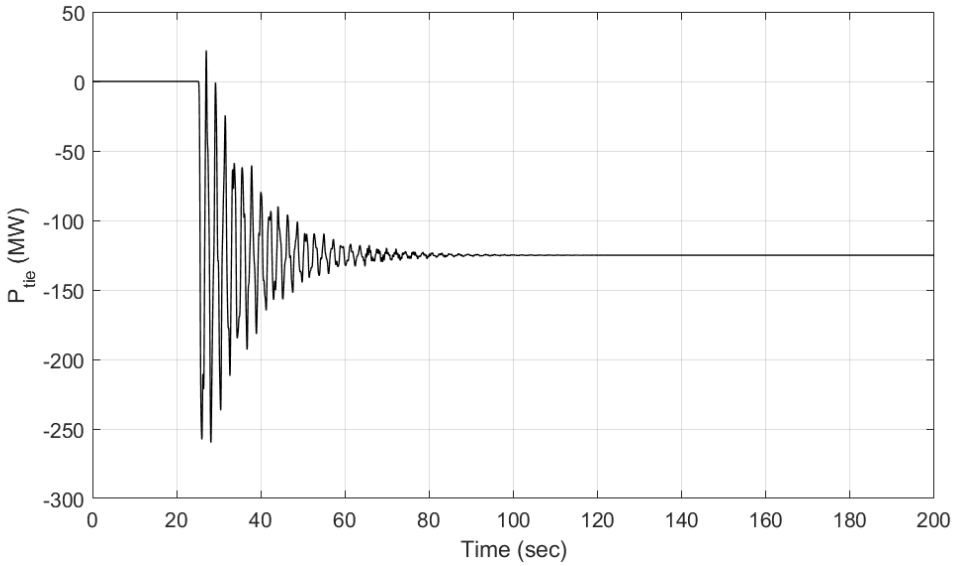


Figure 4-15 Tie-line Flow Out of Area 1

(vii) Automatic Generation Control Model Description for Multiple Areas

With the introduction of multiple areas, automatic generation control must now minimize both frequency and tie-line deviations. Traditional AGC accomplishes this with the use of the Area Control Error (ACE) defined as

$$\text{ACE} = \Delta P_{tie} - 10\text{B}\Delta f_{sys} \qquad (4\text{-}34)$$

where ΔP_{tie} is the sum of tie-line deviations out of the area in MW, B is the frequency bias in MW/0.1 Hz, and Δf_{sys} is the deviation of system frequency in Hz. If the frequency reference f_{set} is a constant 60 Hz, the eACE is the same as the ACE. This value is then minimized with an integral controller, and the resulting governor set point is sent out to the governors of the responsive generation. The block diagram of this AGC system is shown in Figure 4-16.

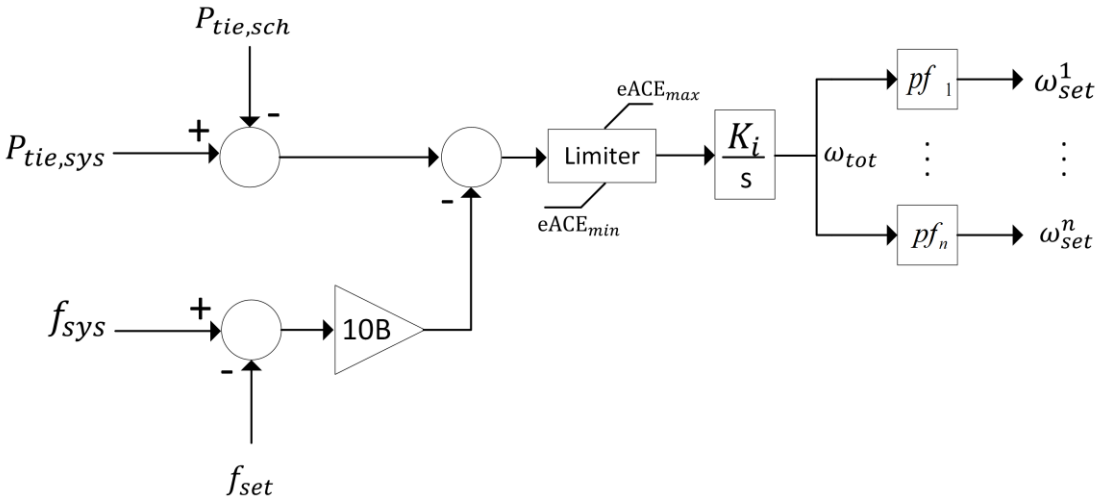


Figure 4-16 Automatic Generation Control for a Single Area in the Multi-area System

The automatic generation control system of an area should only minimally react if a load disturbance occurred outside of its controlled area, only providing initial frequency support. When load disturbances occur outside of an area, the tie-line flow deviation ΔP_{tie}

of the area is the opposite sign of the frequency deviation Δf_{sys}. Recognizing this, the frequency bias B can be set to weight the frequency deviation equally to the tie-line deviation following a disturbance in the power system, in hopes of 'cancelling out' the two ΔP_{tie} and Δf_{sys} values which results in an ACE value of zero when the disturbance is outside of the area. Although the tie-line flows and system frequency do not have the same dynamic behavior and thus are impossible to completely cancel out with one weighting factor, B can be roughly set by balancing the steady-state values of ΔP_{tie} and Δf_{sys} following a disturbance. Details of the derivation can be seen in Kirchmayer's work [113]. After balancing ΔP_{tie} and $10\mathrm{B}\Delta f_{sys}$, it can be seen that a theoretical value for B for the automatic generation control of one area in the constructed two-area system is:

$$\mathrm{B} = 0.1\left(D + \frac{1}{R}\right) = 0.1\left(1 + \frac{0.3}{0.04}\right) = 0.1\left(8.5\ \frac{\text{pu MW}}{\text{pu Hz}}\right) = 354.2\,\frac{\text{MW}}{0.1\ \text{Hz}} \quad (4\text{-}35)$$

where D and R are the load damping and droop of the controlled system, respectively. Section IV in the Appendix B shows the effect of this choice of frequency bias B on the calculation of ACE and tie-line frequency control.

(viii) 2-Area System Response to a Load Increase: With AGC

For this simulation, a load disturbance was modeled as a unit step increase in load at bus 2 in Area 1. The frequency and tie-line flows within the system with a 250 MW load increase (0.01 pu system base) at $t = 25$ seconds is shown in Figure 4-17 and Figure 4-18. The frequency settles to within 0.01 Hz of 60 Hz within 200 seconds of the load disturbance, and the tie-line flow settles to within 5 MW of the nominal flow within 275 seconds.

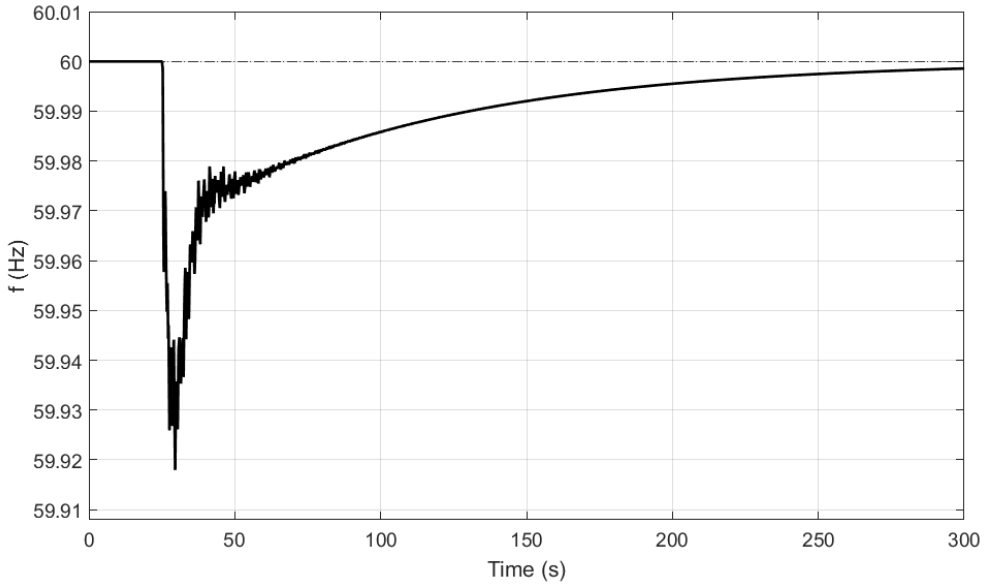


Figure 4-17 Frequency Output of the Power System with Functioning Automatic Generation Control

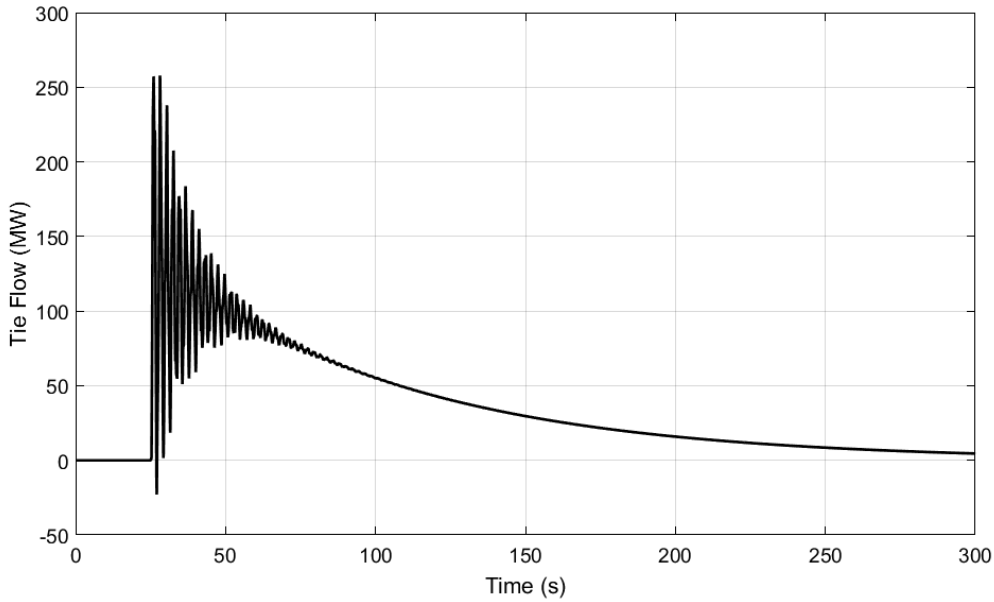


Figure 4-18 Tie-line Flow into Area 1 with Functioning Automatic Generation Control

Responsive turbine outputs in both areas are shown in Figure 4-19, which reflect the proper allocation of changes in steady-state generation as Area 1 begins to picks up all 250 MW of increased load.

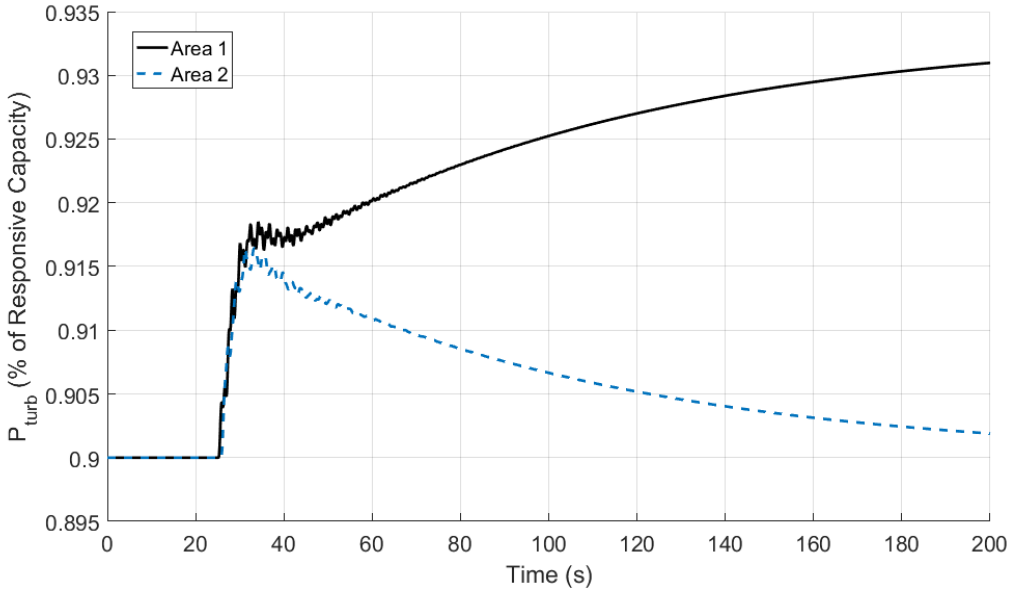


Figure 4-19 Responsive Turbine Changes in Generation in Both Areas

4.3.3. Three-Area System Model Construction & Validation

A 3-area power system was created from three single area systems, connected as shown in Figure 4-20.

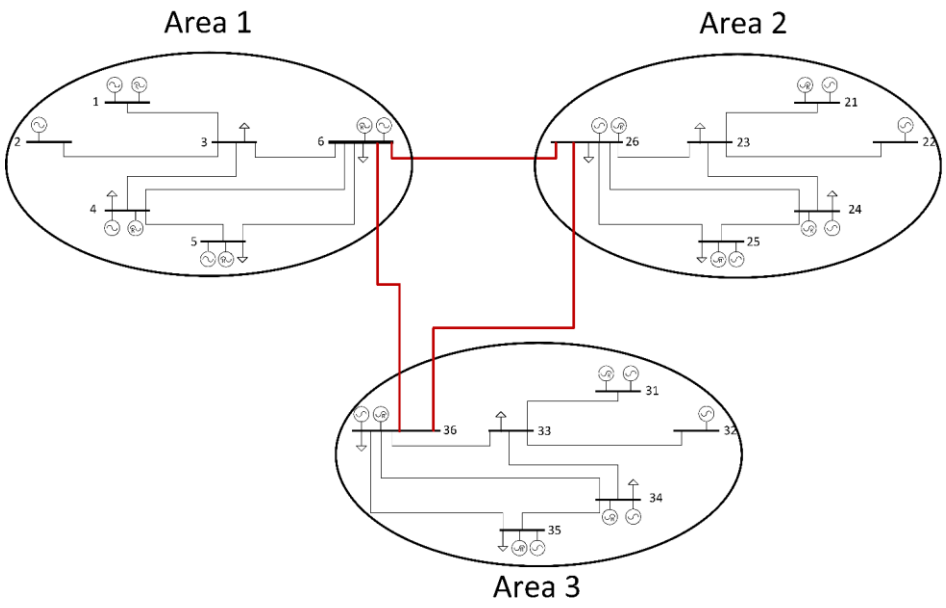


Figure 4-20 The 3-area Test System, with Tie-lines Between Areas Shown in Red

The 3-area system was validated by examining its performance with traditional AGC. With all tie-lines in service, the performance of the system under traditional AGC will be shown.

A 500 MW sudden increase in load is applied to bus 2 in Area 1 at t = 40 seconds. Recall that there is 7500 MW of responsive generation in each area, with a total generation capacity of 25000 MW and an initial load of 24250 MW. Responsive generation outside the area with the load disturbance initially responded to the disturbance to arrest the frequency dive and then returned to the pre-disturbance power output, while responsive generation within Area 1 raised its generation by 500 MW to meet its 500 MW load increase as shown in Figure 4-21.

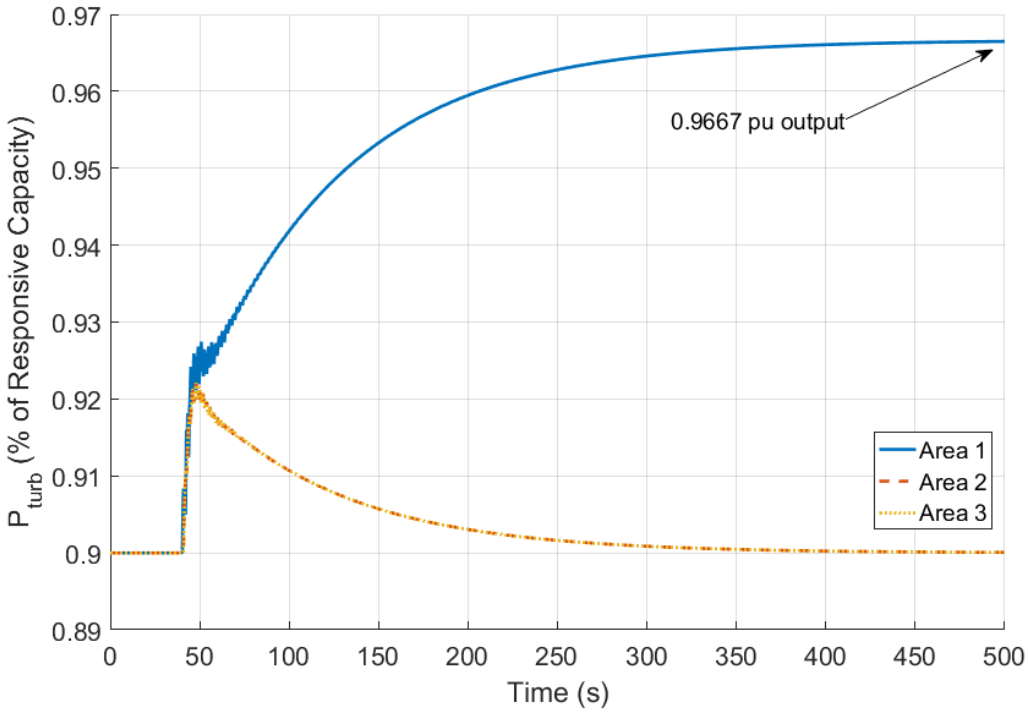


Figure 4-21 Power Output of the AGC-responsive Fraction of Generation in Each Area

Figure 4-22 shows the rotor rotational speeds at bus 6 of each area, representing the frequencies of each area. Though the frequencies are tightly connected, the oscillations in

Area 1 are slightly greater than that of the other areas due to the disturbance taking place in Area 1.

Figure 4-23 shows the tie-line flows. Before the disturbance, there is no flow on any of the tie-lines as each of the 3 areas are identical in generation and load. After the disturbance, tie-flows are brought back to 0 MW by AGC. Note that there is close to no flow on the line connecting Area 2 and Area 3 during or after the disturbance because the two areas are identical. Each of the two areas' responsive generation acts in the exact same manner, so there is no net flow on the line connecting the two areas as the qualities that affect the area dynamics (such as inertia) are identical as well.

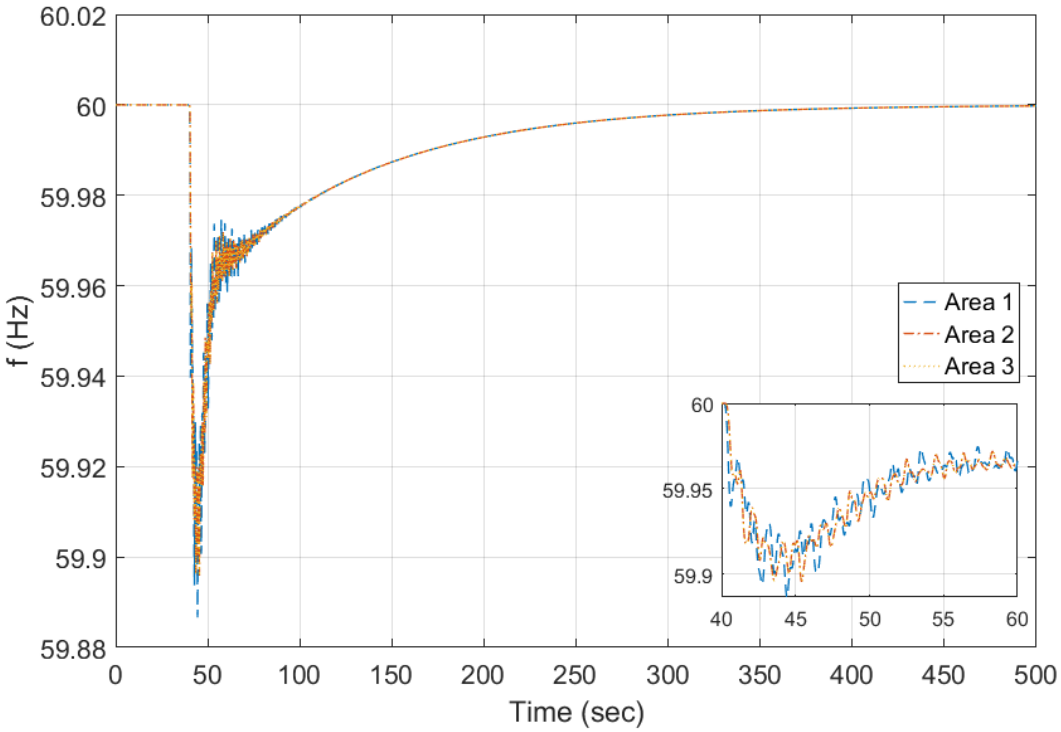


Figure 4-22 Frequencies in Each Area of the System

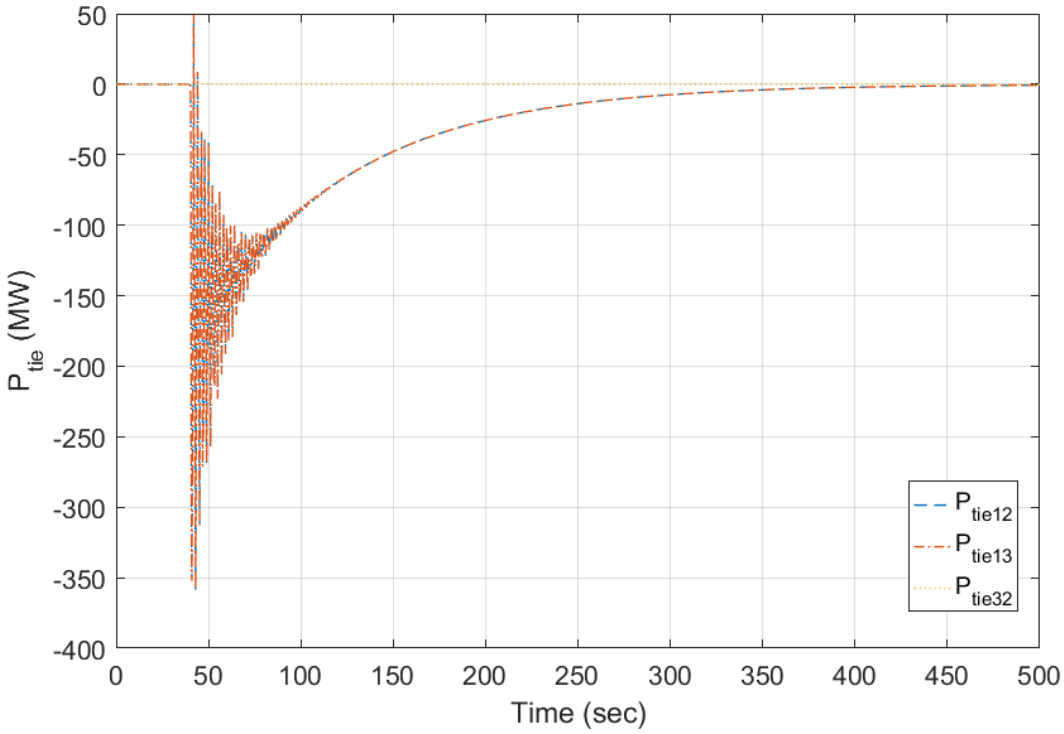


Figure 4-23 Tie-line Flows of Each of the Three Lines Connecting the Three Areas

With the 3-area power system model validated under traditional secondary control, the new anticipatory controller can be augmented into the existing control system and tested for its performance.

4.4. Simulation Studies and Results

To examine the performance of anticipatory secondary control in the 3-area power system, the controller must be integrated into the power system model. In this section, the integration of the MPC controller into the power system is first presented. Then, the performance of anticipatory secondary control with an MPC controller was examined in 1) a single area with a step disturbance, with and without errors in the forecast 2) multi-area response with a step disturbance, with and without broadcasts of a possibly erroneous

anticipated disturbance in an area, and 3) multi-area response with a 5-minute forecast, with an assumed model to interpolate the disturbances between the given 5-minute forecast.

4.4.1. Complete Controller and Plant Formulation

With the construction and validation of the power system models in the previous section, the controller and plant can now be completely defined for a multi-area system. The single-area system formulation will also be shown as a simplified version of the full, multi-area system formulation. Recall the MPC formulation from Section 4.2.2, repeated here for convenience:

$$\min_{f_{set_1} \cdots f_{set_m}} J = \left(\sum_{i=1}^{p} Q_f \left(\Delta f_{sys}\right)^2 + Q_{ptie} (\Delta P_{tie}^m)^2 \right) + \left(\sum_{j=1}^{m} R_f (f_{set})^2 + S_f \left(\dot{f}_{set}\right)^2\right) \quad (4\text{-}36)$$

s.t.

$$X_{i+1} = AX_i + BU_i \quad (4\text{-}37)$$

$$Y_i = CX_i + DU_i \quad (4\text{-}38)$$

$$-0.1 \leq f_{set} \leq 0.1 \quad (4\text{-}39)$$

where the MPC controller step size is set at 2 seconds for this study.

Before presenting the linear state-space plant model, there must be some discussion regarding the estimation of unknown, unmeasured variables of the error in the forecasted disturbance ε_d and the error in tie-line flow ε_{ptie}. Recall the discussion in Section 4.2.2(i), where the process noise covariance matrix Q_{kal} and the measurement noise covariance matrix R_{kal} determines the 'trust' placed in the process model or the measurements themselves. Thus, a row in Q_{kal} with large values will result in a corresponding state estimate that follows the measurements closely, rather than the process model. Because the errors ε_d and ε_{ptie} are unknown and unmeasured, the Kalman filter is used to estimate ε_d and ε_{ptie} by introducing these inputs as actual states in the system and by making the

corresponding diagonal noise entries in the process noise covariance matrix Q_{kal} very large. To estimate the forecasted disturbance error and the tie-line flow error, the plant state-space matrix is augmented with ε_d and ε_{ptie} such that

$$\begin{bmatrix} X_{i+1} \\ \varepsilon_{d,i+1} \\ \varepsilon_{ptie,i+1} \end{bmatrix} = \begin{bmatrix} A & B_d & B_{ptie} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_i \\ \varepsilon_{d,i} \\ \varepsilon_{ptie,i} \end{bmatrix} + \begin{bmatrix} B \\ 1 \\ 1 \end{bmatrix} U_i \qquad (4\text{-}40)$$

where the last two elements of U_i are the white noise inputs w_d and w_{tie}, which have a zero mean and some variance σ_d and σ_{ptie} which will be defined later. These white noise inputs do not actually affect the plant model calculations directly as they are zero mean, but they are present to represent the noise in the Q_{kal} matrix for the Kalman filter. In this way, the variances σ_d and σ_{ptie} of the white noise inputs w_d and w_{ptie} match the diagonal entries in Q_{kal}, which causes the estimates of ε_d and ε_{ptie} to be heavily weighted by measurements and not the process model (which shows that ε_d and ε_{ptie} stay constant over time). Thus, although the errors ε_d and ε_{ptie} are inputs into the system conceptually, the actual implementation results in ε_d and ε_{ptie} as states of the system, being driven by white noise inputs w_d and w_{ptie}. In this work, the variances of w_d and w_{ptie} are both set to 25, in contrast to a noise variance of 1 on every other state in the system, unless otherwise noted. This produces a diagonal Q_{kal} and R_{kal} matrix, which are identity matrices with the exception of the two elements $\sigma_d = 25$ and $\sigma_{ptie} = 25$. Sensitivities to these parameters are shown in Appendix B V.

The linear state-space plant model will now be presented. A full overview of the state variables for the multi-area system model are shown in Table 4-4.

Table 4-4 State-space Model State Variables, in X

State Variable	# of State Variables	Used in Single-area Formulation	Description
$V_{p,i}^0$	4	Yes	Valve position at bus i within own area
$P_{m,i}^0$	4	Yes	Mechanical power output of turbine at bus i within own area
$\Delta\omega_i^0$	6	Yes	Speed deviation of turbine model at bus i within own area
P_{tie-ij}^0	7	Yes	Tie-line flows between connected buses i and j within own area
P_{tot}^0	1	Yes	AGC area governor set point
P_{tie}^{ij}	3	No	Net tie-line flows between areas i and j
V_p^j	2	No	Valve position of equivalent turbine in area j
P_m^j	2	No	Mechanical power output of equivalent turbine in area j
$\Delta\omega^j$	2	No	Speed deviation of equivalent turbine model in area j
P_{tot}^j	2	No	AGC area governor set point in area j
eACE	1	Yes	Measurement of the eACE through a low-pass filter
P_{tie}^m	1	No	Measurement of net tie-line flow within own area through a low-pass filter
ε_d	1	Yes	Error of the disturbance forecast in own area
ε_{ptie}	1	No	Error of the tie-line flow in own area

Each responsive generator is modeled with a classic steam-electric model as shown in Figure 4-5, so the state variables $V_{p,i}^0$, $P_{m,i}^0$, and $\Delta\omega_i^0$ are present at each of the four buses with responsive generation. Bus 2 and bus 3 had no responsive generation, so only $\Delta\omega_i^0$ was represented which models the load with inertia at the bus. In addition, the power flow on the transmission lines were modeled using the formulation shown in (4-29), integrating the differences between $\Delta\omega_i$ and $\Delta\omega_j$ between connected buses i and j. A single area has 7 transmission lines between buses within its own area, so this adds the 7 states P_{tie-ij}^0, where i and j are connected buses as shown in Figure 4-4. The traditional integral control of AGC

which minimizes frequency deviations in the system also has a state P_{tot}^{0}, the area governor set point. Neighboring areas were modeled with a single equivalent classical steam-electric turbine representation connected to a simple integral control as its AGC system, as it was assumed that the details of neighboring areas would not be known. Thus, each of the two neighboring areas each had one state for valve position V_p^j, the mechanical power from the turbine P_m^j, the speed deviation of the turbine $\Delta\omega^j$, and the AGC area governor set point P_{tot}^{j}. Both the eACE and the net tie-line flow within the area P_{tie}^{m} was measured through individual low-pass filters, with time constants of 0.05 Hz and 0.1 Hz, respectively. Lastly, the errors of the disturbance forecast ε_d and the tie-line flow error ε_{ptie} discussed earlier in this section are the last two elements of the vector. All of these states add to a vector of size 37 x 1, which can be written as

$$X \triangleq [V_{p,1}^{0} \dots V_{p,6}^{0}, P_{m,1}^{0}, \dots, P_{m,6}^{0}, \Delta\omega_1^0, \dots, \Delta\omega_6^0, P_{tot}^{0}, P_{tie}^{12}, P_{tie}^{23}, P_{tie}^{13}, \dots \quad (4\text{-}41)$$

$$V_{p,i}^{1}, V_{p,i}^{2}, P_m^1, P_m^2, \Delta\omega^1, \Delta\omega^2, P_{tot}^{1}, P_{tot}^{2}, \text{eACE}, P_{tie}^{m}, \varepsilon_d, \varepsilon_{ptie}]^{\mathrm{T}}$$

making the original system matrix A of dimension 37 x 37. The single-area formulation is a limited case of this formulation, which has a size of 24 x 24 with the variables as denoted in Table 4-4.

The outputs Y of the plant are simply the measured deviation of frequency of the system, measured deviation of net tie-line flow out of the area (measured through a low-pass filter), the turbine power output (said to be equivalent to the electronic power output), and the measured eACE (measured through a low-pass filter). This is shown in Table 4-5.

Table 4-5 State-space Model Output Variables, in Y

State Variable	# of Variables	Description
Δf_{sys}	1	Deviation of system frequency, measured at bus 6
ΔP_{tie}^{m}	1	Deviation of net tie-line flow
P_{turb}^{i}	5	Turbine model power output at bus i within own area
eACE	1	Measurement of the eACE through a low-pass filter

So then, the plant output vector is written as

$$Y \triangleq \left[\Delta f_{sys}, \Delta P_{tie}^{m}, P_{turb}^{1}, P_{turb}^{2}, P_{turb}^{4}, P_{turb}^{5}, P_{turb}^{6}, \text{eACE}\right]^{\text{T}} \quad (4\text{-}42)$$

which is 8 elements long. Recall that bus 2 has a direct electrical injection of generation, which is why P_{turb}^{2} is present without a classic steam-electric turbine representation at the bus. The single-area simulations use an output Y that excludes the ΔP_{tie}^{m} term in (4-41).

The composition of the plant input U depends on whether or not inter-area communication is implemented. Table 4-6 shows all of the variables that can be present in the input vector of the plant model.

Table 4-6 State-space Model Input Variables, in U

State Variable	# of Variables	Description
$\hat{d}$	1	Disturbance forecast for own area
f_{set}	1	Frequency set point within own area
f_{set}^{i}	2	Frequency set point in area i
$\hat{d}^{i}$	2	Disturbance forecast in area i
w_d	1	White-noise input for disturbance error
w_{ptie}	1	White-noise input for tie-line flow error

If inter-area communication exists, then the full vector is

$$U \triangleq \left[\hat{d}, f_{set}, f_{set}^{1}, f_{set}^{2}, \hat{d}^{1}, \hat{d}^{2}, w_d, w_{tie},\right]^{\text{T}} \quad (4\text{-}43)$$

which is 8 elements long, resulting in a B matrix of size 37 x 8 and a D matrix of size 8 x 8, for the multi-area model. If inter-area communication does not exist, then the input is written as

$$U \triangleq \left[\hat{d}, f_{set}, w_d, w_{tie}\right]^{\mathrm{T}} \quad (4\text{-}44)$$

which is 4 elements long, resulting in a B matrix of size 37 x 4 and a D matrix of size 8 x 4 for the multi-area model. The single-area formulation is limited to the U in (4-43) without w_{tie}. The entire 3-area power system was built in Simulink and the MATLAB linearization function was used to create the A, B, C, and D matrices used in this study.

Now, the objective function of the MPC optimization problem will be discussed. The weights R_f and S_f are used to constrain the movement of the frequency set point such that overly aggressive action does not occur. For example, if the controller were to act aggressively for an imperfect forecast, responsive turbine outputs in the system might swing needlessly for a disturbance that was predicted incorrectly. Also, penalties on the movement prevent oscillatory movements in the controller output. Throughout this work, the values of $R_f = 0.15$ and $S_f = 0.75$ were used to produce a smoother controller output that responds moderately to predicted disturbances in the power system, where the output f_{set} is given by the controller in units of Hz.

Measured frequency deviations (measured in Hz) are penalized by the factor Q_f, while tie-line deviations (measured in system pu base) were penalized by the factor Q_{ptie}. Values of $Q_f = 0.75$ and $Q_{ptie} = 3.5$ were the default values used in this study, as those values resulted in a reasonably smooth response by the controller as well as a 'balanced' response given tie-line and frequency deviations. That is, given a disturbance outside of

the controlled area, the effective frequency reference was kept at 60 Hz and the f_{set} controller output was near zero, which mirrors the response of conventional AGC (where only disturbances within its own area require substantial movement of governor set points). However, other values of Q_f and Q_{ptie} are explored in the multi-area simulations, and the weights specific to the study will be presented within the multi-area study.

To reiterate, for the single-area simulations, the formulation is a limited case of the full formulation shown in (4-36) to (4-39) with no inter-area communication. The objective function has no $Q_{ptie}(\Delta P_{tie}^{m})^2$ term, and the state-space formulation contains no inter-area terms, such as the inter-area tie-line flows, which was discussed throughout this section.

4.4.2. Single-area Step Disturbance Simulations

Simulations for a single area power system with anticipatory secondary control are presented in this subsection, comparing the performance of MPC-based secondary control against conventional AGC, an external PI controller, and preplanned human frequency set point adjustments. The effects of perfect disturbance forecasts, erroneous disturbance forecasts, and differing amounts of responsive generation on the operational performance of anticipatory secondary control are presented.

The PI controller was chosen as an alternative real-time external controller as it is robust, easy to implement, and is well-characterized as a traditional method of control. However, unlike MPC-based control, PI control is unable to predict future plant response. The external PI controller uses the system frequency deviation as its input, and it adjusts the area frequency set point in response to errors in frequency such that $f_{set}(t) = K_p^{expi}\Delta f_{sys}(t) + K_i^{expi}\int_0^t \Delta f_{sys}(\tau)d\tau$, where the proportional gain K_p^{expi} and the integral

gain K_i^{expi} are set to ensure the fastest settling time with minimum overshoot of the measured frequency past 60 Hz. Calibration of the external PI set-point controller was subject to 1) limiting its frequency set-point output to the 59.9 Hz – 60.1 Hz limits applied to the MPC controller, 2) allowing a maximum overshoot comparable to that of MPC, and 3) tuning for the fastest settling time. These criteria resulted in a choice of $K_p^{expi} = 0.5$ and $K_i^{expi} = 0.001$ for the gains of the external PI controller. Note that this controller is external to the conventional AGC system, and works in addition to the PI controller which is contained inside conventional AGC.

Preplanned patterns for the frequency set-point f_{set} were chosen as other benchmark comparisons because of the ability for preplanned frequency set point patterns to be implemented in modern power systems. When an event is predicted in advance, it is possible to create a frequency set-point plan that will minimize the system frequency deviations. Although this allows for some fine tuning of the frequency set-point pattern prior to the event, the pattern is assumed to remain unchanged after it is finalized. There is no feedback control using preplanned control, so incorrect modeling of the power system as well as errors in the forecast can cause significant frequency deviations.

Metrics to characterize the performance of the controllers include the frequency settling time, the maximum frequency deviation, the average frequency deviation Δf_{sys} over the simulation horizon, and the integral of the ACE value over the simulation horizon. The settling time is defined as the amount of time between the disturbance and the frequency settling to within 0.0035 Hz (this is approximately 5% of the maximum frequency deviation under conventional control, for the step disturbance), and the

maximum frequency deviation is the frequency nadir, the minimum measured frequency after the disturbance. The metrics of the average frequency deviation and the integral of ACE are inspired by other works regarding the improvement of secondary control systems [54].

The step disturbance anticipated throughout these simulations was a 250 MW (0.01 pu system base) generation power drop, simulating a generation trip or other sudden generation-related event by instantly dropping power within one simulation time step (0.02 seconds) at t = 30 seconds. The actual disturbance in a perfect forecast case matches the expected disturbance, while errors in the disturbance are elaborated upon in its corresponding subsection.

(i) Comparisons with the Perfect Forecast Case

Performance of the MPC controller, an external PI controller, preplanned human operator control actions, and a conventional AGC controller with no external adjustments are first compared with perfect forecasts, in this subsection. After this, the performance of the controllers with errors in both time and magnitude of the forecast are presented.

To establish a base case, a 0.01 pu sudden generation loss at bus 3 at t = 30 seconds was simulated for the single-area power system in Figure 4-4 under conventional AGC, with results similar to the simulation shown in Section 4.3.1. The frequency is shown in Figure 4-24, the responsive turbine output is shown in Figure 4-25, and ACE is shown in Figure 4-26. The settling time was 257 seconds, with a frequency nadir of 59.85 Hz, an average frequency deviation Δf_{avg} of 26.3 mHz, and an integrated ACE value of 32546 MW over the 400 second simulation window.

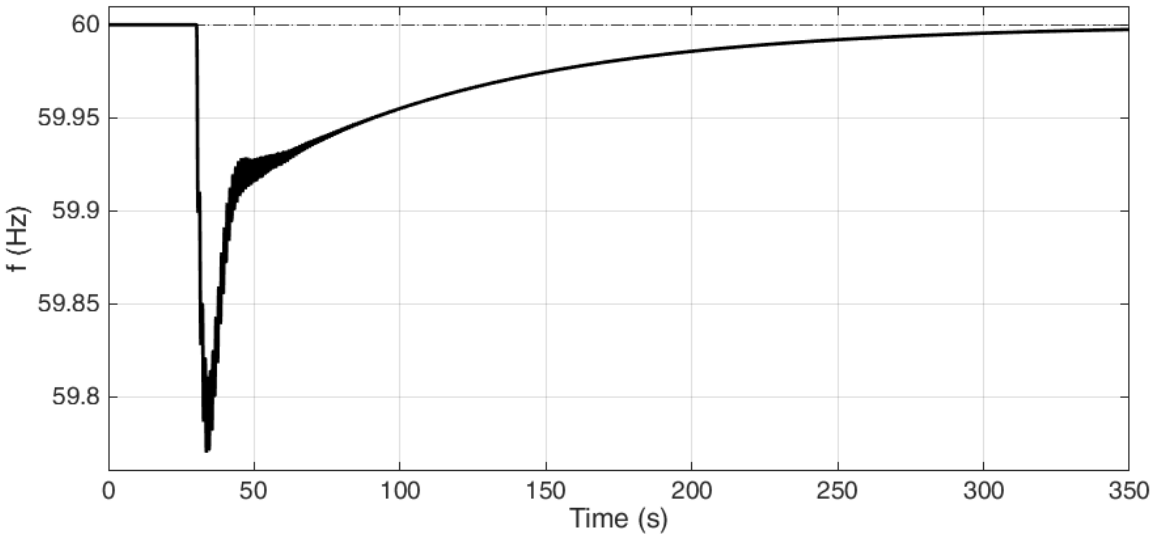


Figure 4-24 Frequency in a Single Area for a 0.01pu Drop in Generation

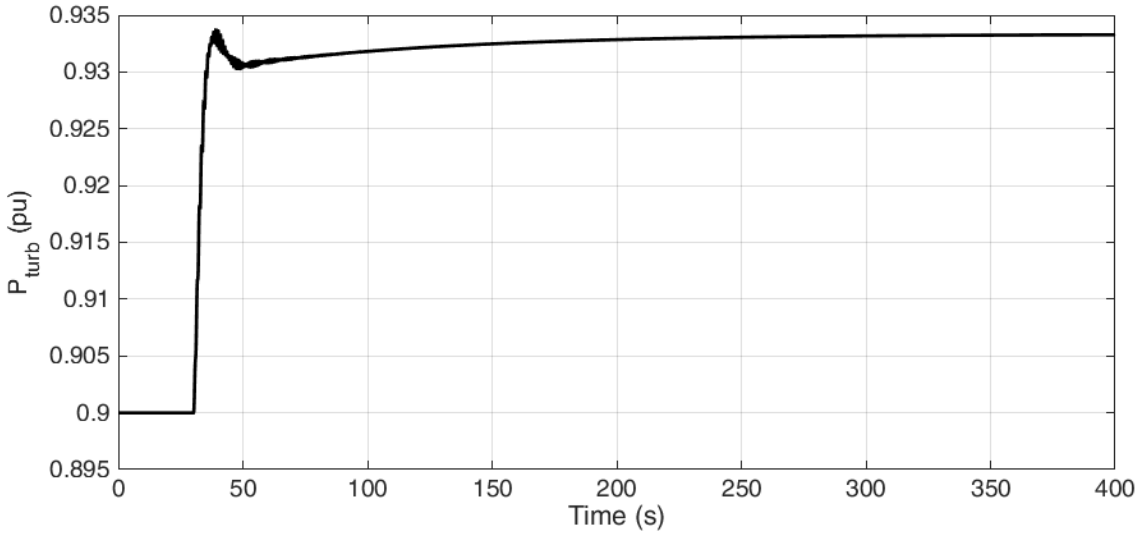


Figure 4-25 Responsive Turbine Output under Conventional Control

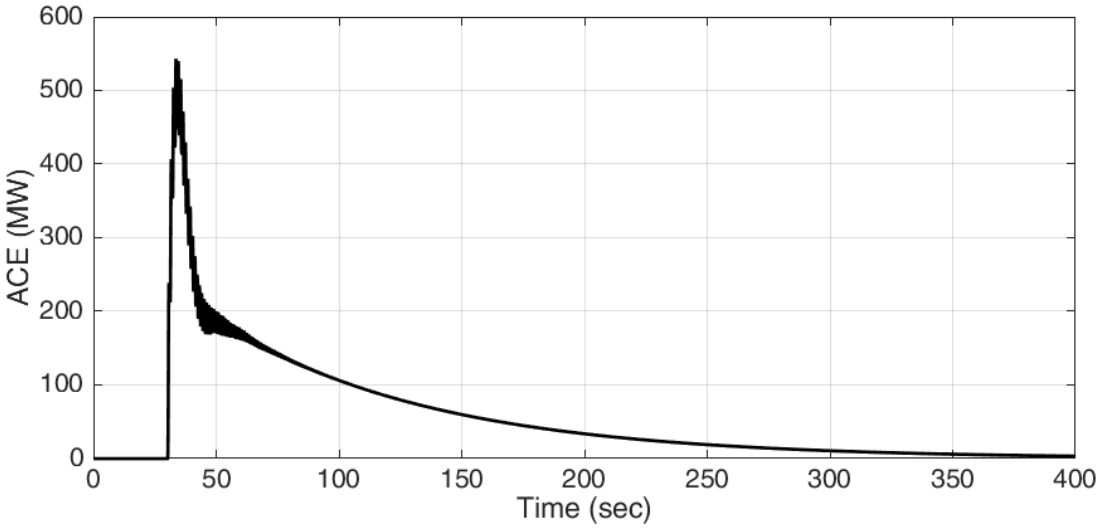


Figure 4-26 ACE under Conventional Control, with No Tie-line Flows

Anticipatory secondary control was then implemented to compare against the performance of conventional AGC. Results are also plotted for use of MPC without any anticipation, which is simply the use of MPC without any prior knowledge of a disturbance, other than what is measured from the system. Lastly, the results are also compared alongside an external PI controller. Results are shown in Figure 4-27, Figure 4-28 and Table 4-7.

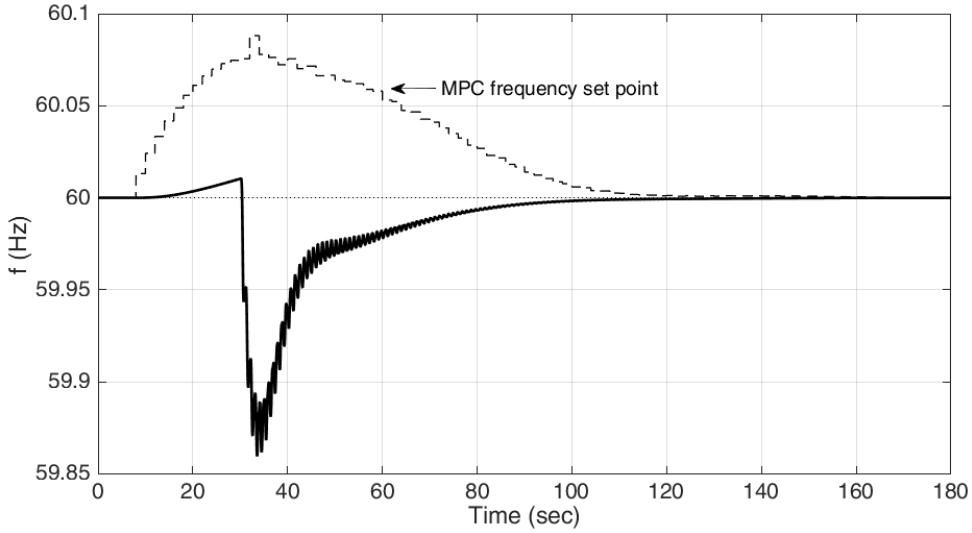


Figure 4-27 Detailed View of Frequency under Anticipatory Secondary Control

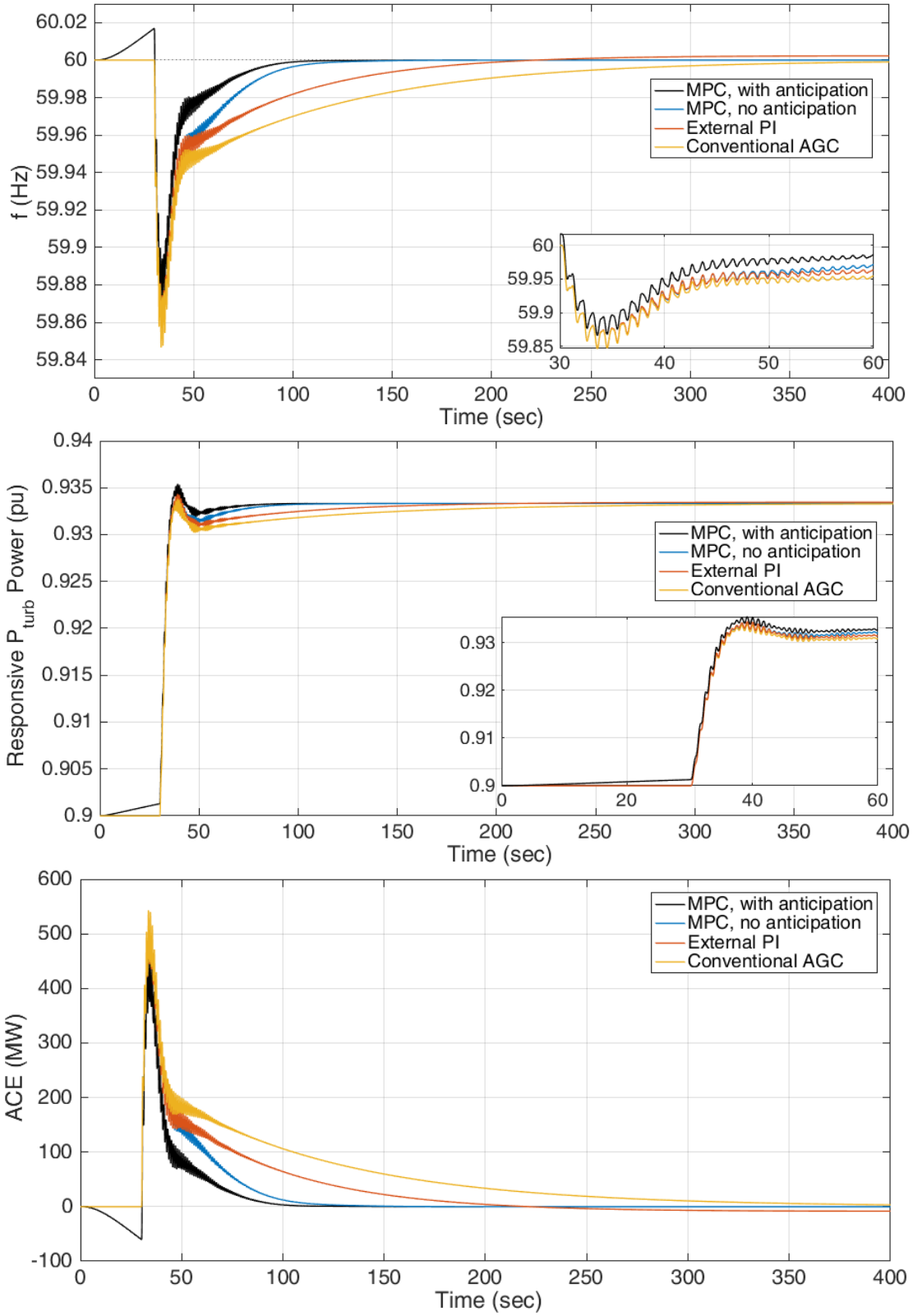


Figure 4-28 Frequency, Turbine Power, and ACE with MPC versus Conventional AGC

Table 4-7 Performance of Single-area System with Various Secondary Control Schemes

Control Type	Settling Time (sec)	Max Freq. Dev. (Hz)	Average Δf_{sys} (mHz)	Integrated ACE (MW)
Conventional AGC	257	0.153	15.5	21926
External PI	142	0.153	10.3	14545
MPC, no anticipation	70	0.153	6.3	8940
MPC, with anticipation	55	0.133	4.6	6518

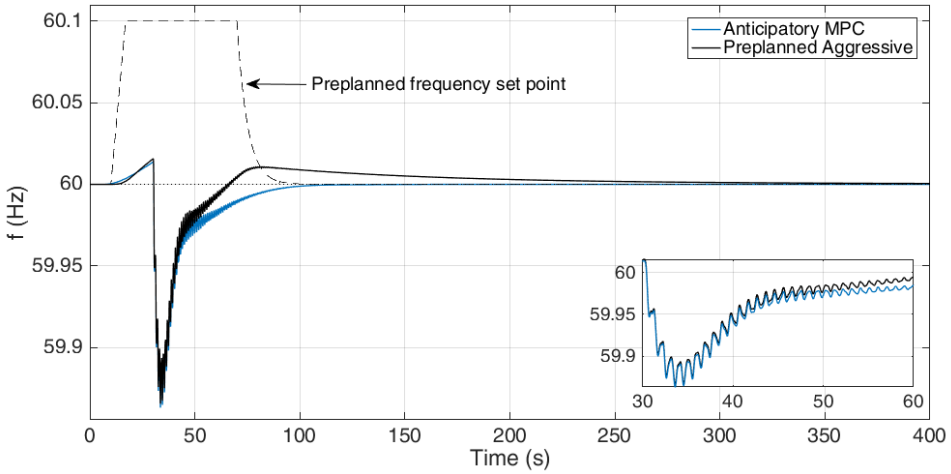
The results show that the power system operational performance is the best under MPC control with the metrics considered. Frequency is shown to recover more quickly with the MPC controller than in the cases with conventional AGC and external PI control, and the maximum frequency deviation under anticipatory MPC control was lower than that of the other compared control methods, as anticipatory control was able to bring the frequency higher in anticipation of the sudden drop in generation. As there is no tie-line flow in this single-area system, the integrated ACE value is directly related to the average frequency deviation in the system. Also, although the external PI controller seems to perform better than conventional AGC, it results in an overshoot of the system frequency, whereas conventional AGC and MPC control does not. See Appendix C I for a simulation with perfect forecasting, but with white noise integrated into the load of the power system.

As the system can be prepared for an anticipated event by raising frequency before an anticipated decrease in generation (or increase in load), anticipatory secondary control can be compared with reasonable, preplanned human operator actions as well. Three types of preplanned actions are considered: 1) aggressive action, 2) conservative action, and 3) near optimal action. Each preplanned action path was restricted to linear or exponential ramps, not allowing for complex shapes of the frequency set point such as that shown in Figure 4-27 for MPC. Each of the three preplanned actions are meant to mimic the possible range of behaviors that would reasonably be seen in a case where the frequency set point

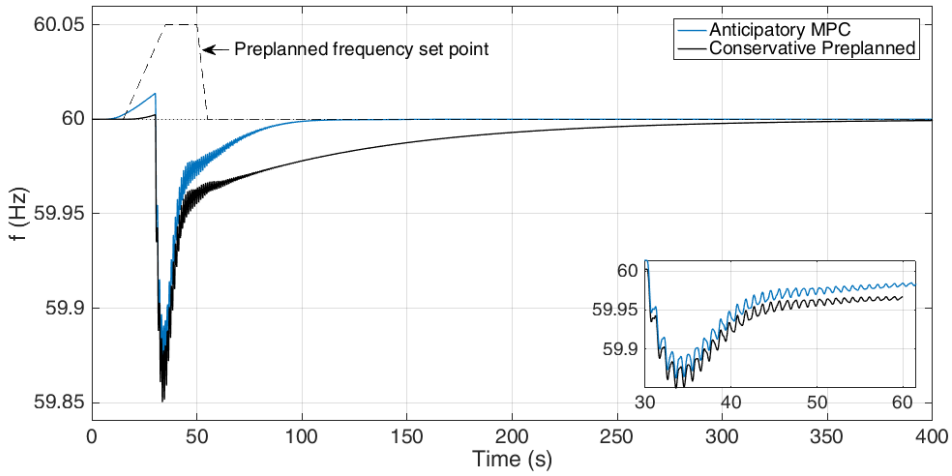
would be directly manipulated by a human. The comparisons with three different types of preplanned operator schemes are shown in Figure 4-29 with the frequency set point paths drawn on each frequency plot, and the performance comparison between them and MPC is shown in Table 4-8. Further results to accompany these figures, such as turbine outputs and ACE comparisons, are shown in Appendix C I.

Table 4-8 Performance of Single-area System with Various Preplanned Schemes

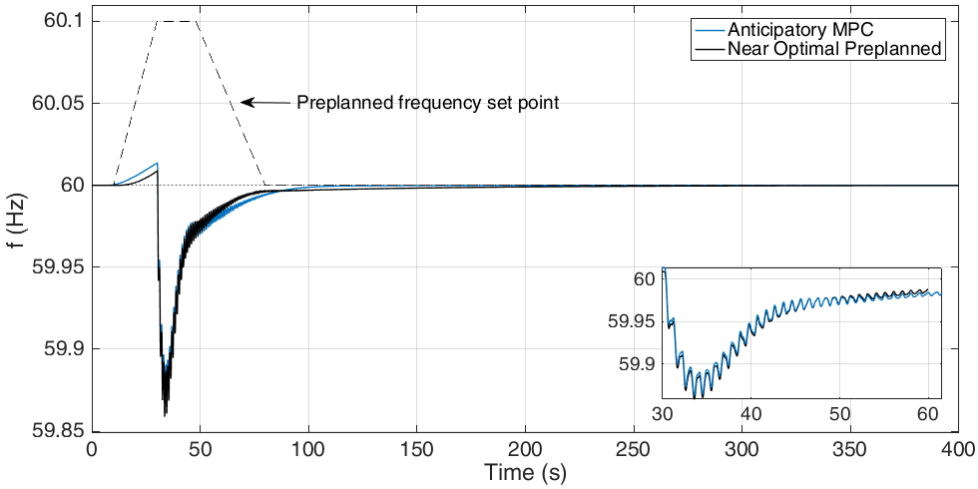
Control Type	Settling Time (sec)	Max Freq. Dev. (Hz)	Average Δf_{sys} (mHz)	Integrated ACE (MW)
Anticipatory MPC	55	0.133	4.6	6518
Aggressive, Preplanned	155	0.134	6.3	8926
Conservative, Preplanned	230	0.150	12.1	17208
Near Optimal, Preplanned	60	0.141	4.9	6999



(a)



(b)



(c)

Figure 4-29 Frequency Results for MPC versus Various Operator's (a) Aggressive, (b) Conservative, or (c) Near Optimal Preplanned Actions

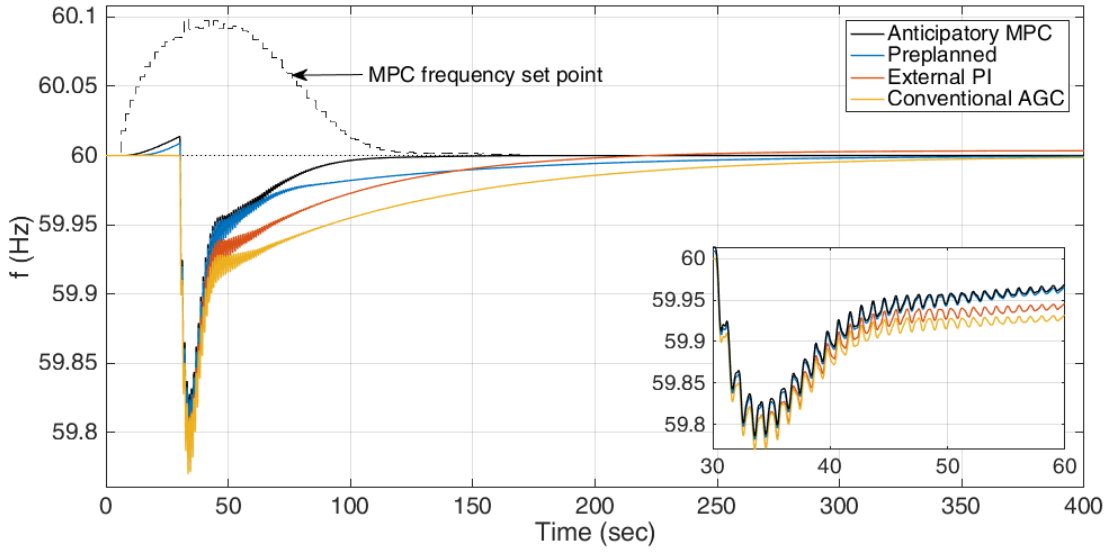
Anticipatory MPC shows better performance than the preplanned frequency set point responses. The near optimal preplanned response will be used to represent the preplanned response in future simulations with errors in the disturbance forecast. Unless otherwise noted, anticipatory MPC control will be referred to as MPC control, as non-anticipatory control is not discussed further.

(ii) Comparisons with the Erroneous Forecast Case

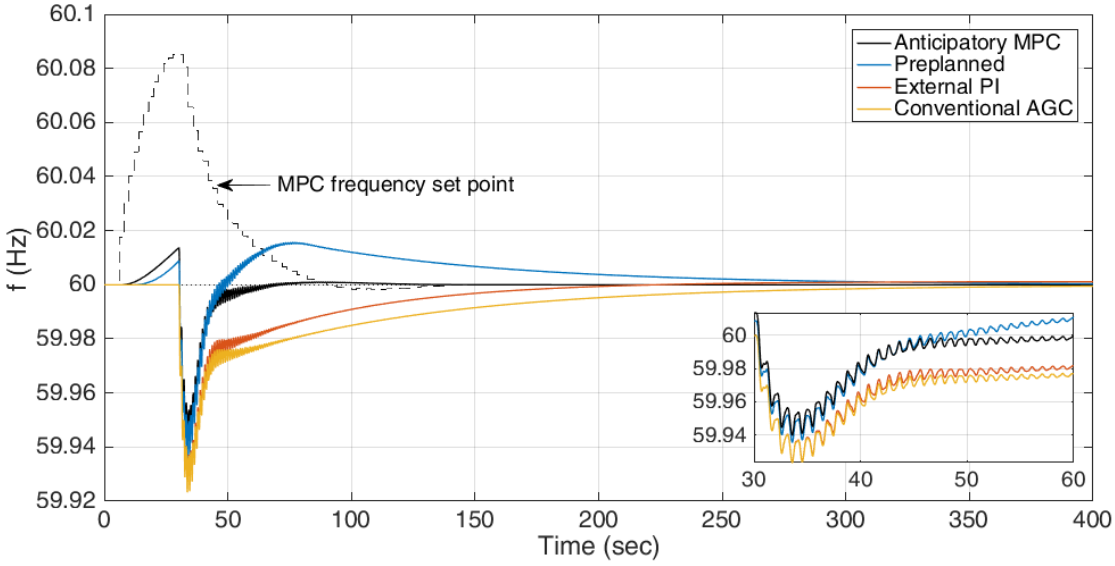
Uncertainties in the disturbance forecast, both in time and magnitude of the event, are likely to exist. The simulations are split into two groups: one study on the effects of magnitude errors on operational performance, and one study on the effects of temporal errors on operational performance. Anticipatory secondary control is compared against preplanned control, external PI control, and conventional AGC. In these simulations, near-optimal action was chosen to represent preplanned action.

First, the study concerning magnitude errors is presented. It is assumed that a ±50% error occurred in the disturbance forecast, where the disturbance forecast was 0.01 pu, but the true disturbance was of a 0.005 pu magnitude (such that the actual disturbance is -50% of the forecast) or 0.015 pu magnitude (such that the actual disturbance is +50% of the forecast). Results for simulations with magnitude errors in the disturbance forecast are shown in Figure 4-30 and Table 4-9, comparing conventional control, external PI control, preplanned control, and anticipatory MPC control. For preplanned control, the exact profile of the frequency set point as seen in Figure 4-29(c) is used. Only the frequency set point trace of MPC is shown in Figure 4-30, as the actions of preplanned control do not change with forecasting errors and controller actions of external PI control are simply scaled

versions of the frequency deviation. Plots of turbine output and ACE are shown in Appendix C III.



(a)



(b)

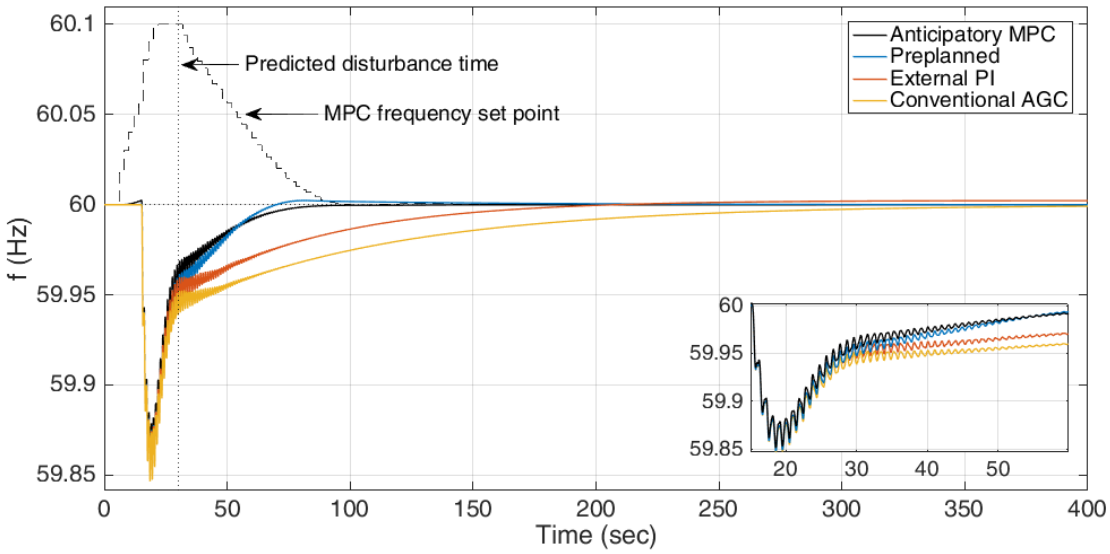
Figure 4-30 Frequency Results from Various Controllers in When (a) a 250 MW Generation Drop was Predicted for an Actual 375 MW Drop and (b) a 250 MW Drop was Predicted for an Actual 125 MW Drop

Table 4-9 Performance of Various Control Schemes with Forecast Errors of ±50%

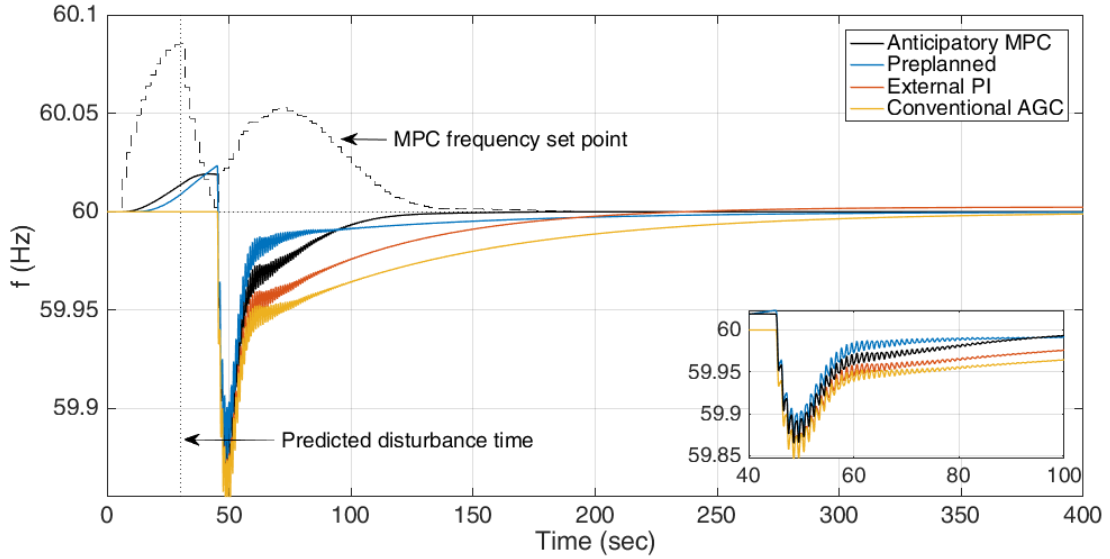
	Control Type	Settling Time (sec)	Max Freq. Dev. (Hz)	Average Δf_{avg} (mHz)	Integrated ACE (MW)
Actual is 1.5x Forecast	Conventional	292	0.230	23.2	32889
	External PI	155	0.229	15.4	21815
	Preplanned	212	0.218	12.8	17961
	MPC	70	0.214	8.1	11447
Actual is 0.5x Forecast	Conventional	197	0.077	7.7	10963
	External PI	115	0.076	5.1	7272
	Preplanned	178	0.064	5.5	7585
	MPC	26	0.060	1.6	2241

The results now show a greater performance divide between MPC and preplanned control than in the perfect disturbance forecast case. Preplanned control does not adjust to the errors in the forecast, so its performance is much worse than the other methods with external control. External PI control outperforms preplanned control in many metrics, as the PI controller was unaware of any forecast to begin with and so only reacts to the disturbance as seen in the system. MPC control consistently performs better than the other control methods, as it quickly changes its behavior after the disturbance occurs from the measured generation output, which is then passed to the estimator within the MPC module. This allows MPC to adjust the frequency set point to minimize any further deviations in the system.

The next set of simulations involve a timing error in the disturbance forecast, such that the actual disturbance occurs 15 seconds before or after the forecasted disturbance time. The magnitude of the disturbance remains the same in these simulations (0.01 pu system base). The frequencies of the system are shown in Figure 4-31 and Table 4-10, and the responsive turbine power output and ACE plots are shown in Appendix C III.



(a)



(b)

Figure 4-31 Plots of System Frequency for the MPC, PI, and Preplanned Near-optimal Controller with a Forecast Error, where the Actual Event Happens (a) Early, 15 Seconds Before the Predicted Time and (b) Late, 15 Seconds Past the Predicted Time

Table 4-10 Performance of Various Control Schemes with Temporal Forecasting Errors of ± 15 seconds

	Control Type	Settling Time (sec)	Max Freq. Dev. (Hz)	Average Δf_{sys} (mHz)	Integrated ACE (MW)
Disturbance -15 sec of Forecast	Conventional	257	0.153	15.5	21973
	External PI	142	0.153	10.4	14667
	Preplanned	49	0.152	5.7	8136
	MPC	55	0.149	5.0	7026
Disturbance +15 sec of Forecast	Conventional	257	0.153	15.5	21871
	External PI	142	0.153	10.3	14422
	Preplanned	133	0.127	6.1	8702
	MPC	65	0.135	6.0	8468

Timing errors in the disturbance forecast result in less of a gap between preplanned and MPC control than magnitude errors, as preplanned control still raises the frequency set point by the same pattern and eliminates the frequency error relatively quickly in the simulation time window, regardless of when the disturbance occurs. The MPC controller, on the other hand, reacts to the disturbance (or lack of a disturbance) by changing its frequency set point path. Although the ramping of the frequency set point is limited by weights in the MPC optimization objective function, it is still seen that the trace of the MPC frequency set point varies significantly during the simulation. Increased or decreased weightings on frequency set point changes in the objective function of the MPC controller would exacerbate or lessen this behavior.

(iii) Simulations with Sweeps of Percent Responsive Generation

In addition to improved operational performance, the use of anticipatory secondary control may allow for less responsive generation to be kept online. That is, a power system with less responsive generation using anticipatory secondary control may be able to achieve the same results as a power system with a greater amount of responsive generation

under conventional AGC. The effect of responsive generation is studied in this subsection, where both the primary and secondary response are swept over a range of 10% to 30% of the system generation capacity. The same generation drop of 0.01 pu at t = 30 seconds is performed in this system as a perfectly forecasted disturbance for MPC. Results are shown in Figure 4-32 and Table 4-11, with plots of the turbine power output and ACE in Appendix C IV.

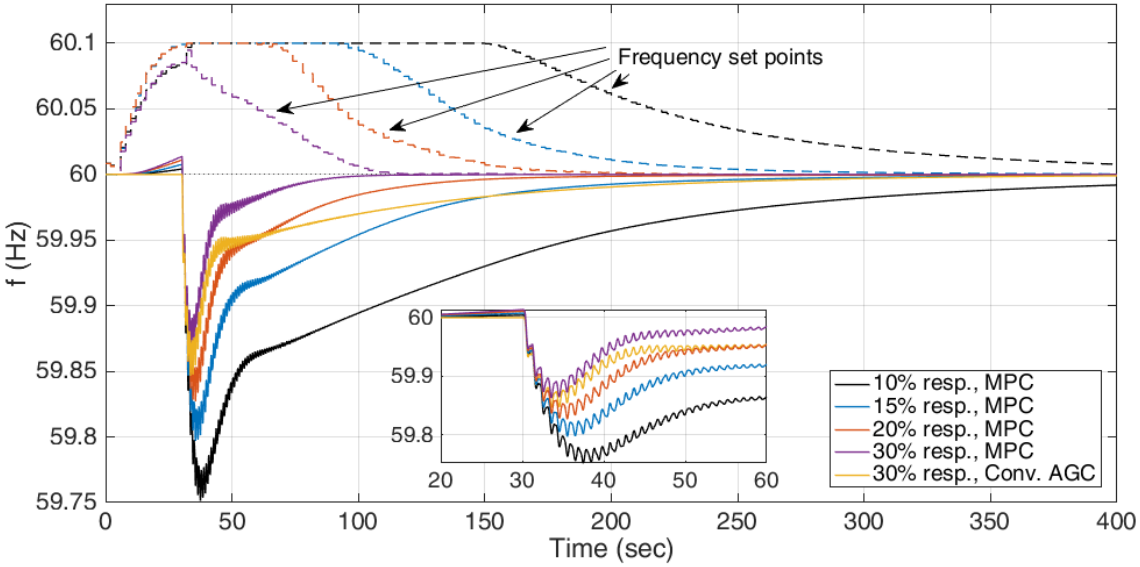


Figure 4-32 Frequency Comparison Between Conventional AGC and MPC under Different Amounts of Responsive Generation in the Power System

Table 4-11 Responsive Generation Sweep Results

Control Type	Settling Time (sec)	Max Freq. Dev. (Hz)	Average Δf_{sys} (mHz)	Integrated ACE (MW)
Conventional AGC 30% Responsive Gen.	257	0.153	15.5	21926
MPC 10% Responsive Gen.	370	0.249	53.1	30999
MPC 15% Responsive Gen.	227	0.203	22.0	17407
MPC 20% Responsive Gen.	112	0.173	11.2	11186
MPC 30% Responsive Gen.	55	0.133	4.6	6518

The results indicate that the use of anticipatory secondary control may require less responsive generation when compared with a similar case under conventional AGC, which could lead to a reduction in operating costs. In this scenario, MPC outperforms conventional AGC when the amount of responsive generation in the system is similar (at 30%), and conventional AGC in a system with 30% responsive generation outperforms MPC in a system with 10% responsive generation. However, the amount of responsive generation necessary in an MPC-controlled system to match performance with conventional AGC depends on the performance metric. For example, matching the settling time or integrated ACE metric between a system with conventional AGC and a system with MPC control would allow as little as 10-15% of responsive generation in the MPC-controlled system. However, matching the average frequency deviation would require somewhere between 15-20% of responsive generation, and matching the maximum frequency excursion would require somewhere between 20%-30% of responsive generation.

Power system operational performance for anticipatory secondary control has been compared against the performance of preplanned control, external PI control, and conventional AGC. In most metrics, anticipatory secondary control met or exceeded the performance of other control schemes, even with extremely inaccurate forecasts. However, these results are limited to a single-area system. Many power systems are connected with other systems through tie-lines, forming a multi-area power system with multiple, local secondary control schemes. The next subsection will present anticipatory secondary control within a multi-area power system.

4.4.3. Multi-area Step Disturbance Simulations

Simulations with anticipatory secondary control within a 3-area power system will be presented in this subsection. Distributed MPC will be considered here, with each of the three areas in Figure 4-20 controlled by a separate secondary control system. This is in contrast to centralized MPC, where one MPC controller is aware of a complete model and measurements from all areas in the interconnection. Distributed control with and without inter-area communication will be shown with a perfect forecast, then performance of anticipatory secondary control will be analyzed with an imperfect forecast. The controllers in each area are identical, as each area is identical.

(i) Perfect Disturbance Forecast

The performance of the distributed MPC controllers are studied with and without communication with a perfect disturbance forecast. The disturbance is the same 0.01 pu sudden decrease in generation as in the previous section, only occurring in Area 1 at bus 2. All inter-area communication assumed a 1 second delay in the exchange of information

between areas every 2 seconds, and the variables transmitted were the present frequency set point f_{set}^{a} as well as the predicted time series of the forecasted disturbance $\hat{d}^{a}$, from the present time t until 25 seconds from then $t + 25$, of area a. This resulted in 1 set point value and 26 forecasted disturbance points being transmitted every 2 seconds, from each area to the other two areas. All inputs must be defined from some time $t = 0$ to the prediction horizon p, so f_{set}^{a} is assumed to be held constant throughout the prediction horizon.

With the introduction of a net tie-line deviation term in the objective function of the MPC controller, the performance gains of communication rely heavily on the weights in the objective function for MPC given to the frequency or tie-line deviations terms. The default balanced weights of $Q_f = 0.75$ and $Q_{ptie} = 3.5$ are used to weight the frequency and tie-line deviations in the objective function of the MPC controller in all areas, though other weights are explored later in this subsection. Further comments on determining Q_f and Q_{ptie} weights are discussed in Appendix B VI. Figure 4-33 and Figure 4-34 show the frequency and net tie-line flows as seen from Area 1, with more plots contained in Appendix C V. Metrics such as the integrated ACE are presented at the end of this subsection. Note that because Area 2 and Area 3 is identical, the tie-line flows between the areas is 0, so the net tie-line deviations are half of that shown in Figure 4-34 and so are not shown.

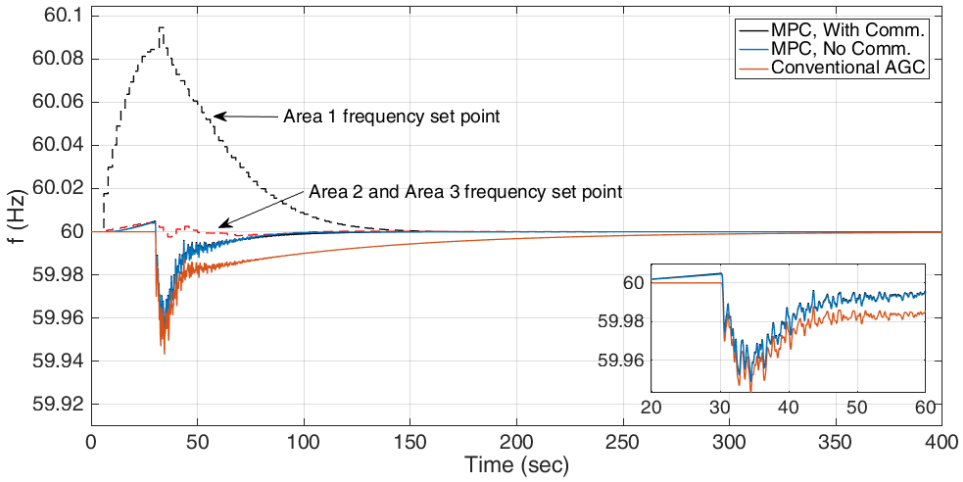


Figure 4-33 Frequency With and Without Inter-area Communication, Balanced

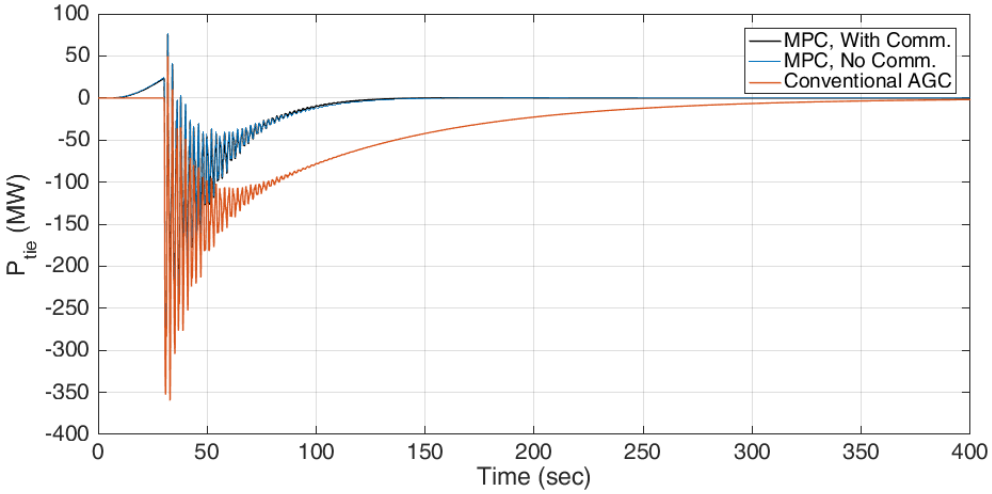
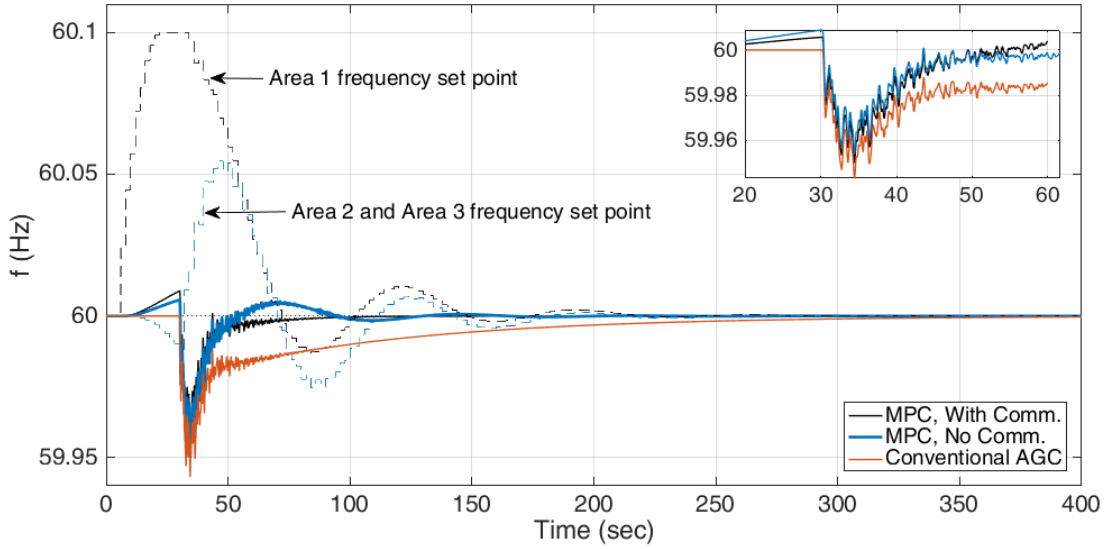


Figure 4-34 Tie-line Flow With and Without Inter-area Communication, Balanced

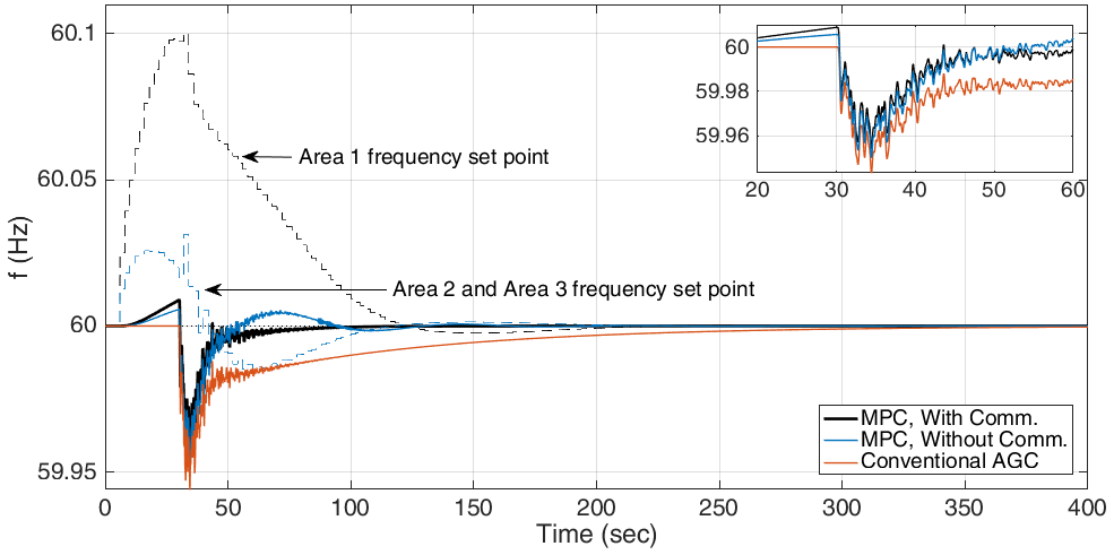
The frequency set points of Area 2 and Area 3 do not vary much in response to a disturbance outside of their areas, so the frequency set point traces of simulations with and without inter-area communication are nearly identical. Hence, the operational performance of the areas with and without inter-area communication with balanced weights on frequency and tie-line deviations are nearly identical.

Now unbalanced weights in the objective function of the MPC controller are introduced. If the value of Q_f is doubled to $Q_f = 1.5$, frequency deviations are now heavily

penalized. The controllers now have more incentive to minimize frequency deviations at the cost of some more tie-line flow or tie-line flow overshoot. The results are shown in Figure 4-35 and Figure 4-36.



(a)



(b)

Figure 4-35 Frequency Results with Increased $Q_f = 1.5$ Where (a) the Frequency Set Points are Shown for a Case Without Communication and (b) the Frequency Set Points are Shown for a Case with Communication

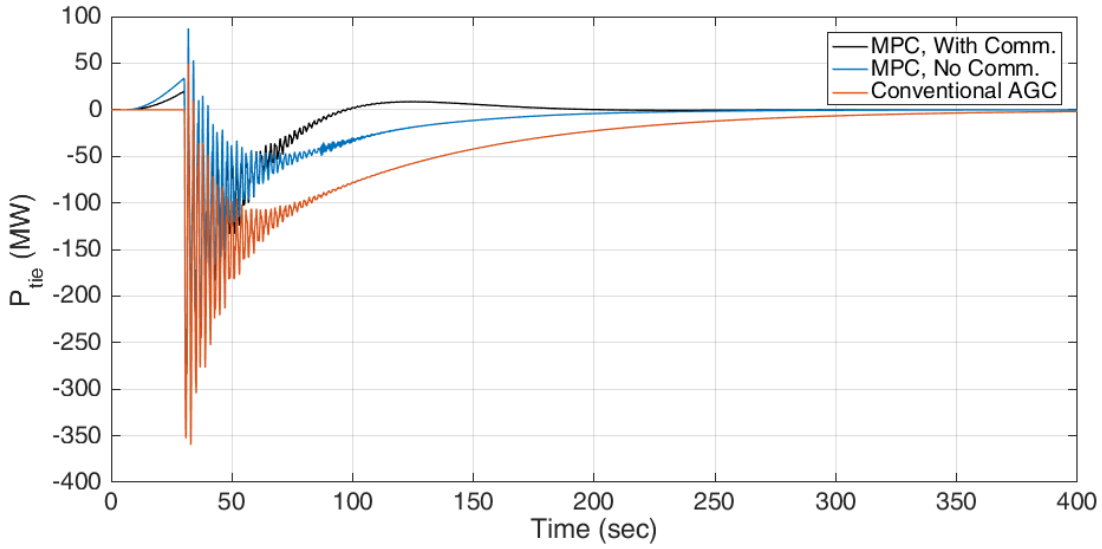
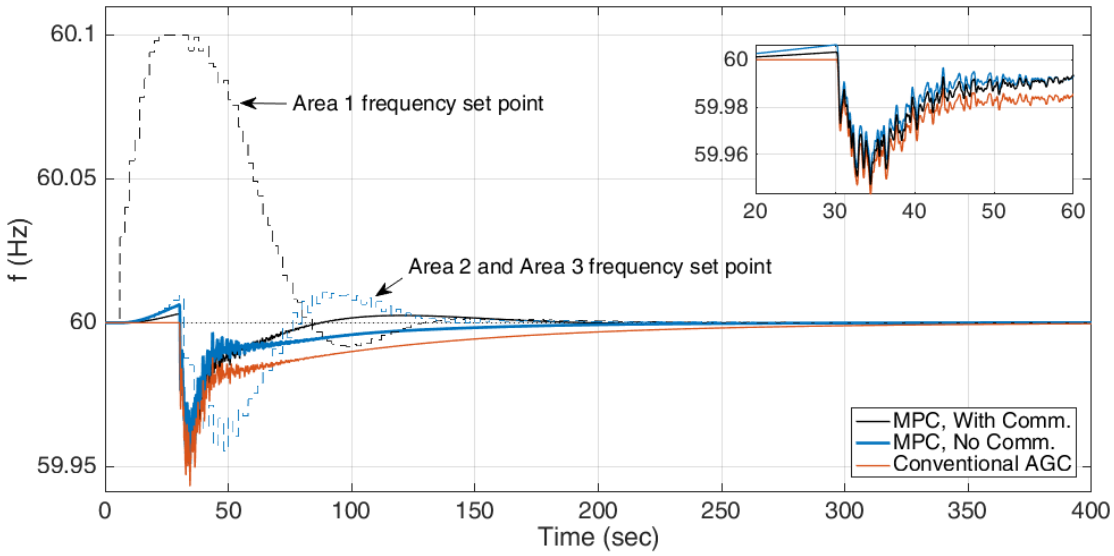


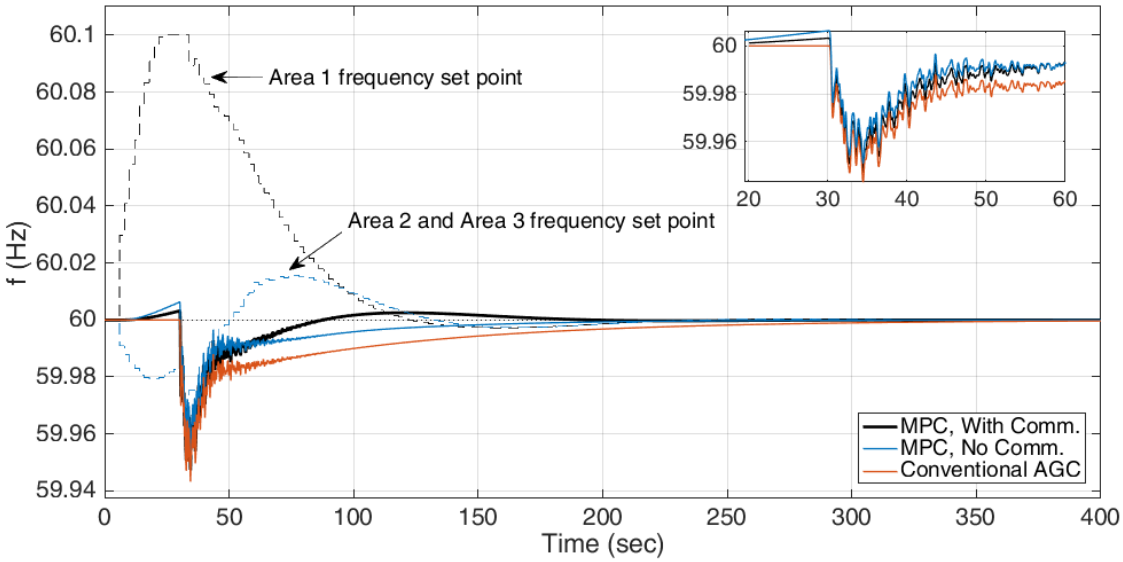
Figure 4-36 Net Tie-line Flows With and Without Inter-area Communication, Large Q_f

Now, controllers outside of the area work to bring frequency back to nominal, which results in some overshoot in the tie-line flows. In the scenario with inter-are communication, the frequency set points of all areas raise before the disturbance in anticipation of the disturbance, while without communication, Area 2 and Area 3 frequency set points act opposite to Area 1's frequency set point. This is because without knowledge of a future disturbance in Area 1, the controllers in other areas simply see an increase in frequency before the disturbance, and so act to bring the frequency back to 60 Hz. In the scenario without communication, the swinging frequency set points of all areas cause an oscillatory frequency trace after the disturbance, while in the scenario with communication, the controllers work together to bring the frequency quickly to nominal without any overshoot.

Restrictions on tie-line flow may be desired as well, so in the next set of simulations, $Q_f = 0.75$ and Q_{ptie} is doubled to $Q_{ptie} = 7$. Results are shown in Figure 4-37 and Figure 4-38, with more results in Appendix C V.



(a)



(b)

Figure 4-37 Frequency Results with Increased $Q_{ptie} = 7$ Where (a) the Frequency Set Points are Shown for a Case Without Communication and (b) the Frequency Set Points are Shown for a Case With Communication

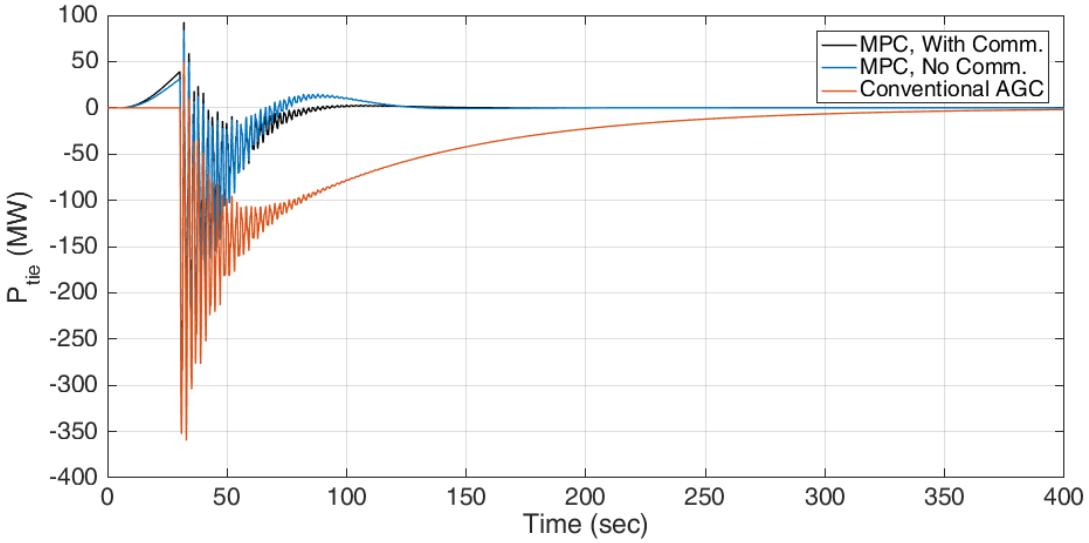


Figure 4-38 Net tie-line Flows With and Without Inter-area Communication, Large Q_{ptie}

Results of the large Q_{ptie} value mirror the results of a large Q_f, with tighter control seen in minimizing ΔP_{tie} for a scenario with communication. Table 4-12 shows the summary of frequency-related results for the simulations with perfect forecasting, varying Q_f and Q_{ptie}. Table 4-13 shows the summary of tie-line flow related metrics for the same simulations, where the P_{tie} Settling Time is defined as the time between the disturbance and the time when the net tie-line flow deviation settles to within 15 MW. Both tables contain the Integrated ACE metric for ease of comparison. As expected, the system showed better performance through the frequency metrics when Q_f was doubled in value, and better performance through the tie-line metrics when Q_{ptie} was doubled in value. However, in these scenarios with unbalanced Q_f and Q_{ptie}, the presence of inter-area communication had a large impact on the performance of the anticipatory controllers, as frequency set points in Areas 2 and 3 moved significantly in response to a disturbance in Area 1. In fact, in scenarios with unbalanced weights, inter-area communication is necessary to ensure robust operation with no frequency or tie-line flow overshoot. In other words, inter-area

communication is highly beneficial in power systems using anticipatory secondary control with unbalanced weights Q_f and Q_{ptie}, and does not change the performance metrics in systems with balanced weights Q_f and Q_{ptie}.

Table 4-12 Multi-area Results With and Without Communication, Frequency Metrics

	Control Type	Freq. Settling Time (sec)	Max Freq. Dev. (Hz)	Average Δf_{sys} (mHz)	Integrated ACE (MW)
	Conv. AGC	161	0.057	5.2	21974
Balanced $Q_f = 0.75$ $Q_{ptie} = 3.5$	MPC, no comm.	41	0.051	1.6	6886
	MPC, with comm.	41	0.051	1.6	6884
$\boldsymbol{Q_f = 1.5}$ $Q_{ptie} = 3.5$	MPC, no comm.	50	0.050	1.5	7818
	MPC, with comm.	31	0.046	1.2	6549
$Q_f = 0.75$ $\boldsymbol{Q_{ptie} = 7}$	MPC, no comm.	78	0.049	2.5	6105
	MPC, with comm.	43	0.053	2.1	6135

Table 4-13 Multi-area Results With and Without Communication, Tie-line Flow Metrics

	Control Type	P_{tie} Settling Time (sec)	Average ΔP_{tie} (MW)	Integrated ACE (MW)
	Conv. AGC	203	36.7	21974
Balanced $Q_f = 0.75$ $Q_{ptie} = 3.5$	MPC, no comm.	95	11.8	6886
	MPC, with comm.	95	11.7	6884
$\boldsymbol{Q_f = 1.5}$ $Q_{ptie} = 3.5$	MPC, no comm.	105	16.7	7818
	MPC, with comm.	55	12.3	6549
$Q_f = 0.75$ $\boldsymbol{Q_{ptie} = 7}$	MPC, no comm.	59	8.7	6105
	MPC, with comm.	45	8.0	6135

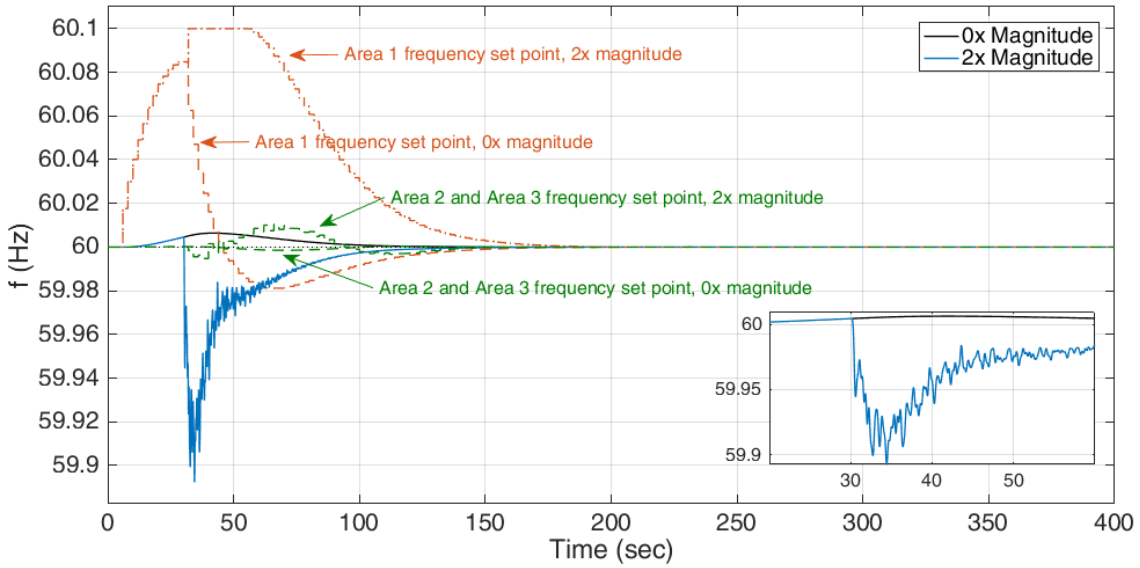
As shown in Figure 4-33, the anticipatory secondary control system where the Δf_{sys} term and ΔP_{tie} weights are balanced sees almost no difference in performance between scenarios with and without inter-area communication because of the similar frequency set points. However, if more weight is placed on either Δf_{sys} or ΔP_{tie}, inter-area communication improves the performance of the controllers immensely. More accurately, the lack of communication in these two situations (where the MPC controller weights either Δf_{sys} or ΔP_{tie} more than the other) hinders the ability of MPC to perform well.

These results were all achieved with the assumption of a perfect disturbance forecast. The effects of imperfect disturbance forecasts on distributed MPC control is investigated in the next subsection.

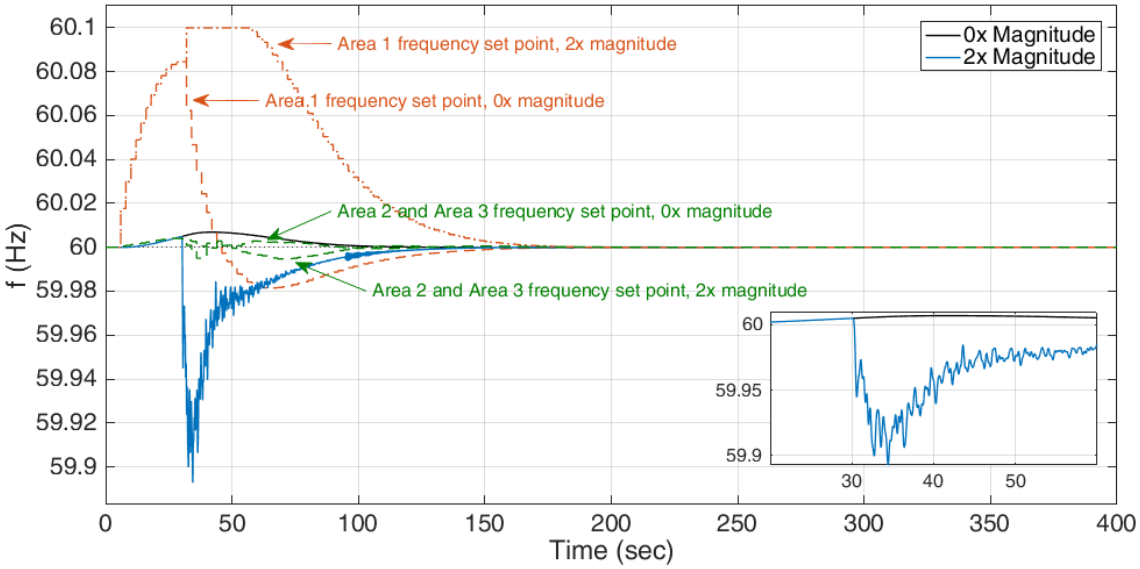
(ii) Imperfect Disturbance Forecasts

Large errors in the magnitude of the forecasted disturbance will be investigated, where a 0.01 pu disturbance is always anticipated. The actual disturbance ranges from 0 pu to 0.02 pu, which means that the actual disturbance was 0% of the forecast (referred to as 0x magnitude) or 200% of the forecast (referred to as 2x magnitude). First, simulations will be performed with balanced weights $Q_f = 0.75$ and $Q_{ptie} = 3.5$, such that the frequency set points of areas without a disturbance should only move minimally. Figure 4-39 shows the frequency and frequency set points of the extreme scenarios with either a 0 pu or 0.02 pu drop in generation when a 0.01 pu drop in generation was anticipated. Figure 4-40 shows the settling time, average frequency deviation, average tie-line deviation, and integrated ACE values for various values of the actual disturbance, when the predicted disturbance is 0.01 pu. A magnitude multiplier of 1 means the actual disturbance is exactly the predicted

disturbance, while a multiplier of 0 means there was no actual disturbance and a multiplier of 2 means that the actual disturbance was twice the magnitude of the predicted disturbance.



(a)



(b)

Figure 4-39 System Frequencies (a) Without and (b) With Inter-area Communication, Balanced

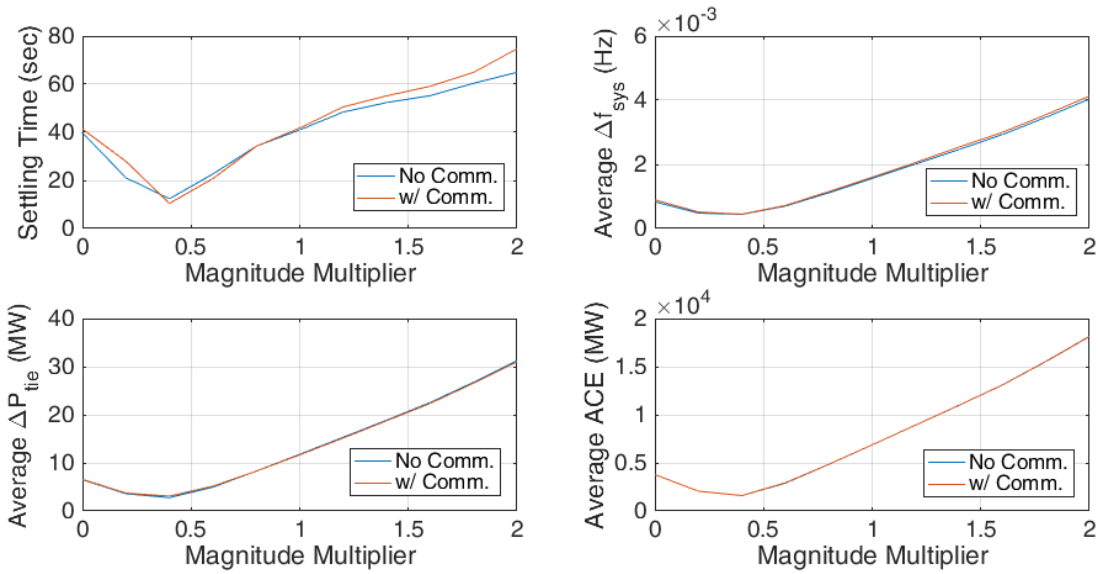
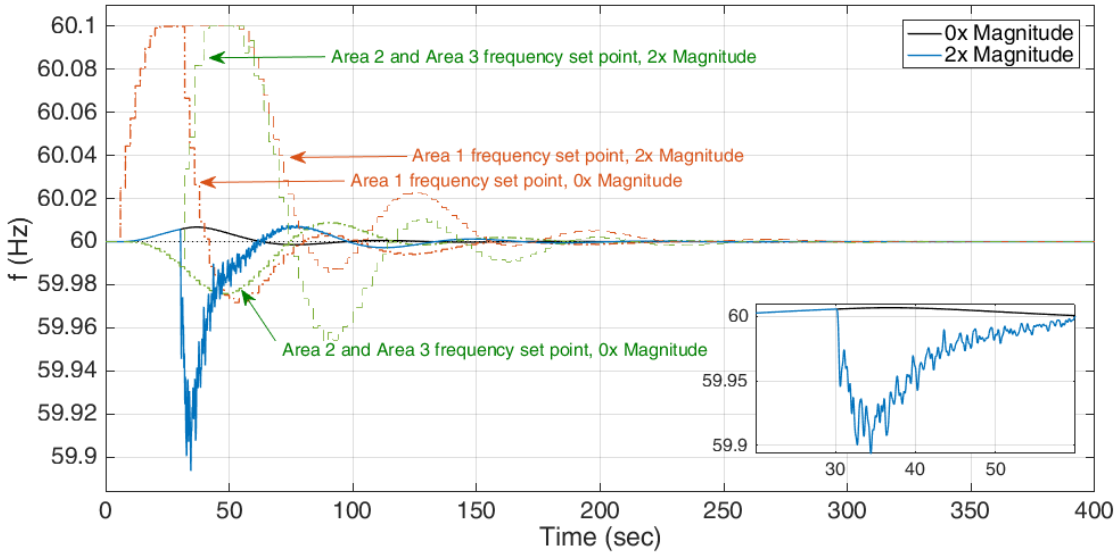


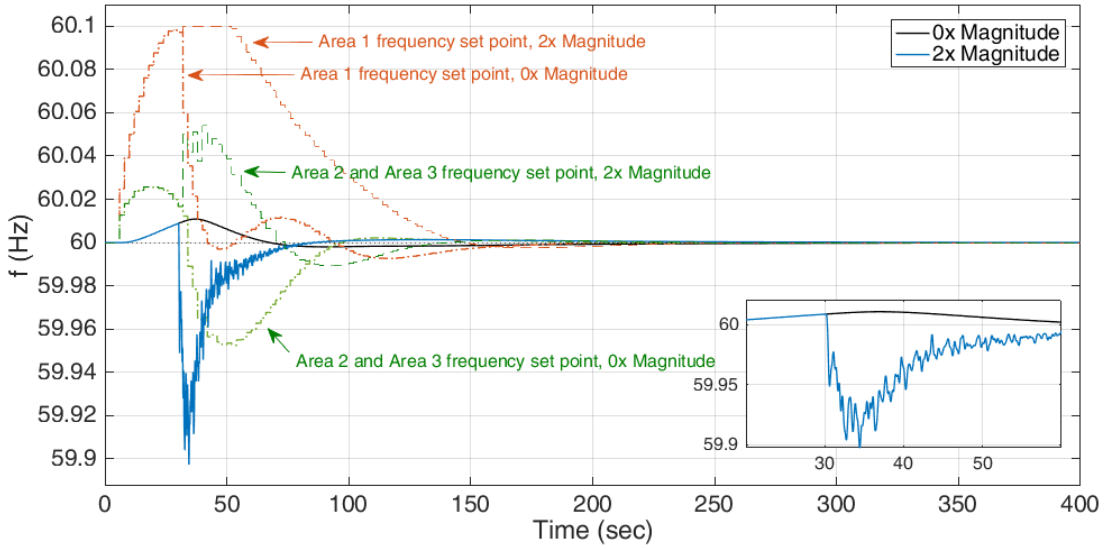
Figure 4-40 Operating Metrics with Imperfect Forecasting, under Equal Weighting

As expected, the operational performance for situations with and without communication are very similar because of the balanced weighting of the Δf_{sys} and ΔP_{tie} values. As seen in Figure 4-39, the frequency set points of Area 2 and Area 3 are always close to zero, as areas outside of the disturbed area only contribute in primary response, avoiding heavy contribution in secondary response. This results in similar metrics between the systems with and without inter-area communication.

Simulations with large $Q_f = 1.5$ are now presented, with the same sweep of actual disturbances ranging from 0 pu to 0.02 pu and a predicted disturbance of 0.01 pu. Figure 4-41 shows the frequencies and frequency set points in the extreme scenarios of no disturbance and a 0.02 pu disturbance, and Figure 4-42 shows performance metrics swept over a range of disturbances.



(a)



(b)

Figure 4-41 Frequencies and Set Points for Imperfect Forecasting with Heavy Weight on Frequency Deviations $Q_f = 1.5$

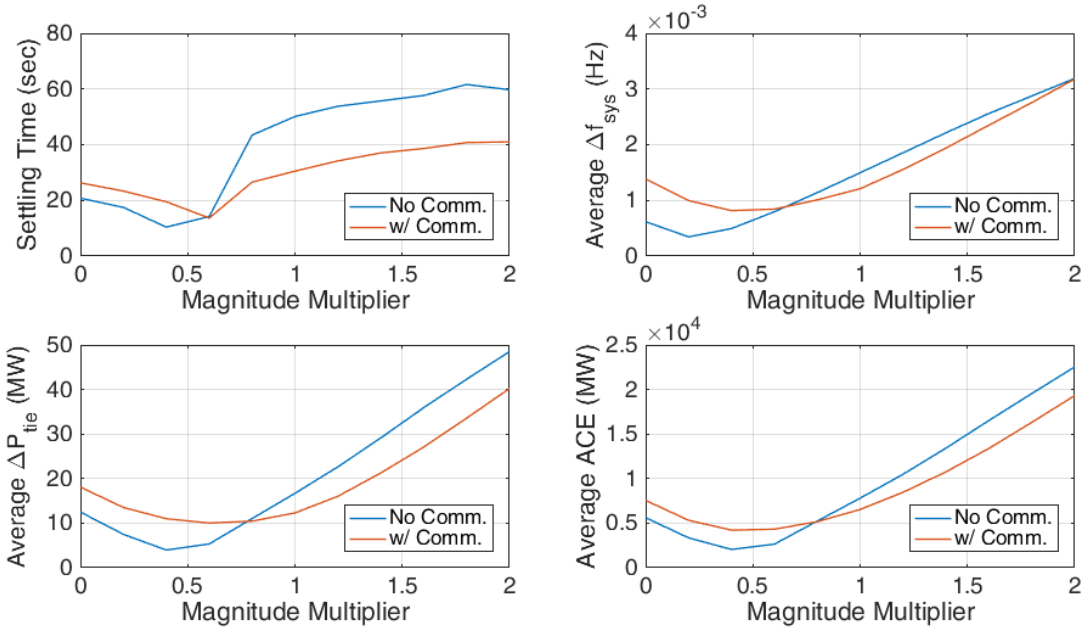


Figure 4-42 Operating Metrics with Imperfect Forecasting, with $Q_f = 1.5$

With imperfect forecasts, communication of the anticipated disturbance actually hurts operational performance, as areas outside of the Area 1 (which predicted the disturbance) react prior to the predicted disturbance. In situations with communication where the actual disturbance magnitude was less than ~75% of the predicted disturbance magnitude, neighboring areas over-prepared for the disturbance by excessively raising frequency. That is, in these simulations, if the true disturbance is less than approximately 75% of the anticipated forecast, performance suffers if areas are communicating. Thus, if areas set non-balanced Q_f and Q_{ptie} weights, it becomes important to only broadcast a forecasted disturbance with some degree of accuracy.

The multi-area simulations in this subsection offered insight on the performance of distributed MPC for anticipatory secondary control in situations with a perfect forecast and an imperfect forecast. It was seen that the weights of the frequency and tie-line deviations heavily affect the results, but that in general, inter-are communication is beneficial if the

forecasted disturbance is relatively accurate. Results so far have been limited to the cases of a discrete step disturbance, which can reflect the sudden disruption of generation from a transmission or generation outage. However, short-term forecasts for renewable energy are generally given in discrete time intervals, such as at a 5-minute basis. The next section will present results on MPC actions given a 5-minute-ahead forecasted disturbance.

4.4.4. 5-Minute Ahead Forecasts with Intra-Forecast Uncertainty

Anticipatory secondary control can utilize a prediction less than a minute in advance of rare, sudden events, as is shown in the previous subsection. This sub-minute prediction includes generation trip events, given through a signal given to the operator before the trip, as well as renewable energy events, which can be predicted through a weather station a kilometer or less upstream from a solar or wind farm.

In contrast to rare and sudden discrete events, short-term renewable energy forecasts are given consistently to wind farm and power system operators in set intervals (e.g. on a 5-minute basis). These point forecasts can also be used with anticipatory control, and simulations incorporating these forecasts into anticipatory secondary control are presented. Although short-term forecasts (e.g. 30-second forecasts) may be available from the use of weather towers that surround a wind farm, this study is limited to use of a 5-minute window because of the present availability of 5-minute-ahead wind power forecasting data.

Given an initial operating point of renewable energy generation P_{init}, the point forecast will provide the expected value P_{final} the generation will be at after some time duration T_d. However, $T_d = 5$ minutes, and AGC operates on a smaller timescale, so the

shape of the generation between these two points P_{init} and P_{final} is must be assumed. A predicted disturbance $P(t)$ over time t is assumed to be:

$$P(t) = P_{\Delta}\left(\frac{t}{T_d}\right)^{\gamma} + P_{init} \quad (4\text{-}45)$$

where P_{init} is the initial value of the generation at the beginning of the disturbance, P_{Δ} is the delta change of the disturbance value $P_{final} - P_{init}$, T_d is the duration of the disturbance, and γ is a controllable parameter determining the shape of the disturbance, controlling the shape of the change, as shown in Figure 4-43. Note that $\gamma \in [0, \text{inf})$.

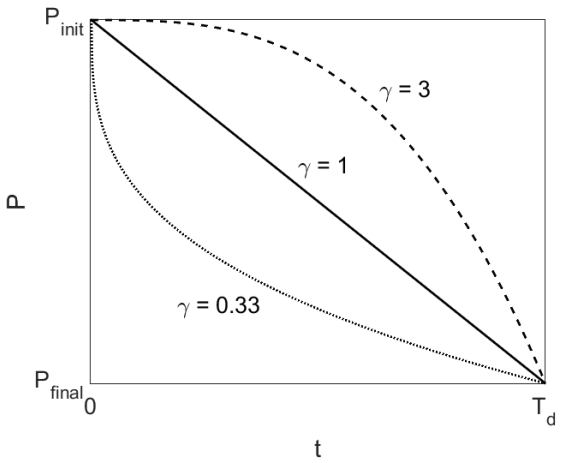


Figure 4-43 Shape of the Disturbance Curve Between the Initial Value and the Final Value

It is important to note that the controller only has knowledge of the next point forecast, and it is assumed in this work that one-step-ahead forecasts are given every 5 minutes. Simulations in this section will use the profile of the forecast shown in Figure 4-44, where the unknown window is where the shape determined by γ in Figure 4-43 will occur. This mirrors the generation points before and after a disturbance seen in prior simulations in this work, where the generation at bus 2 in Area 1 drops from 2500 MW to 2250 MW.

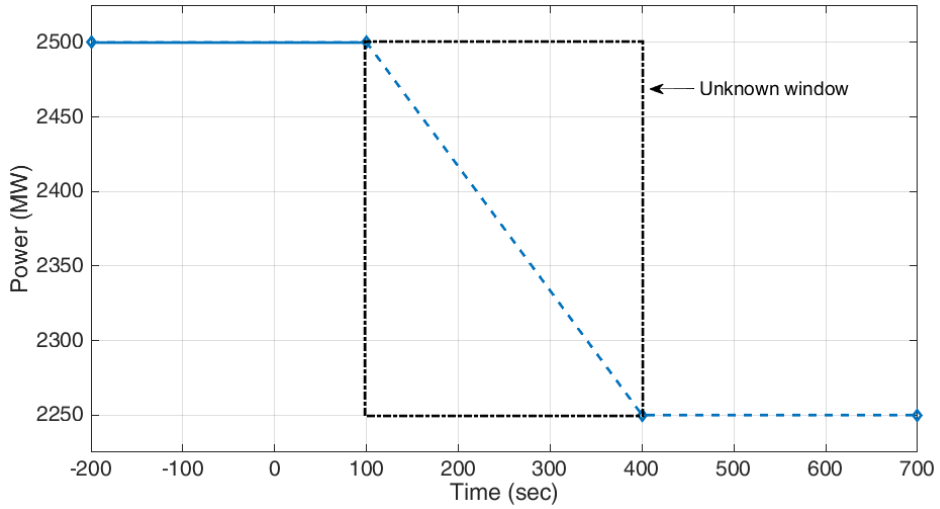


Figure 4-44 Assumed Generation Shape with a Window that Contains a Predicted Path

With the forecast in Figure 4-44, when the controller operates in the interval of t $\in$ [0 100) seconds the controller will anticipate the point forecast for 2500 MW at t = 100 seconds, which is no change. When operating in the interval of t $\in$ [100 400) seconds, the controller will know that there is a forecast of 2250 MW at t = 400 seconds. The controller is given some type of interpolation defined by the γ value to fill in the forecasted power output gaps between the 5-minute-ahead forecasts. It is assumed that the power output is steady at 2500 MW before t = 100 seconds, and that the output is steady at 2250 MW after t = 400 seconds.

To evaluate the effectiveness of anticipatory secondary control using 5-minute forecasts, the shape γ of the actual disturbance was assumed to be one of the three values shown in Figure 4-43, which are 0.33, 1, and 3. Multiple 600 second simulations were ran to compare anticipatory secondary control with conventional AGC and external PI control. Because of the longer forecast (5-minutes), it was beneficial to increase the prediction and control horizons of the MPC controller, so the forecast horizon for the controller was set

to 120 seconds, while the control horizon was set to 60 seconds. Figure 4-45 and Figure 4-46 show the frequency and net tie-line flow for Area 1 for various γ values using conventional AGC. Next, a set of anticipated values $\hat{\gamma}$ were tested with an actual linear ramp (where the actual $\gamma = 1$), with frequency and tie-line results shown in Figure 4-47 and Figure 4-48, and responsive turbine power and ACE outputs shown in Appendix C VI. Table 4-14 shows the integrated ACE values (integrated over the simulation time of 600 seconds) for anticipatory control assuming it anticipated $\hat{\gamma} = 0.33$, $\hat{\gamma} = 1$, or $\hat{\gamma} = 3$ for all three scenarios of the actual $\gamma = 0.33$, $\gamma = 1$, or $\gamma = 3$, and compares them with external PI and conventional AGC.

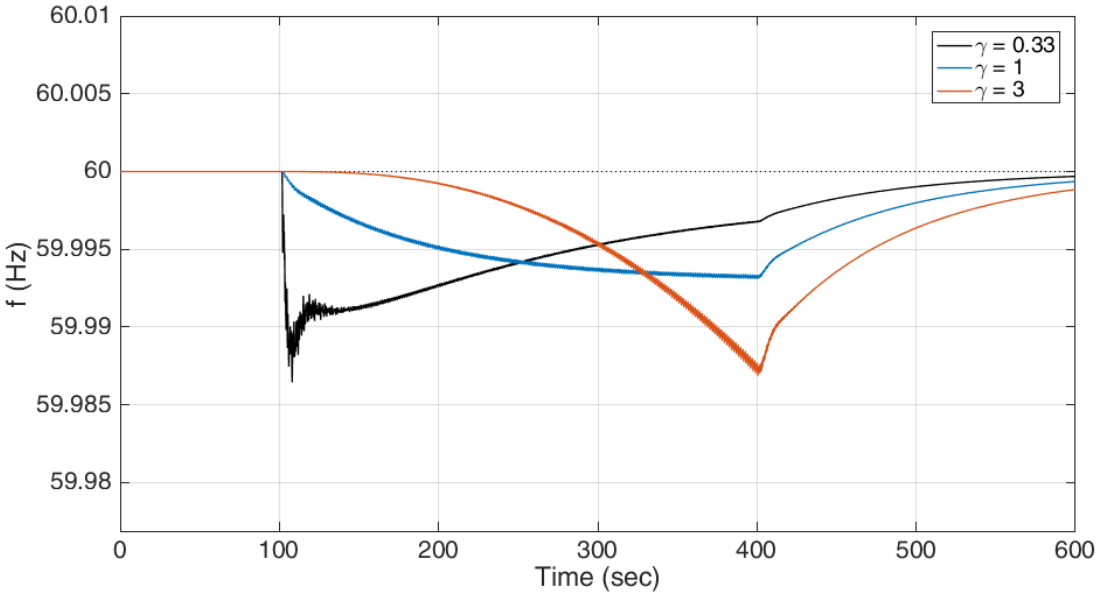


Figure 4-45 Frequency of Multi-area System under Conventional AGC for Varied γ

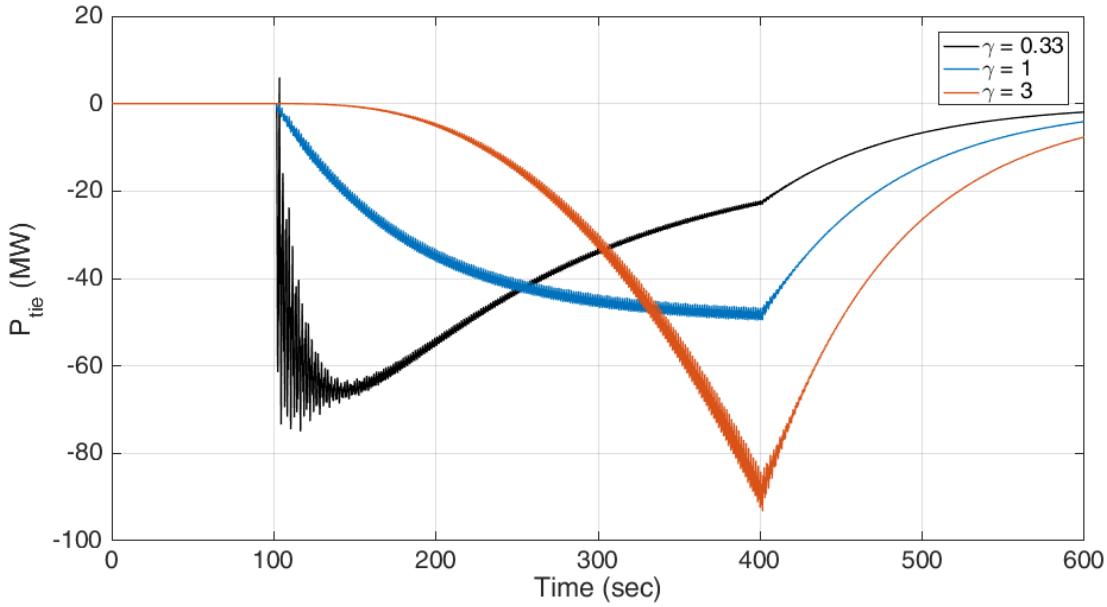


Figure 4-46 Net tie-line Flow of Area 1 under Conventional AGC for Varied γ

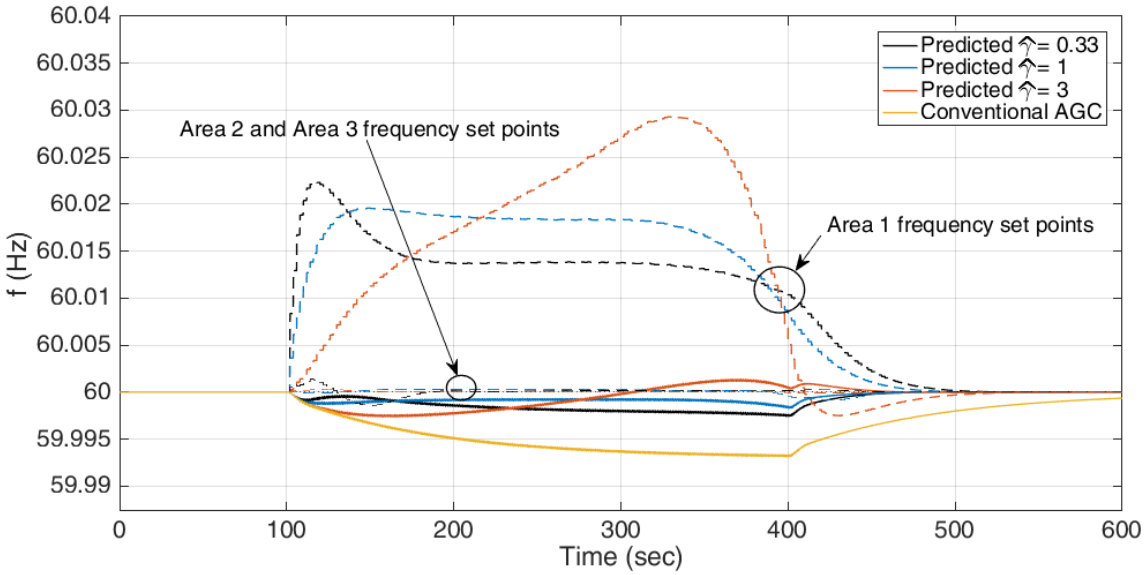


Figure 4-47 Frequency and Set Points for MPC with Varied Predicted $\hat{\gamma}$, given $\gamma = 1$

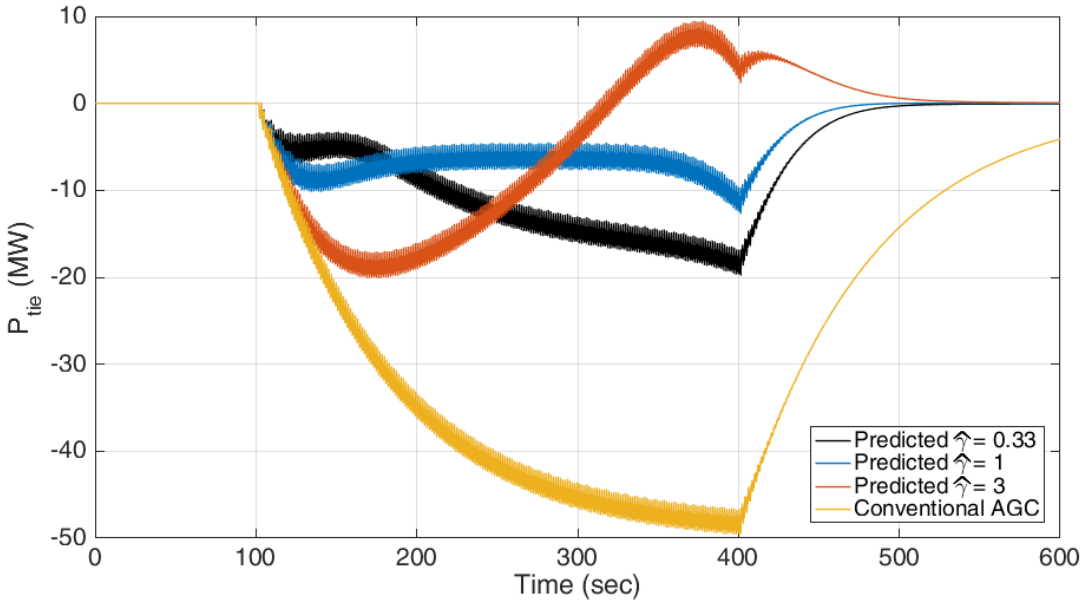


Figure 4-48 Net Tie-line Flows for MPC with Varied Predicted $\hat{\gamma}$, Given $\gamma = 1$

Table 4-14 Integrated ACE for Various Predicted vs. Actual Wind Ramps

		Actual γ		
		0.33	1	3
Anticipated $\hat{\gamma}$ for MPC	0.33	5791	5749	9103
	1	6222	3382	7165
	3	10409	4974	2811
Conventional		21973	21697	21256
External PI		18650	18657	18559

Note that given an actual γ, MPC produces a lower integrated ACE than other methods, even if the anticipated $\hat{\gamma}$ was the inverse of the actual γ. That is, in every column of the table, the integrated ACE for MPC under any of the anticipated $\hat{\gamma}$ values is the lowest of all methods. In practice, keeping $\hat{\gamma} = 1$ for MPC with the lack of any knowledge on the inter-forecast shape of the disturbance would work well, based on these results.

As mentioned earlier, very strong assumptions about the shape of the wind power curve between the point forecasts were used, as readily available wind power output data is at a 5-minute or greater timescale, though secondary control operates on a minute and

sub-minute timescale. In addition to showing the performance of anticipatory secondary control under such assumptions, this simulation highlights the need for available data on a 30-second-ahead or 1-minute-ahead basis to 1) test the performance of anticipatory secondary control under realistic conditions without assumptions on the behavior of renewable generation between 5-minute point forecasts, and 2) take advantage of the strengths of anticipatory control with a minute or less advance notice of a disturbance, as shown in prior simulations in this chapter.

4.5. Conclusions on Anticipatory Secondary Control

Anticipatory secondary control was compared with various forms of secondary control in both single-area and multi-area simulations in this work, using a power system model that was validated through simulation. Single area simulations showed the basic functions of MPC in a system, while multi-area simulations highlighted the importance of the weights in the objective function of its MPC controller, balancing deviations in frequency with deviations in net tie-line flow. In both the single-area and multi-area scenarios, errors in the forecast were introduced, which showed that MPC still was able to outperform traditional methods of secondary control with large forecast errors. Varying objective function weights in the MPC were shown to change the behavior of distributed MPC within a multi-area system, showing that a system with unbalanced objective function benefits heavily from inter-area communication. Integration of this controller would not require completely replacing the existing infrastructure, as it was designed as an add-on module that only manipulated the effective area frequency reference. In addition, the simulations in this section show advantages to using anticipatory control with a minute or

less notice of a disturbance, highlighting the benefits for available 30-second ahead or minute-ahead forecasts. Further work with this controller would involve the use of a more detailed power system model, in addition to examination of its performance with measured disturbances, as expanded upon in the next chapter.

5. CONCLUSIONS & FUTURE RESEARCH

5.1. Conclusions

This dissertation has presented work focused on characterization and improvements in short-term wind power forecasting along with the introduction of a new, anticipatory method of secondary control that can enhance power system operational performance. With increasing penetrations of wind energy, limiting the risk of integrating large amounts of stochastic generation is important from both an economic and reliability standpoint, and characterizing the extreme forecasting errors in expected generation is helpful in planning the future of the power system, covered in Chapter 2. Short-term forecasts are an integral part of power system operations with high amounts of wind energy, and so a novel metric that characterizes the amount of non-stationarity along with a proposed algorithm to find optimal training windows for statistical forecasting methods was presented in Chapter 3. An anticipatory secondary control scheme that makes use of prior knowledge of a major disturbance was proposed and tested on simulated power systems in Chapter 4.

Because of the stochastic nature of renewable resources, thorough characterization of wind activity is necessary to maintain grid stability and reliability. Results from extreme wind power ramp characterization indicated that standard Gaussian assumptions about wind power ramps would not capture low probability, high impact events that would raise

problems with power system economics or reliability. The proper characterization of extreme wind ramp events using EVA allows for proper risk assessment in a wind farm, indicating that a power system is exposed to more risk than previously modeled from integration of a wind farm. In the event that persistence forecasts are used, for example, these extreme wind ramps represent actual forecasting error, which would be previously underrepresented. Quantifying these risks will spur more research interest in stochastic generation, as this model poses new questions about reliability and pricing concerns surrounding wind power generation. In addition, the work can be extended to model extreme, rare occurrences in the power system, providing a mechanism to quantify risk for low-probability events.

Statistical forecasts, which are essential to power system operations in power systems with large amounts of wind energy, use a set of training data that is commonly determined through a purely heuristic approach. Also, wind power data is may be highly non-stationary, which may result in different optimal sets of training data over time. This work introduced the EDNS metric and showed that the proposed metric can quantify the degree of non-stationarity in a signal, which was previously not explicitly defined in time series wind power data analysis. An algorithm was also presented that used the EDNS to find the optimal window of training data, which was designed as a module for use in any statistical time series forecasting method. Use of the dynamic window algorithm with a statistical forecasting approach lowered the mean absolute error versus a carefully selected, static training window, given some highly non-stationary data. The introduction of the EDNS and an algorithm to determine optimal training windows in real time helps to

quantify the effects of the training window on short-term forecasts and provides a more rigorous method to determine a suitable training window.

Large disturbances such as generation trip events or large ramps in renewable generation can aversely effect power systems, resulting in large ACE values over time. The creation and characterization of an anticipatory secondary control scheme showed that, for both single-area and multi-area power systems, frequency deviation metrics can be improved with anticipation of an event over the use of conventional AGC or other proposed methods of secondary control. In a multi-area system, it was seen that the performance of the distributed MPC controllers are highly dependent on the weights in the MPC objective function as well as the existence of inter-area communication. Inter-area communication was seen to be almost essential in scenarios where an anticipatory secondary controller would act to remedy a disturbance outside of its own area. As the anticipatory controller is designed to be added onto an existing AGC system, this work provides insight into an effective method of utilizing prior knowledge of an event to minimize frequency deviations using existing secondary control infrastructure.

This dissertation provided insight and methods for improving the stability of the power grid with a large penetration of wind energy. Natural directions for future research will be presented in the next section.

5.2. Future Research

Future research for the work presented in the dissertation is divided into three sections, each following one of the chapters for extreme wind ramp characterization, the quantification of non-stationarity, as well as anticipatory secondary control.

Although a method for extreme wind ramp characterization has been shown, the impacts of this on operations or planning in a power system have not been investigated. Future research can include the construction of a new wind power output model based upon an EVA-characterization of historical wind power output data, and the differences it would ultimately make in forecasting studies or planning studies in some geographic region. This approach could be compared to the existing wind speed models as shown by NREL, which provide high-resolution wind speed and wind power output models across much of the United States [131].

The quantification of non-stationarity with the EDNS uses the spectrum, arrived at through an empirical approach called empirical mode decomposition (EMD). However, recent work from N. Huang addressed some limitations of the EMD method with a proposed method, Holo-Hilbert spectral analysis [132]. Use of the new method to arrive at a more accurate spectrum of frequencies over time can provide a more accurate way of characterizing the non-stationarity of a signal, which would ultimately lead to more accurate optimal training windows in short-term wind power forecasting. Also, the algorithm that uses the EDNS may be improved, as for example, it only takes into account the most recent data. Empirical studies have shown that the use of seasonal data, cut into epochs, may show improved forecasting results [31]. For example, instead of using the most recent 48 hours, forecasting may be improved with the use of data from midnight to noon on the two most recent days, because of the cyclic behavior of wind.

Future research in anticipatory secondary control may involve the use of more detailed power system models. Voltage was assumed to be at 1 pu throughout the system in this work, and a simplified model of a power system was used. Full models may be used

to study the numerical efficiency of the method, testing the viability of using anticipatory secondary control with the tradeoffs of controller performance and power system model complexity. Also, further work can include the application of the anticipatory secondary controller as a constantly working distributed MPC controller, versus one that is only used when a large disturbance is predicted. As an extension of the 5-minute ahead forecast simulations in this work, an MPC controller that is constantly fed load and renewable energy forecasts must be examined for its robustness to error in both forecasts and its tangible improvements in reducing the cost to operate the power system. For example, comparisons over a year of operation for an actual interconnection between conventional AGC and anticipatory secondary control would provide important insights into the costs and benefits of integrating anticipatory secondary control.

Although the methods proposed in this dissertation are not immediately ready for implementation in the actual power system, further research should effectively quantify their value in improving power system operations. These methods support future innovations in wind power modeling, wind power forecasting, and power system controls in scenarios with a large amount of installed wind capacity, and these innovations would improve power system operational performance in the future power grid.

REFERENCES

[1] U.S. Department of Energy, "Annual Energy Outlook 2013 Early Release Overview," Jan. 2013.

[2] M. Hand, N. Blair, M. Bolinger, R. Wiser, R. O'Connell, T. Hern and B. Miller, "Power System Modeling of 20% Wind-Generated Electricity by 2030," in *IEEE Power Engineering Society General Meeting*, July 2008.

[3] U.S. Department of Energy, "20% Wind Energy by 2030," Washington D.C., 2008.

[4] S. Joo, J. Kim and C. Liu, "Empirical Analysis of the Impact of 2003 Blackout on Security Values of U.S. Utilities and Electrical Equipment Manufacturing Firms," *IEEE Trans. Power Sys.*, vol. 22, no. 3, pp. 1012-1080, Aug. 2007.

[5] National Transportation Safety Board, "Pipeline Accident Report," Oct. 2002.

[6] B.-M. Hodge, A. Florita, J. Sharp, M. Margulis and D. Mcreavy, "The Value of Improved Short-Term Wind Power Forecasting," NREL, Feb. 2015.

[7] K. Orwig et al., "Enhanced Short-Term Wind Power Forecasting and Value to Grid Operations," in *11th Annual International Workshop on Large-Scale Integration of Wind Power*, NREL CP-5500-56258, Nov. 2012.

[8] K. Parks, Y. Wan, G. Wiener and L. Y., "Wind Energy Forecasting: A Collaboration of the National Center for Atmospheric Research (NCAR) and Xcel Energy," NREL SR-5500-52233, Oct. 2011.

[9] L. Xie, P. M. S. Carvalho, L. A. F. M. Ferreira, J. Liu, B. Krogh, N. Popli and M. Ilic, "Wind Integration in Power Systems: Operational Challenges and Possible Solutions," *Proc. of IEEE*, vol. 99, no. 1, pp. 214-232, Oct. 2010.

[10] D. Lew, M. Milligan, G. Jordan and R. Piwko, "The Value of Wind Power Forecasting," NREL, Jan. 2011.

[11] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee and M. Anitescu, "Unit Commitment with Wind Power Generation: Integrating Wind Forecast Uncertainty and Stochastic Programming," Argonne National Laboratory, Mathematics and Computer Science Division, Sep. 2009.

[12] R. Sevlian and R. Rajagopal, "Wind power ramps: Detection and statistics," in *IEEE PES General Meeting*, 2012.

[13] C. Ferreira, J. Gama, L. Matias, A. Botterud and J. Wang, "A Survey on Wind Power Ramp Forecasting," Dec. 2010.

[14] B. Greaves, J. Collins, J. Parkes and A. Tindal, "Temporal Forecast Uncertainty for Ramp Events," in *European Wind Energy Conference & Exhibition*, Marseille, Frances, 2009.

[15] AWS Truewind, "AWS Truewind's Final Report for the Alberta Forecasting Pilot Project," Jun. 2008.

[16] GE Power & Water Renewable Energy, "2.5 MW Wind Turbine Series," 2010.

[17] Y. Gao and R. Billinton, "Adequacy assessment of generating systems containing wind power considering wind speed correlation," *IET Renewable Power Generation,* vol. 3, no. 2, pp. 217-226, 2009.

[18] D. A. Bechrakis and P. D. Sparis, "Correlation of wind speed between neighboring measuring stations," *IEEE Trans. on Energy Conversion,* vol. 19, no. 2, pp. 400-406, 2004.

[19] F. Vallee, J. Lobry and a. O. Deblecker, "Impact of the Wind Geographical Correlation Level for Reliability Studies," *IEEE Trans. Power Sys.,* vol. 22, no. 4, pp. 2232-2239, Nov. 2007.

[20] J. Lerner, M. Grundmeyer and a. M. Garvert, "The importance of wind forecasting," *Renewable Energy Focus,* vol. 10, no. 2, pp. 64-66, Mar. 2009.

[21] H. Bludszuweit, J. A. Dominguez-Navarro and a. A. Llombart, "Statistical Analysis of Wind Power Forecast Error," *IEEE Trans. Power Sys.,* vol. 23, no. 3, pp. 983-991, Aug. 2008.

[22] N. E. Evmorfopoulos, G. I. Stamoulis and J. N. Avaritsiotis, "A Monte Carlo approach for maximum power estimation based on extreme value theory," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems,* vol. 21, no. 4, pp. 415-432, 2002.

[23] B. R. L. Cour, "Statistical characterization of active sonar reverberation using extreme value theory," *IEEE Journal of Oceanic Engineering,* vol. 29, no. 2, pp. 310-316, 2004.

[24] G. Sideratos and N. D. Hatziargyriou, "Wind Power Forecasting Focused on Extreme Power System Events," *IEEE Trans. on Sustainable Energy.*, vol. 3, no. 3, pp. 445-454, Jul. 2012.

[25] H. You, V. Vittal and Z. Yang, "Self-Healing in Power Systems: An Approach Using Islanding and Rate of Frequency Decline-Based Load Shedding," *IEEE Trans. Power Sys.*, vol. 18, no. 1, pp. 174-181, Feb. 2003.

[26] X. Zhang and M. Lai, "Risk analysis of power bidding market based on extreme value theory," in *Int. Conf. on Wireless Communications, Networking and Mobile Computing*, 2008.

[27] H. Waldl and P. Brandt, "Anemos.Rulez: Rule based extreme event prediction and alarming to support the integration of wind power," in *Int. Conf. on Intelligent System Application to Power Systems*, 2011.

[28] Y. Wu and J. Hong, "A literature review of wind forecasting technology in the world," in *IEEE Lausanne Power Tech*, 2007.

[29] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis and P. S. Dokopoulos, "A Fuzzy Model for Wind Speed Prediction and Power Generation in Wind Parks Using Spatial Correlation," *IEEE Trans. Energy Conversion*, vol. 19, no. 2, pp. 352-361, Jun. 2004.

[30] K. Bhaskar and S. N. Singh, "AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network," *IEEE Trans. on Sustainable Energy*, vol. 3, no. 2, pp. 306-315, Apr. 2012.

[31] M. He, L. Yang, J. Zhang and V. Vittal, "A Spatio-Temporal Analysis Approach for Short-Term Forecast of Wind Farm Generation," *IEEE Trans. Power Sys.*, vol. 29, no. 4, pp. 1611-1622, Jul. 2014.

[32] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, Time Series Analysis: Forecasting and Control, Englewood Cliffs, NJ: Prentice Hall, 1994.

[33] G. T. Heydt, R. Ayyanar, K. W. Hedman and V. Vittal, "Electric Power and Energy Engineering: The First Century," *Proc. of the IEEE, Special Centennial Issue*, vol. 100, pp. 1315-1328, 2012.

[34] N. Jaleeli, L. S. VanSlyck, D. N. Ewart, L. H. Fink and A. G. Hoffmann, "Understanding Automatic Generation Control," *IEEE Trans. Power Sys.*, vol. 7, no. 3, pp. 1106-1122, 1992.

[35] Ibraheem, P. Kumar and D. P. Kothari, "Recent Philosophies of Automatic Generation Control Strategies in Power Systems," *IEEE Trans. Power Sys.,* vol. 20, no. 1, pp. 346-357, Feb. 2005.

[36] J. L. Rodriguez-Amenedo, S. Arnalte and J. C. Burgos, "Automatic generation control of a wind farm with variable speed wind turbines," *IEEE Trans. Energy Conversion,* vol. 17, no. 2, pp. 279-284, 2002.

[37] H. Banakar, C. Luo and B. T. Ooi, "Impacts of Wind Power Minute-to-Minute Variations on Power System Operation," *IEEE Trans. Power Sys.,* vol. 23, no. 1, pp. 150-160, 2008.

[38] C. Luo et al., "Estimation of Wind Penetration as Limited by Frequency Deviation," *IEEE Trans. Energy Conversion,* vol. 22, no. 3, pp. 783-791, Sep. 2007.

[39] T. Gjengedal, "System Control of Large Scale Wind Power by use of Automatic Generation Control (AGC)," in *IEEE PES Int. Symposium of Quality and Security of Electric Power Delivery Systems,* 2003.

[40] M. D. Ilic, Y. Makarov and D. Hawkins, "Operations of Electric Power Systems with High Penetration of Wind Power: Risks and Possible Solutions," in *IEEE General PES Meeting,* 2007.

[41] F. F. Wu and V. S. Dea, "Describing-function Analysis of Automatic Generation Control System with Governor Deadband," *Elect. Power Syst. Res.,* vol. 1, no. 2, pp. 113-116, Apr. 1978.

[42] B. Oni, H. Graham and L. Walker, "Investigation of Nonlinear Tie-line Bias Control of Interconnected Power Systems," *IEEE Trans. Power App. Syst.,* Vols. PAS-100, no. 5, pp. 2350-2356, May 1981.

[43] S. C. Tripathy, T. S. Bhatti, C. S. Jha, O. P. Malik and G. S. Hope, "Sampled Data Automatic Generation Control Analysis with Reheat Steam Turbines and Governor Dead Band Effects," *IEEE Trans. Power App. Syst.,* Vols. PAS-103, no. 5, pp. 1045-1051, May 1984.

[44] Y. L. Abdel-Magid and M. M. Dawoud, "Optimal AGC Tuning with Genetic Algorithms," *Elect. Power Syst. Res.,* vol. 38, no. 3, pp. 231-238, Sep. 1996.

[45] Y. L. Karnavas and D. P. Papadopoulos, "AGC for Autonomous Power System Using Combined Intelligent Techniques," *Elect. Power Syst. Res.,* vol. 62, no. 3, pp. 225-239, Jul. 2002.

[46] D. Rerkpreedapong, A. Hassanovic and A. Feliachi, "Robust Load Frequency Control Using Genetic Algorithms and Linear Matrix Inequalities," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 855-861, May 2003.

[47] H. J. Kunish, K. G. Kramer and H. Dominik, "Battery Energy Storage – Another Option for Load-frequency Control and Instantaneous Reserve," *IEEE Trans. Energy Convers.*, Vols. EC-1, no. 3, pp. 46-51, Sep. 1986.

[48] S. C. Tripathy, R. Balasubramanian and P. S. C. Nair, "Adaptive Automatic Generation Control with Superconducting Magnetic Energy Storage in Power Systems," *IEEE Trans. Energy Convers.*, vol. 7, no. 3, pp. 434-441, Sep. 1992.

[49] A. Demiroren, "Automatic Generation Control with Fuzzy Logic Controllers in the Power System Including SMES Units," *Int. J. Elect. Power Energy Syst.*, vol. 26, no. 4, pp. 291-305, May 2004.

[50] C. E. Fosha and O. I. Elgerd, "The Megawatt Frequency Control Problem: A New Approach Via Optimal Control Theory," *IEEE Trans. Power App. Syst.*, Vols. PAS-89, no. 4, pp. 563-577, Apr. 1970.

[51] E. C. Tacker, C. C. Lee, T. W. Reddoch, T. O. Tan and P. M. Julich, "Optimal Control of Interconnected Electric Energy Systems: A New Formulation," *Proc. IEEE*, vol. 60, no. 10, pp. 1239-1241, Oct. 1972.

[52] C. He, H. Wang and X. Cao, "Optimal Dynamic Load Dispatching with the Coordination of Wind Farm and AGC Units," in *IEEE PES General Meeting*, Jul. 2014.

[53] H. Guolian, Z. Xiaobin, J. Pengcheng and Z. Jianhua, "Study of Modeling and Intelligent Control on AGC System with Wind Power," in *26th Chinese Control and Decision Conf.*, Jun. 2014.

[54] A. M. Ersdal, L. Imsland and K. Uhlen, "Model Predictive Load-Frequency Control," *IEEE Trans Power Sys*, vol. 31, no. 1, pp. 777-785, Jan 2016.

[55] A. N. Venkat, I. A. Hiskens, J. B. Rawlings and S. J. Wright, "Distributed MPC Strategies With Application to Power System Automatic Generation Control," *IEEE Trans. on Cont. Sys. Tech.*, vol. 16, no. 6, pp. 1192-1206, Nov 2008.

[56] H. Jiang, J. Lin, Y. Song and D. J. Hill, "MPC-Based Frequency Control With Demand-Side Participation: A Case Study in an Isolated Wind-Aluminum Power System," *IEEE Trans. Power Sys.*, vol. 30, no. 6, pp. 3327-3337, Nov 2015.

[57] S. M. Kaplan, "Electric Power Transmission: Background and Policy Issues," 2009.

[58] H. Y. Yamin, "Review on methods of generation scheduling in electric power systems," *Electric Power Systems Research*, vol. 69, pp. 227-248, May 2004.

[59] "Energy Policy Act of 2005," 2005.

[60] D. Mooney, "Integrating Large-Scale Renewable Energy into the Nation's Electricity System," Sep. 2011.

[61] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest and smallest member of a sample," *Proc. of the Cambridge Philosophical Society*, pp. 180-190, 1928.

[62] B. Gnedenko, "Sur la distribution limite du terme maximum d'une serie aleatoire," *Ann. of Math*, vol. 44, no. 3, pp. 423-453, 1943.

[63] E. J. Gumbel, Statistics of Extremes, New York: Columbia University Press, 1958.

[64] A. Balkema and L. d. Haan, "Residual life time at great age," *Annals of Probability*, vol. 2, pp. 792-804, 1974.

[65] J. Pickands, "Statistical inference using extreme order statistics," *Annals of Statistics*, vol. 3, pp. 119-131, 1975.

[66] P. Friederichs, "An Introduction to Extreme Value Theory," Jul. 2007.

[67] V. P. Fernández, "Extreme value theory: value at risk and returns dependence around the world," in *Centro de Economía Aplicada, Universidad de Chile*, No. 161, 2003.

[68] R. R. Holmes and K. Dinicola, "100-Year Flood - It's All About Chance," 2010.

[69] S. Coles, An Introduction to Statistical Modeling of Extreme Values, London: Springer-Verlag, 2001.

[70] E. Omey, F. Mallor and E. Nualart, "An introduction to statistical modelling [sic] of extreme values," *Hub Research Paper*, Nov. 2009.

[71] J. Danielsson and C. G. d. Vries, "Where do Extremes Matter?," *RiskResearch*, Feb. 2002.

[72] J. Danielsson and C. G. d. Vries, "Beyond the Sample: Extreme Quantile and Probability Estimation," *IDEAS*, Dec. 1997.

[73] C. M. Goldie and R. L. Smith, "Slow variation with remainder: Theory and applications," *Quarterly Journal of Mathematics, Oxford,* vol. 2, no. 38, pp. 45-71, 1987.

[74] P. Lv and L. Yue, "Short-Term Wind Speed Forecasting Based on Non-stationary Time Series analysis and ARCH model," in *Int. Conf. Multimedia Technology*, 2011.

[75] S. Murugesan, J. Zhang and V. Vittal, "Finite State Markov Chain Model for Wind Generation Forecast: A Data-driven Spatiotemporal Approach," in *2012 IEEE PES Innovative Smart Grid Technologies*, 2012.

[76] [Online]. Available: http://people.duke.edu/~rnau/411diff.htm.

[77] O. Burke, "Statistical Methods. Autocorrelation. Non-Stationary Series," 2011.

[78] P. Chen, T. Pedersen, B. Bak-Jensen and Z. Chen, "ARIMA-Based Time Series Model of Stochastic Wind Power Generation," *IEEE Trans. Power Sys.,* vol. 25, no. 2, pp. 667-676, 2010.

[79] M. G. Genton and A. S. Hering, "Blowing in the Wind," *Significance,* vol. 4, no. 1, pp. 11-14, Mar. 2007.

[80] Australian Energy Market Operator, "Generation and Load," [Online]. Available: http://www.aemo.com.au/Electricity/Data/Market-Management-System-MMS/Generation-and-Load. [Accessed 2016].

[81] G. Schwarz, "Estimating the Dimension of the Model," *The Annals of Statistics,* vol. 6, no. 2, pp. 461-464, 1978.

[82] J. Dowell and P. Pinson, "Very-Short-Term Probabilistic Wind Power Forecasts by Sparse Vector Autoregression," *IEEE Transactions on Smart Grid,* vol. PP, no. 99, pp. 1-8, 2015.

[83] B. M. Schreiber, "Asymptotically Stationary and Related Processes," *Pure and Appl. Math.,* vol. 238, pp. 363-397, 2004.

[84] P. J. Brockwell and R. A. Davis, "Stationarity and Strict Stationarity," in *Time Series: Theory and Methods*, New York, NY, Springer, 2006, pp. 11-14.

[85] L. Yang, M. He, J. Zhang and V. Vittal, "Support-Vector-Machine-Enhanced Markov Model for Short-Term Wind Power Forecast," *IEEE Trans. Sust. Energy*, vol. 6, no. 3, pp. 791-799, Jul. 2015.

[86] B. Hobijn, "Generalization of the KPSS-test for Stationarity," Econometric Institute, Erasmus University Rotterdam, Rotterdam, Jan. 1998.

[87] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," *J. of Econometrics*, vol. 54, pp. 159-178, 1992.

[88] S. E. Said and D. A. Dickey, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, vol. 71, no. 3, pp. 599-607, 1984.

[89] P. C. B. Phillips and P. Perron, "Testing for a Unit Root in Time Series Regression," *Biometrika*, vol. 75, no. 2, pp. 335-346, 1988.

[90] Y. Cheung and K. S. Lai, "Practitioner's Corner: Lag Order and Critical Values of a Modified Dickey-Fuller Test," *Oxford Bulletin of Economics and Statistics*, vol. 57, no. 3, pp. 411-419, 1995.

[91] M. B. Priestley and S. T. Rao, "A Test for Non-Stationarity of Time-Series," *Journal of the Royal Statistical Society*, vol. 31, pp. 140-149, 1969.

[92] G. P. Nason, "A Test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series," *J. of the Royal Stat. Society*, vol. 75, pp. 1-36, 2013.

[93] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A*, vol. 454, pp. 903-995, 1998.

[94] W. Chang, "A Literature Review of Wind Forecasting Methods," *J. of Power and Energy Eng.*, vol. 2, pp. 161-168, Apr. 2014.

[95] G. Eshel, "The Yule Walker Equations for the AR Coefficients," Wharton Statistics Department, 2016.

[96] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974.

[97] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing,* vol. 14, pp. 199-222, 2004.

[98] B. Scholkopf and A. J. Smola, "Regression Estimation," in *Learning with Kernels*, Cambridge, Massachusetts, The MIT Press, 2002, pp. 251-279.

[99] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* vol. 2, pp. 121-167, 1998.

[100] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning,* vol. 20, pp. 273-297, 1995.

[101] T. Hofmann, B. Scholkopf and A. J. Smola, "A Review of Kernel Methods in Machine Learning," Max Planck Institute for Biological Cybernetics, Germany, Dec. 2006.

[102] C. Hsu, C. Chang and C. Lin, "A Practical Guide to Support Vector Classification," National Taiwan University, Taipei, 2003.

[103] V. Cherkassky, "Support Vector Machines," in *Predictive Learning*, 2013.

[104] C. Chang and C. Lin, "LIBSVM -- A Library for Support Vector Machines," 14 Dec. 2015. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/libsvm/. [Accessed 2016].

[105] O. Chapelle, "Training a support vector machine in the primal," *Neural Comput.*, vol. 19, no. 5, pp. 1115-1178, May 2007.

[106] D. G. Long, "Comments on Hilbert Transform Based Signal Analysis," Brigham Young University, MERS 04-001, 2004.

[107] R. C. Sharpley and V. Vatchev, "Analysis of the intrinsic mode functions," Industrial Mathematics Institute, Dec. 2004.

[108] Y. Wang, C. Yeh, H. V. Young, K. Hu and M. Lo, "On the computationa complexity of the empirical mode decomposition algorithm," *Physica A*, vol. 400, pp. 159-167, 2014.

[109] PJM Interconnection, "Security Constrained Economic Dispatch System (SCED)," 2011.

[110] L. S. VanSlyck, N. Jaleeli and W. R. Kelley, "A Comprehensive Shakedown of an Automatic Generation Control Process," *IEEE Trans. Power Sys.*, vol. 4, no. 2, pp. 771-781, May 1989.

[111] "Balancing and Frequency Control: A Technical Document," Jan. 2011.

[112] N. Cohn, "Control of Interconnected Power Systems," in *Handbook of Automation, Computation, and Control, 2nd ed, vol. 3*, New York, Wiley, 1971.

[113] L. Kirchmayer, "Control Analysis of Power Systems," in *Economic Control of Interconnected Systems, 1st ed.*, New York, Wiley, 1959, pp. 1-47.

[114] C. Guan, P. B. Luh, L. D. Michel, Y. Wang and P. B. Friedland, "Very Short-Term Load Forecasting: Wavelet Neural Networks with Data Pre-Filtering," *IEEE Trans. Power Sys.*, vol. 28, no. 1, pp. 30-41, Feb. 2013.

[115] D. Lew and G. Brinkman, "The Western Wind and Solar Integration Study Phase 2: Executive Summary," Sep. 2013.

[116] J. Undrill, "Power and Frequency Control as it Relates to Wind-Powered Generation," Dec. 2010.

[117] A. J. Wood and B. F. Wollenberg, "Control of Generation," in *Power Generation Operation and Control, 2nd ed.*, New York, John Wiley & Sons, Inc., 1996, pp. 345-355.

[118] NERC Training Resources Working Group, "NERC Training Document Understand and Calculate Frequency Response," Feb. 2003.

[119] P. Etingov, "Frequency Response Analysis Tool," Pacific Northwest National Laboratory, Washington, DC, 2015.

[120] NERC Resources Subcommittee, "Balancing and Frequency Control: A Technical Document," North American Electric Reliability Corporation, Princeton, NJ, 2011.

[121] North American Electric Reliability Corporation, "Real Power Balancing Control Performance," *Standard BAL-001-0.1a*, 2009.

[122] M. Morari, J. H. Lee and C. E. Garcia, "Dynamic Matrix Control – The Basic Algorithm," in *Model Predictive Control*, Mar. 2002.

[123] A. Jadbabaie and J. Hauser, "On the Stability of Receding Horizon Control With a General Terminal Cost," *IEEE Trans. Auto. Cont.*, vol. 50, no. 5, pp. 674-678, May 2005.

[124] F. Borrelli, A. Bemporad and M. Morari, "Receding Horizon Control," in *Predictive Control for Linear and Hybrid Systems*, 2014, pp. 260-273.

[125] MATLAB, "Controller State Estimation," MATLAB, 2016. [Online]. Available: http://www.mathworks.com/help/mpc/ug/controller-state-estimation.html.

[126] R. R. Labbe, Kalman and Bayesian Filters in Python, Dec 2015.

[127] C. Schmid and L. T. Biegler, "Quadratic programming methods for reduced Hessian SQP," *Computers & Chemical Engineering*, vol. 18, no. 9, pp. 817-832, 1994.

[128] MATLAB, "Documentation for mpcqpsolver," 2016. [Online]. Available: http://www.mathworks.com/help/mpc/ref/mpcqpsolver.html.

[129] E. Grid Information. [Online]. Available: http://www.ercot.com/gridinfo/csc/index.html.

[130] J. Undrill, "Steam Plants," in *Power System Dynamics and Control*, 2014, pp. 47-52.

[131] NREL, "Wind Integration National Dataset (WIND) Toolkit," 2016. [Online]. Available: http://www.nrel.gov/electricity/transmission/wind_toolkit.html.

[132] N. Huang et al., "On Holo-Hilbert spectral analysis: a full informational spectral representation for nonlinear and non-stationary data," *Phil. Trans. R. Soc. A*, vol. 374, no. 2065, Apr. 2016.

[133] G. Rilling, P. Flandrin and P. Goncalves, "On Empirical Mode Decomposition and Its Algorithms," IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, Grado, Jun. 2003.

[134] M. Klein and P. K. G. J. Rogers, "A Fundamental Study of Inter-Area Oscillations," *IEEE Trans. Power Sys.*, vol. 6, no. 3, pp. 914-921, Aug. 1991.

[135] M. He, V. Vittal and J. Zhang, "A Sparsified Vector Autoregression Model for Short-term Wind Farm Power Forecasting," in *IEEE PES Gen. Meeting*, 2015.

[136] B. J. Kirby et al., "Frequency Control Concerns in The North American Electric Power System," Dec. 2002.

APPENDIX A

NOTES ON THE EDNS METRIC FOR NON-STATIONARITY

I. Training Windows with SVR

A range of training window sizes were swept over the same wind farm power output data outlined in Section 3.1.1 using an SVR forecasting approach. The features of the SVR model were defined to be the $p = 4$ past power output measurements, which corresponds to the results of the BIC method, and the label was defined to be the difference between the past and present power output. As a result, the matrix of training data was of dimension $(T^* - 3)$ x 4, while the label vector was of dimension $(T^* - 3)$ x 1 for the creation of an SVR model. A new SVR model was created at every 5-minute time step based on the most recent data of length T^*, which was contrasted with an SVR created at every 5-minute time step based on a static training window T_{st}^*, the optimal static training window. MATLAB 9.0 was used for these simulations in conjunction with the *LIBSVM* library [104]. In this work, $C = \max y - \min y = 1$ after data normalization, and a grid search is performed over $\varepsilon = 0{,}1, \ldots 5$ and $\gamma = 2^{-2}, 2^{-1}, \ldots, 2^{7}$. A 5-fold cross validation was performed at every time step to optimize the SVR parameters. See Section 3.2.3 for further details on the SVR forecasting model.

A selected month within the AEMO data was used to generate Figure A-1, which shows that a minimum error can be achieved if the training window is selected appropriately for SVR. This is similar to Figure 3-1 and Figure 3-2, in which the effect of the training window length is shown on AR forecasting performance.

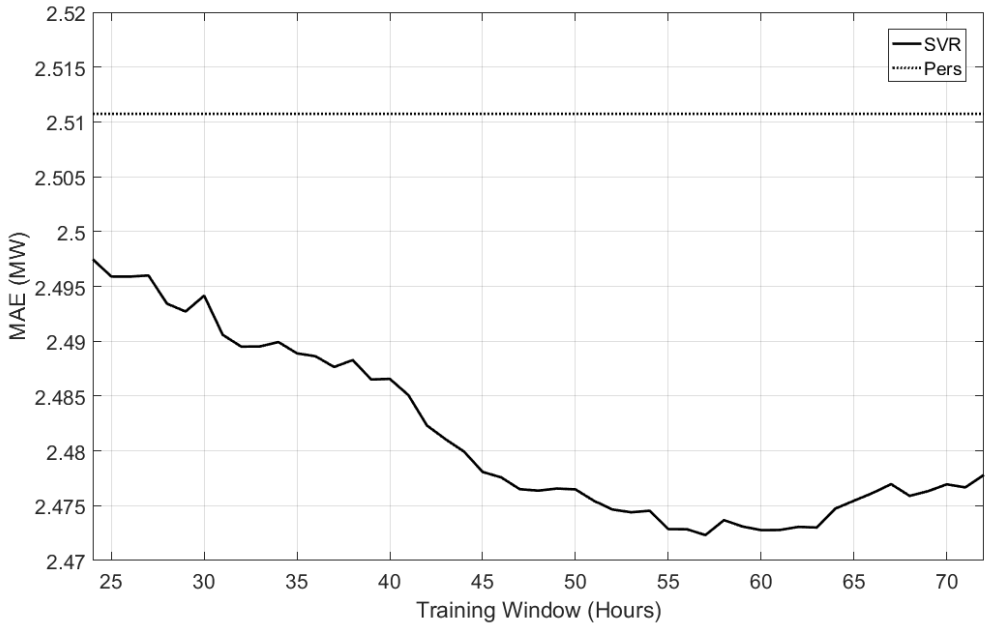


Figure A-1 MAE vs. Training Window Size for SVR, 1st Subset of Data

II. Stopping Criteria for Sifting in IMF Construction

Recall that the two criteria for defining an IMF are as follows: 1) the number of extrema and zero crossings must equal or differ at most by one and 2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. Implementing the check for the first criterion during the sifting process is straightforward and comes with no issues, but strictly enforcing the second criterion during sifting would result in a pure frequency modulated signal of constant amplitude. Thus, the amplitudes throughout the IMF, which carry information about the strength of an oscillation over time, would be erased. To ensure that the IMF components retain enough information on amplitude modulations, a stopping criterion for the sifting process was created. The criterion is a limit on the standard deviation, SD, computed from two consecutive sifting results defined as

$$SD = \sum_{t=0}^{T} \frac{\left|h_{1,k-1}(t) - h_{1,k}(t)\right|^2}{h_{1,k-1}(t)} < L_{sd} \quad \text{(A-1)}$$

where the limit L_{sd} is typically set between 0.2 and 0.3.

Although the SD-based stopping criterion worked, further improvements to the method were made by introducing a new approach that was meant to guarantee globally small fluctuations in the mean while taking into account locally large fluctuations [133]. This requires the definition of what is called the mode amplitude $a(t) \coloneqq \frac{e_{max}(t) - e_{min}(t)}{2}$ where $e_{max}(t)$ and $e_{min}(t)$ are the upper and lower envelopes of the signal, respectively. Then, the evaluation function is defined as

$$\sigma(t) \coloneqq \left|\frac{m(t)}{a(t)}\right| \quad \text{(A-2)}$$

where sifting continues until $\sigma(t) < \theta_1$ for some fraction $(1 - \alpha)$ of the total duration and $\sigma(t) < \theta_2$ for the rest of the signal. This allows the sifting procedure to preserve some large variations that may exist for some percentage θ_1 of the time series. Typical values are $\alpha = 0.05, \theta_1 = 0.05,$ and $\theta_2 = 10\theta_1 = 0.5,$ which are the default values used throughout the entire study with the EDNS metric.

III. Comments on the HHT Hilbert amplitude spectrum Output

Notice that the Hilbert amplitude spectrum for the entire example signal is not 'correct' in that it does not correctly identify the underlying, piecewise creation of the example time series, with the example time series and Hilbert amplitude spectrum shown again in Figure A-2 and Figure A-3 for the reader's convenience. Though it shows the existence of a 1/20 Hz frequency during the first 200 seconds and variable frequencies through the last 200 seconds due to the random walk process, it fails to show the existence

of the 1/20 Hz frequency through the entire time series. In addition, the spectrum shows multiple low frequency components where there were none explicitly introduced in the creation of the signal. These errors are introduced because of the underlying assumption in the EMD process that the signal is composed entirely of oscillatory modes (the IMFs), and this piecewise example time series does not agree with this assumption. This is one of the fundamental downfalls of the EMD process, which is indeed still being improved upon by the original authors. However, this behavior is acceptable in determining the stationarity of a process, especially as the EDNS reaches high values when EMD is used to arrive at the Hilbert amplitude spectrum.

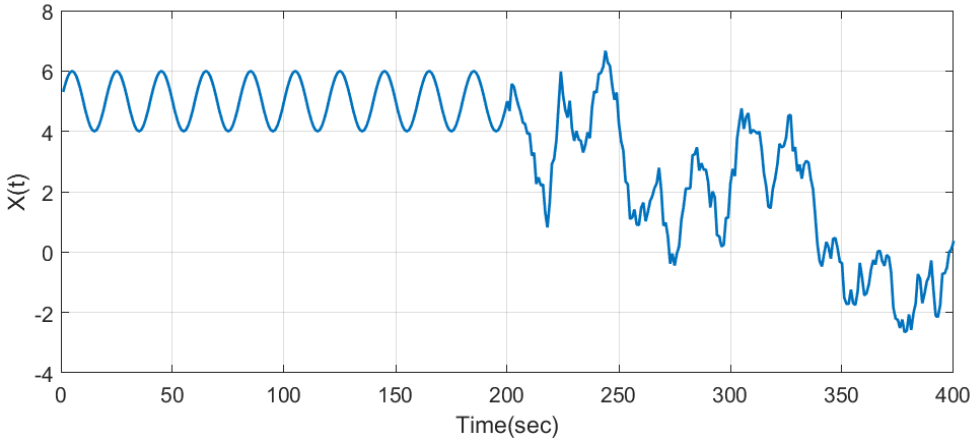


Figure A-2 Example Time Series Signal, Repeated from Figure 3-5 for Convenience

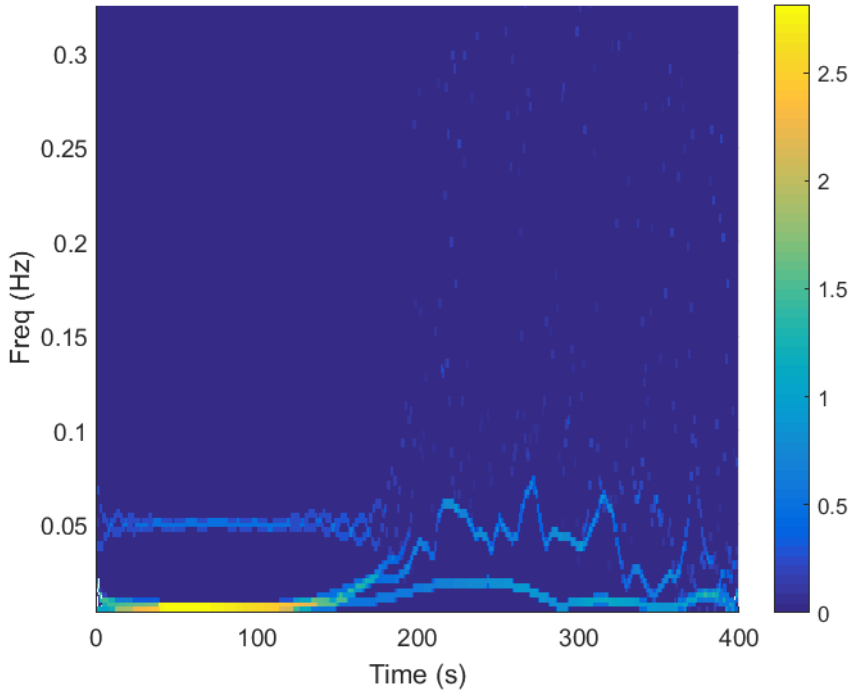


Figure A-3 Hilbert Amplitude Spectrum for the Example Signal, Repeated from Figure 3-11 for Convenience

IV. PDFs of EDNS for Different Training Windows

The probability density functions (PDFs) for the EDNS values in the AEMO wind farm power output data for various lengths of training data are shown in Figure A-4, for a set epsilon of 27. The PDF spreads out over larger sections of data, which indicates that, for a set epsilon over a short period of time, the EDNS values are very similar. If using the dynamic window algorithm proposed in this work, the PDF results suggest that the optimal training windows only fluctuate very little within a small time frame, though optimal training windows may change substantially over a long time frame.

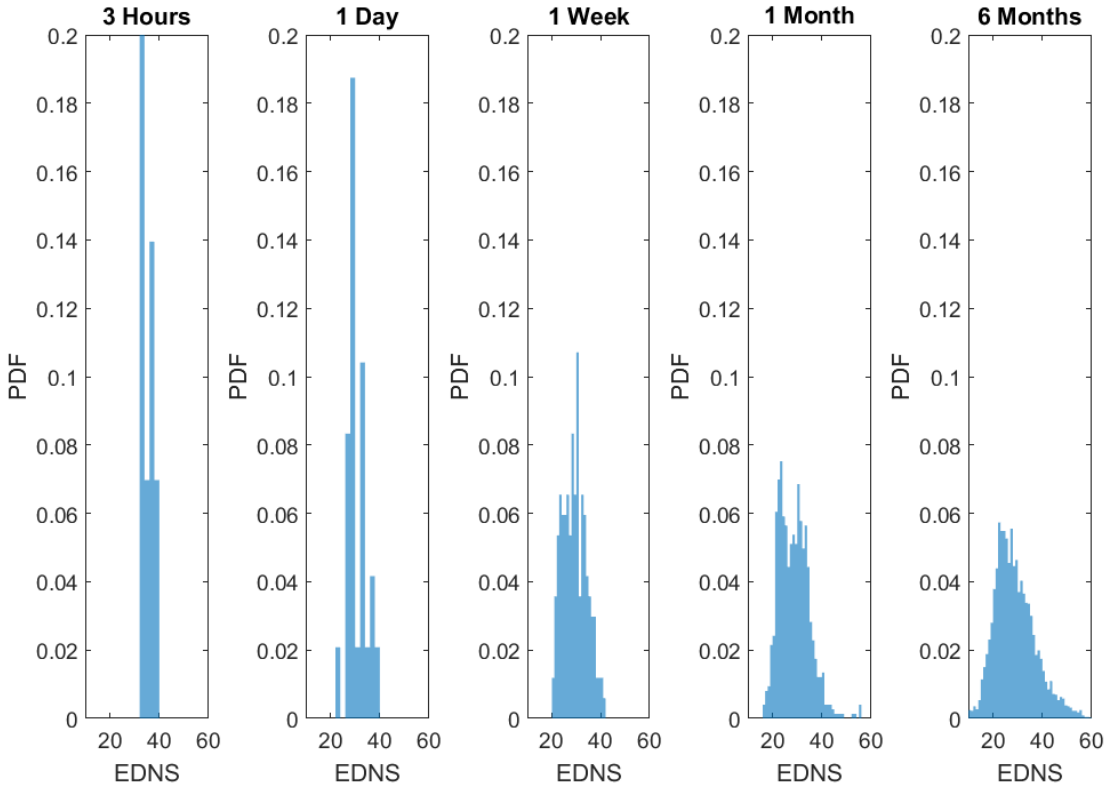


Figure A-4 PDFs of the EDNS Values for Different Sizes of Training Data Sets

APPENDIX B

NOTES ON POWER SYSTEM CONTRUCTION

I. Transmission Line Impedance Calculations

A simplified power system based off of the geography of Texas was created as shown in Figure B-1, with labeled bus numbers corresponding to the one-line diagram bus labels in Figure 4-4. Each transmission line was assumed to be a 500 kV line with a reactance of 0.5 Ω/mile. Each path between connected buses were assumed to contain two transmission line connections as shown in Figure B-1, except for the Lubbock-Dallas connection and the Sherman-Dallas connection. Approximate distances between pairs of buses and equivalent reactances between those buses are shown in Table B-1. Note that $Z_{base} = \frac{V_{base}^2}{S_{base}} = \frac{(500\text{ kV})^2}{25000\text{ MW}} = 10\Omega$, and that X_{pu} is the equivalent per-unit reactance of the path between a pair of buses.

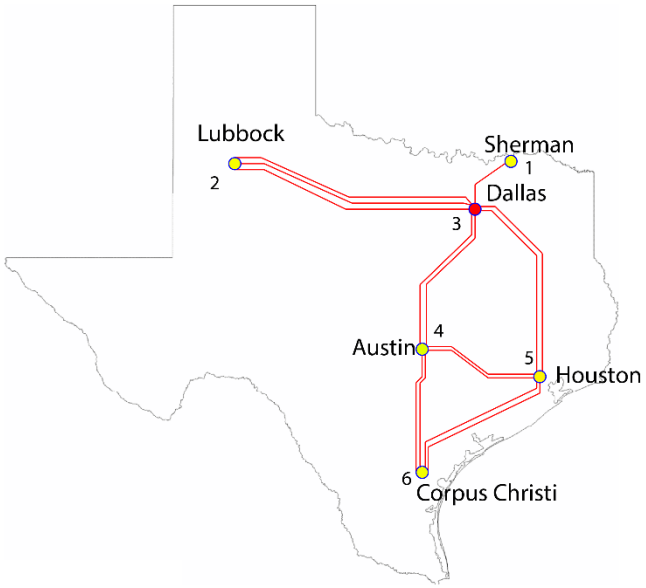


Figure B-1 The 6-bus System Based Loosely Off of the Geography of Texas

Table B-1 Approximate Distances Between Buses

Bus Pair	Distance (miles)	Total Reactance per Line (Ω)	Total Reactance of Path (Ω)	X_{pu} (System Base)
1-3	100	50	50	5
2-3	300	150	50	5
3-4	200	100	50	5
3-5	200	100	50	5
4-5	200	100	50	5
4-6	200	100	50	5
5-6	200	100	50	5

Thus, the value K is calculated to be $K = \frac{1}{X_{pu}} = 0.2$ pu for each path.

II. Measured Rotor Speeds and Tie-line Flows in Single-Area System

A frequency time series plot is shown in Figure B-1 and Figure B-2 for a 0.01 pu generation decrease at bus 2. Buses 1 and 2 show larger oscillations while the other buses show smaller oscillations about the same path. Figure B-3 shows the power across the paths between connected buses, showing the larger oscillations across the power lines between buses 1 & 2 and 2 & 3.

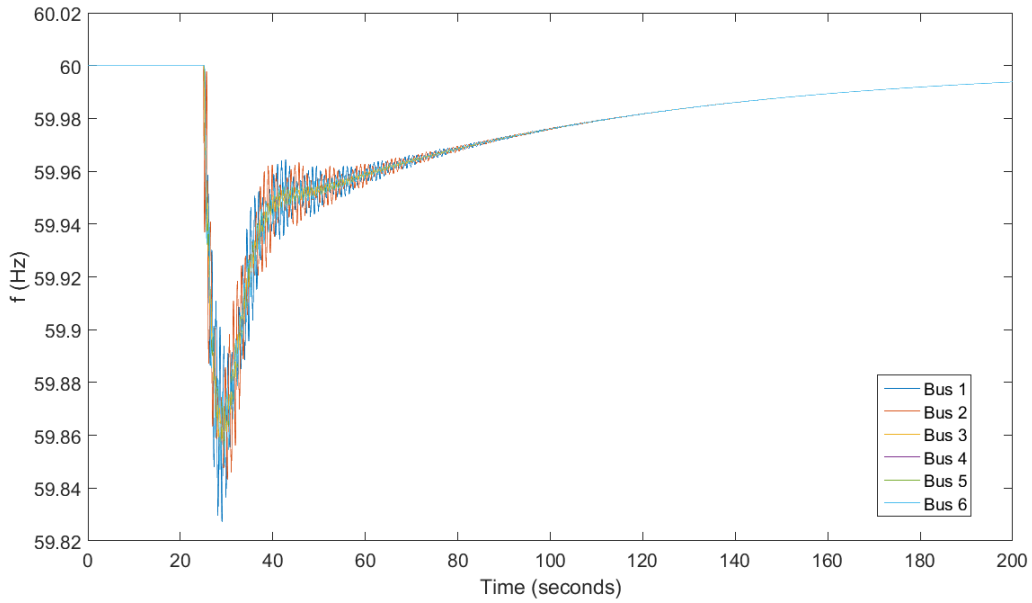


Figure B-2 Frequencies at Every Bus after a 0.01 pu Disturbance at Bus 2

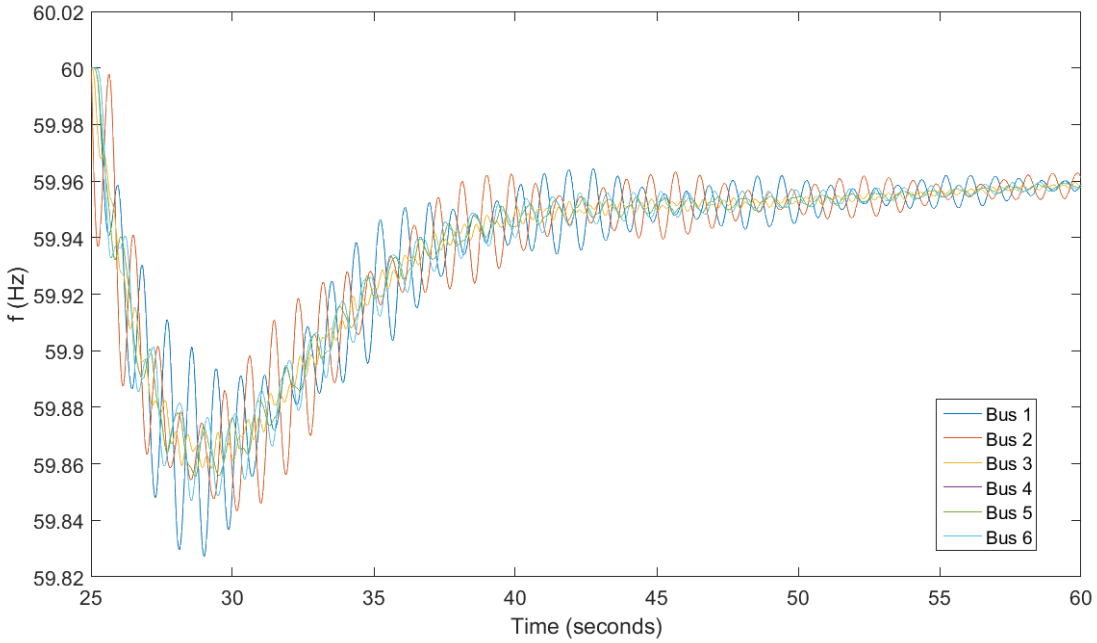


Figure B-3 Frequencies at Every Bus after a Disturbance at Bus 2

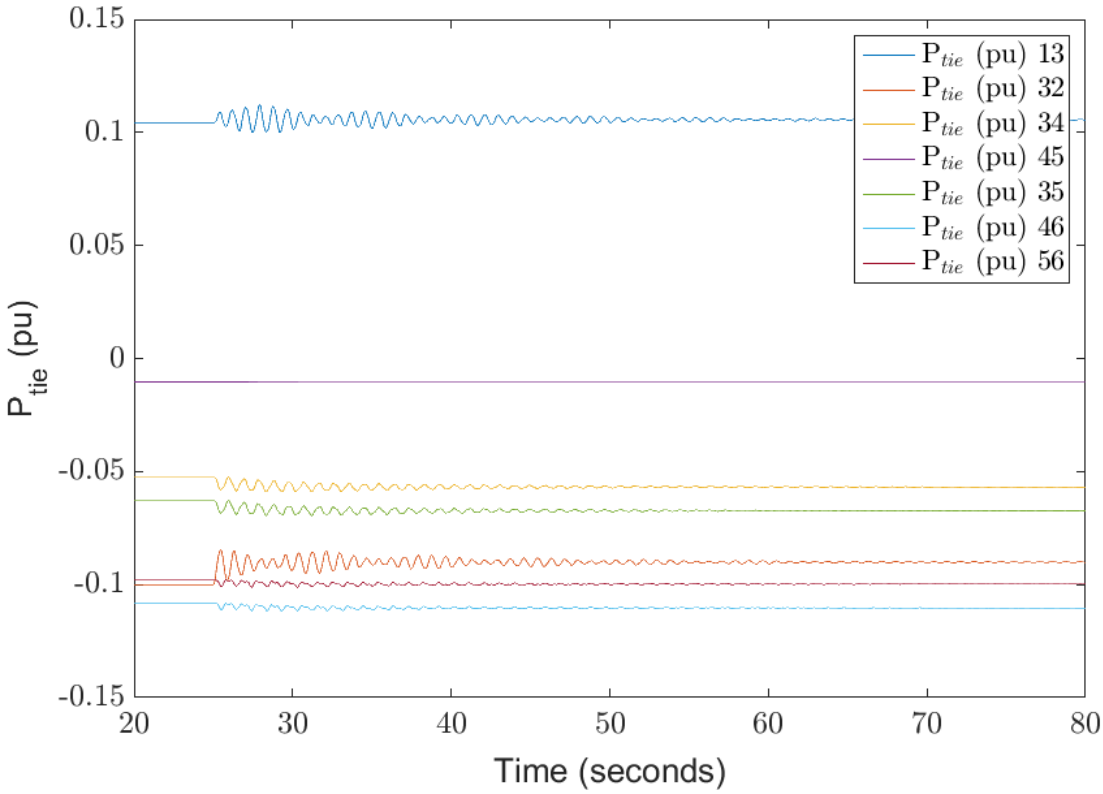


Figure B-4 Power (pu) Across Transmission Line Paths in the System

III. Two-Area System Rotor Rotational Speeds

For the two-area system in Section 4.3.2, a frequency time series plot showing frequency at all 12 buses throughout the two areas is shown in Figure B-5 and Figure B-6. This is in response to a sudden 250 MW (0.01 pu) generation decrease at bus 2 in area 1 at t = 25 seconds. Buses 1 and 2 of Area 1, in red and blue as denoted in Figure B-5, show larger oscillations as they are electrically closer to the generation decrease.

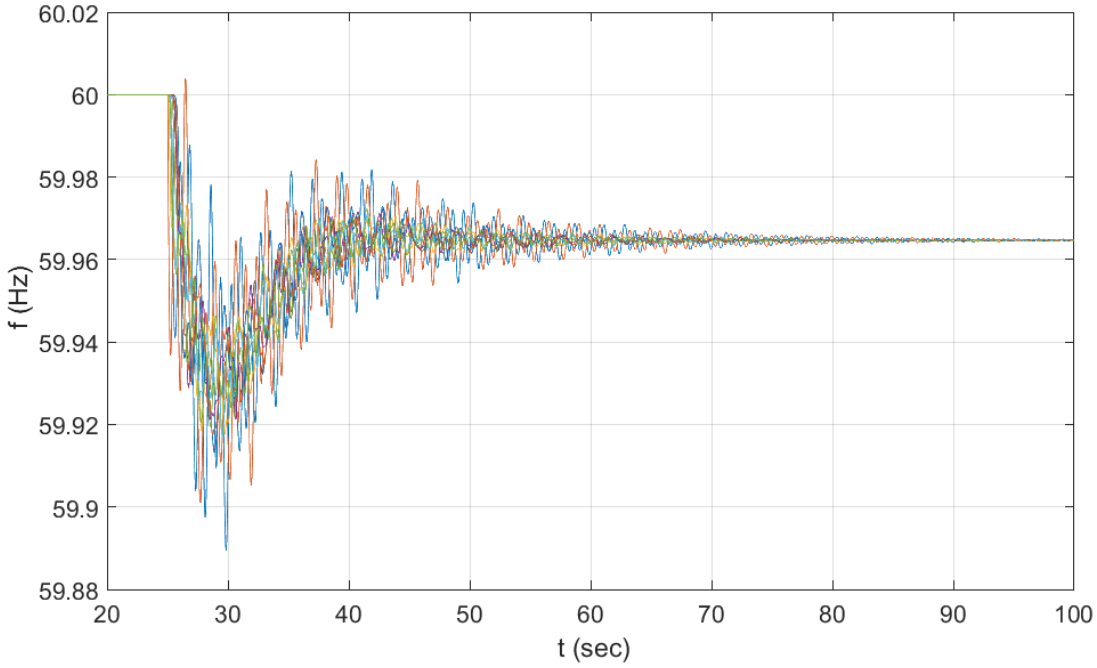


Figure B-5 Frequency Plot of the 2-area System in Response to a Sudden Generation Decrease

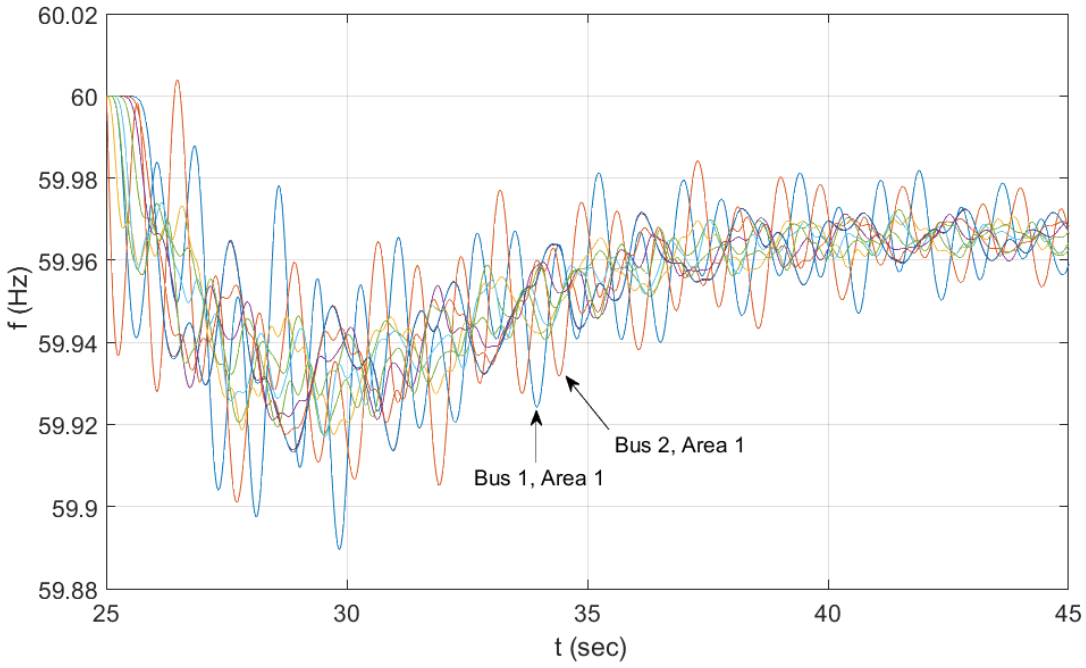


Figure B-6 A Close-up of the Two-area Frequency Time Series Immediately after the Generation Decrease

IV. Effects of Frequency Bias Values in Secondary Control

The value of the frequency bias term B can change the behavior of secondary control in a system. Figure B-7 shows the calculated ACE values for the two-area system with AGC turned off when a 250 MW load increase is seen at bus 2 in area 1 using the theoretical optimal value of $B = 354.2 \frac{\text{MW}}{0.1\ \text{Hz}}$. Notice that area 1's ACE shows that area 1 shows a 250 MW generation deficiency, while area 2 correctly shows that there is no increase in generation needed at steady state. With other values of B, an area's ACE would no longer represent the true MW deficiency in that area *in the steady state*. Figure B-8 shows the ACE values with $B = 250 \frac{\text{MW}}{0.1\ \text{Hz}}$ and $B = 450 \frac{\text{MW}}{0.1\ \text{Hz}}$, where the ACE values in either area do not reflect the true MW deficiency in the area.

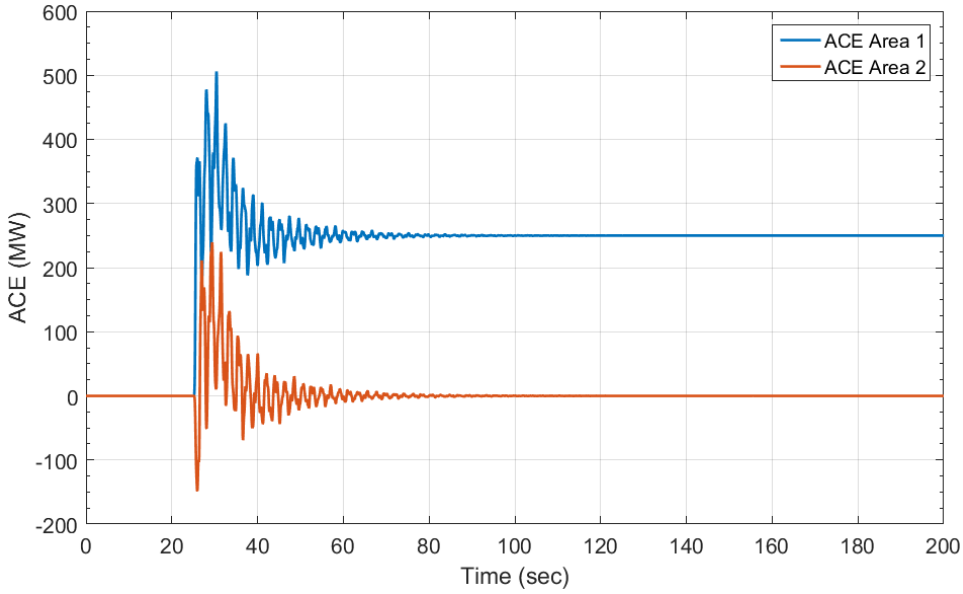


Figure B-7 ACE values in Both Areas with B $= 354.2 \frac{\text{MW}}{0.1\ \text{Hz}}$

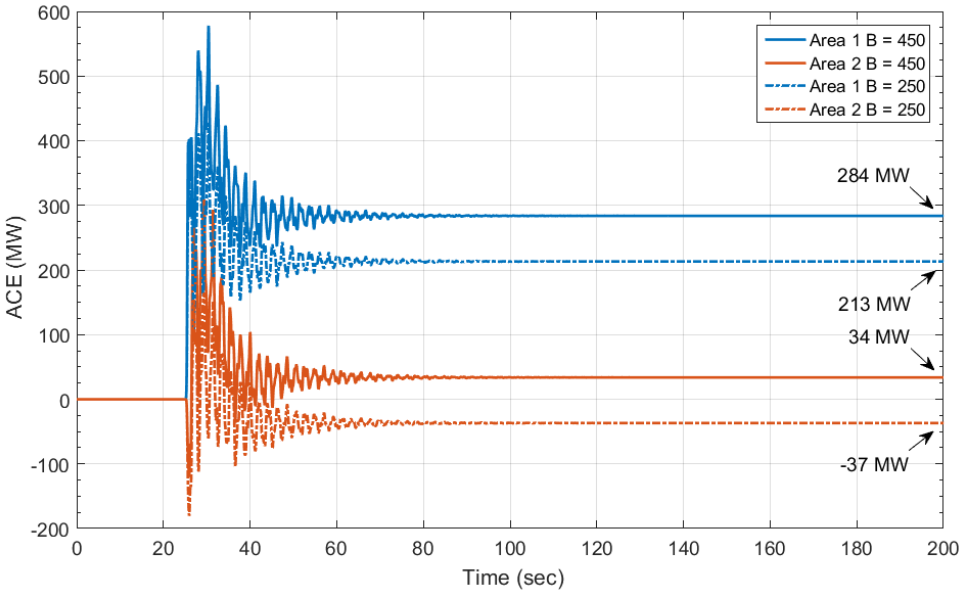


Figure B-8 ACE Values in Both Areas with B $= 250 \frac{\text{MW}}{0.1\ \text{Hz}}$ and $450 \frac{\text{MW}}{0.1\ \text{Hz}}$

Frequency and tie-line results are shown in Figure B-9 and Figure B-10 for values of frequency deviation and tie-line gains. To allow for comparisons of these gains, note that the integral controller has a gain of K_i, and so acts on the value $K_i(\Delta P_{tie} -$

$10\mathrm{B}\Delta f_{sys}\big) = K_i\Delta P_{tie} - 10\mathrm{B}K_i\Delta f_{sys}$. To show the system's sensitivity to the value of frequency bias, a range of B values will be tested keeping K_i constant, and to show sensitivity to tie-line gains, K_i will be swept while keeping $10\mathrm{B}K_i$ constant.

Various frequency deviation gains are shown in Figure B-9 and Figure B-10 for values of $\mathrm{B} = 25, 100, 354.2, 1000, 4500$ with a constant integral gain of $K_i = 0.0015$. Various tie-line gains are shown in Figure B-11 and Figure B-12 for values of $K_i = 0.00015, 0.001, 0.0015, 0.002, 0.01$ with a constant value $10\mathrm{B}K_i = 10 * 354.2 * 0.0015 = 5.313$.

As expected, the frequency is not very sensitive to changes in the effective tie-line deviation gain, and the tie-line flow is not very sensitive to changes in the frequency bias. However, because there was gain placed on both ΔP_{tie} and Δf_{sys}, both the tie-line flow and system frequency eventually returns to nominal. This shows that turbine outputs in both areas fall to their correct final outputs (where area 1 turbines pick up all of the load increase).

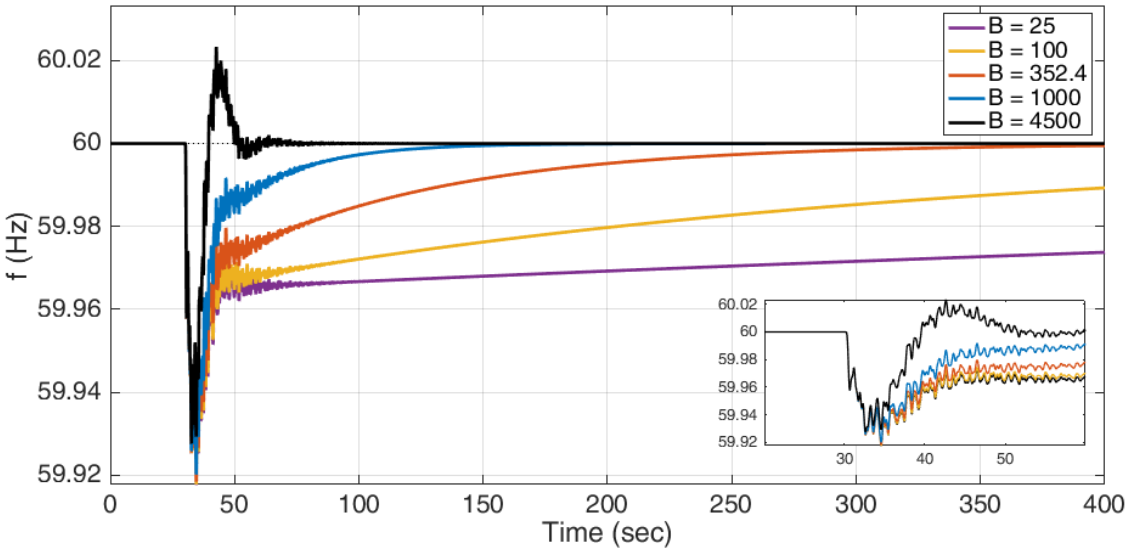


Figure B-9 Frequencies for Various Values of B

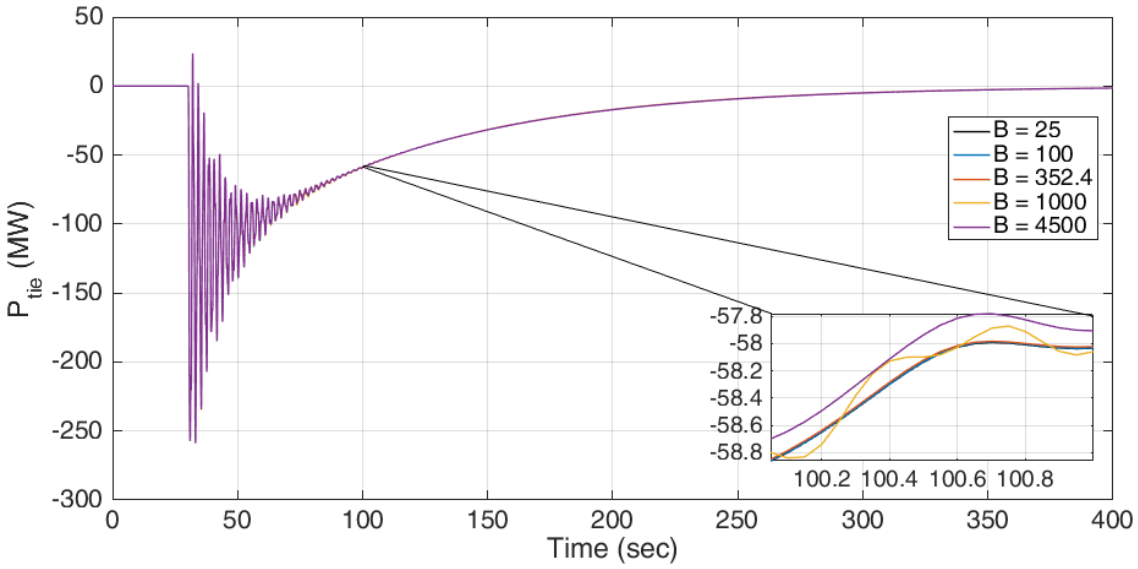


Figure B-10 Tie-line Flow Out of Area 1 for Various Values of B

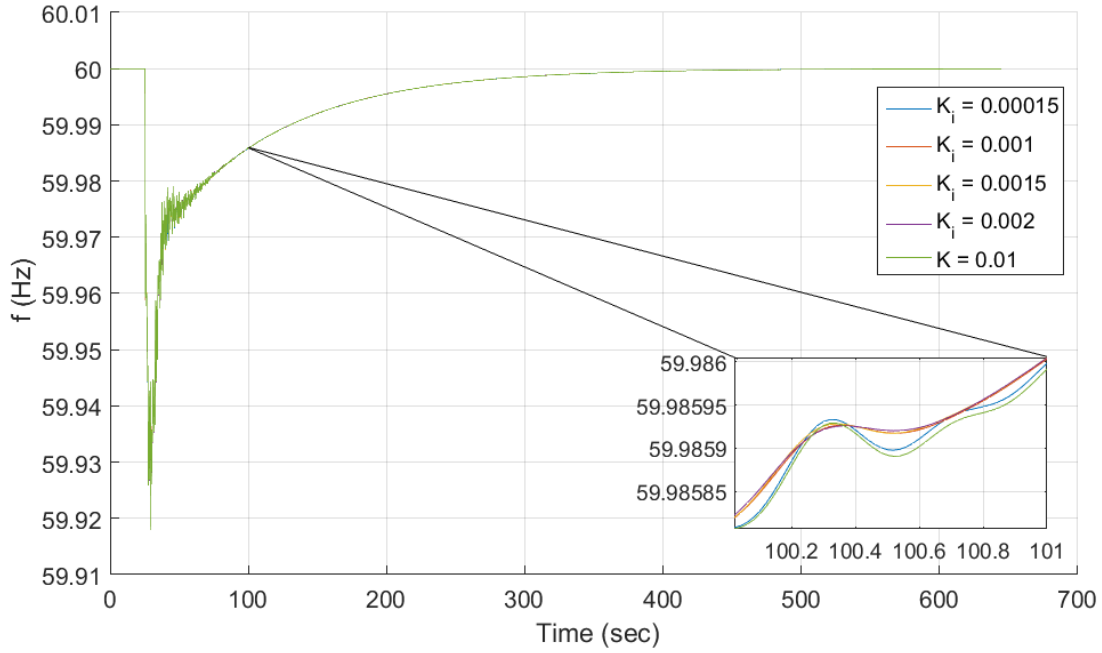


Figure B-11 Frequency for a Sweep of Effective Tie-line Gain

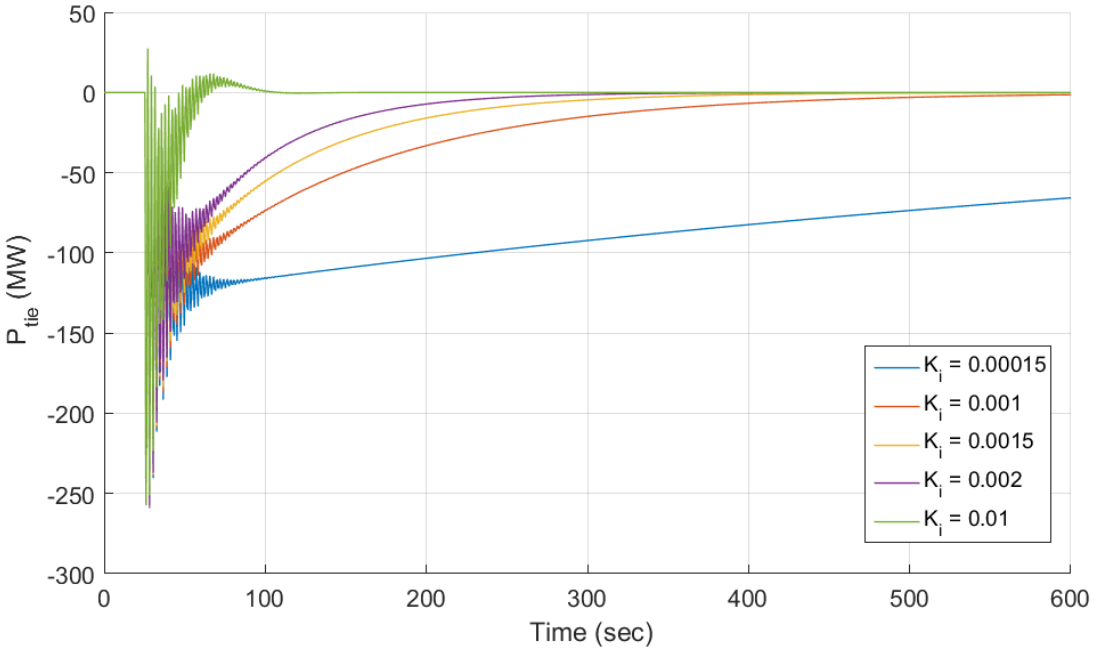


Figure B-12 Tie-line Flow Out of Area 1 for a Sweep of Effective Tie-line Gain

Plots for a sweep of integral gain in the AGC controller, holding the frequency bias value constant, are shown in Figure B-13 and Figure B-14.

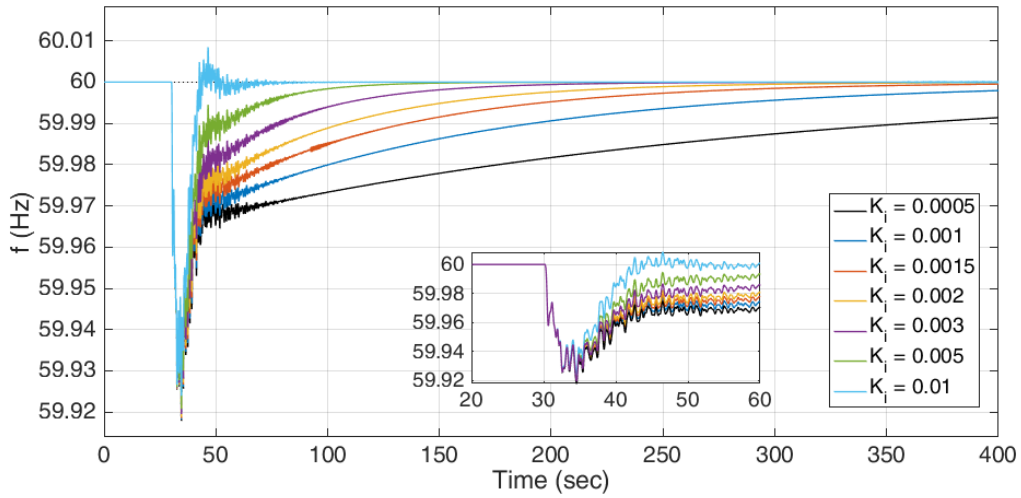


Figure B-13 Frequency for a Sweep of Integral Gain in the AGC Controller

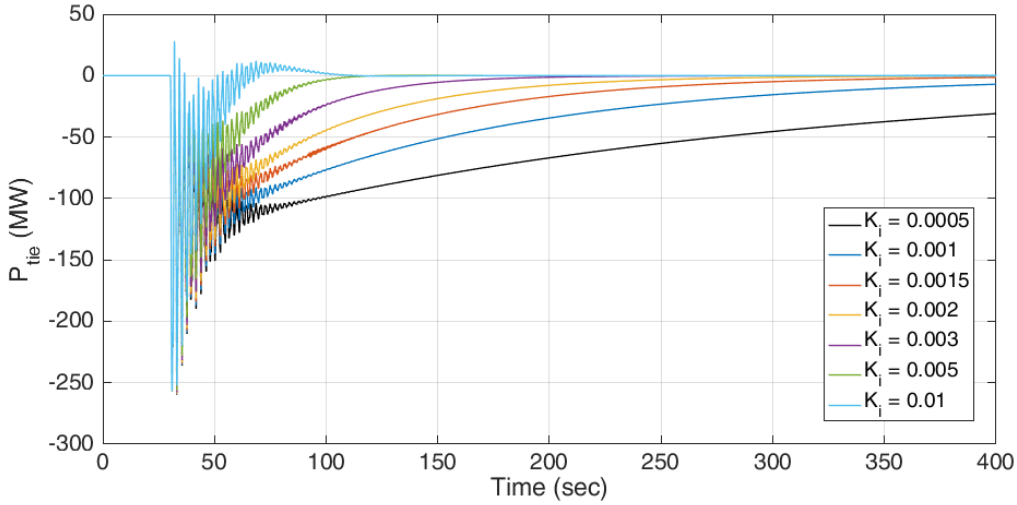


Figure B-14 Tie-line Flows for a Sweep of Integral Gain in the AGC Controller

V. Effects of Noise Covariance Values in State Estimation

To study the effect of changing the noise covariance values on the performance of the MPC controller, a single-area MPC controller was used with the simulation parameters shown in Section 4.4.2(ii), which involve a sudden drop in generation at bus 2 in the single-area power system model. In this simulation, a 0.01 pu drop is expected, while an actual 0.02 pu drop occurs. Figure B-15 shows the state estimate over time for the disturbance forecast error ε_d, for varying values of the noise parameter σ_d in the Q_{kal} noise covariance matrix, while the corresponding system frequencies are shown in Figure B-16. Recall that the noise parameter σ_d is a user-defined input that changes the Kalman gains, such that a larger noise parameter results in a greater weight placed on measurements instead of the process model. Because the unmeasured disturbance is estimated with much greater accuracy with large noise parameter, the frequency recovers quicker when a large noise parameter is used.

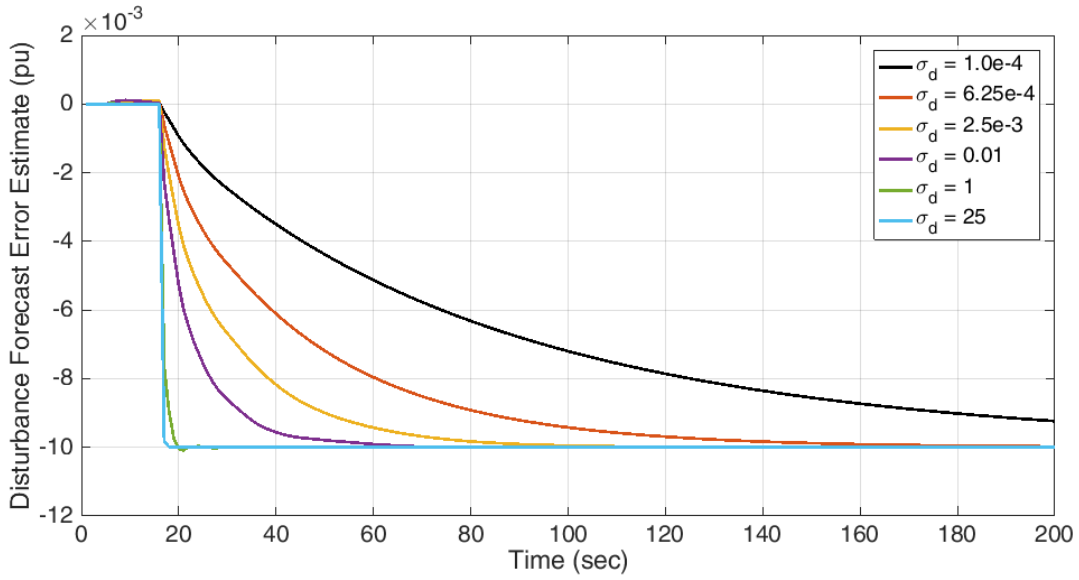


Figure B-15 Error Estimate for the Disturbance Forecast for Varying Modeled Noise Values

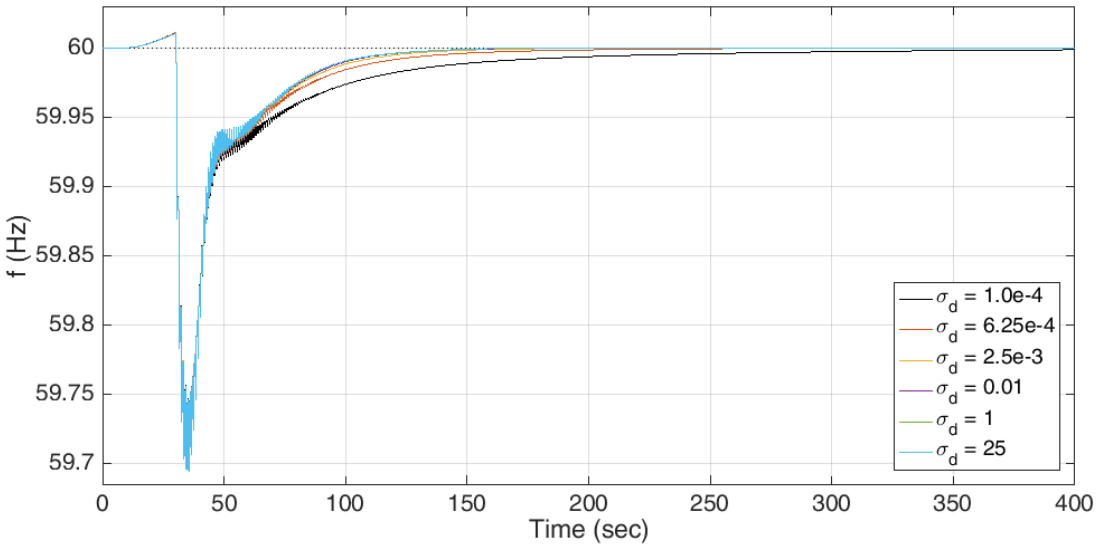


Figure B-16 Error Estimate for the Disturbance Forecast for Varying Modeled Noise Values

VI. Discussion on Determining Balanced Weights of Q_f and Q_{ptie}

As shown in the multi-area Chapter 4 simulations, the MPC objective function weights Q_f and Q_{ptie} have a large influence on the performance of anticipatory secondary control, and especially determine the importance of inter-area communication of the frequency set points and forecasted disturbances. It is shown in Section 4.4.3 that the balanced weights of $Q_f = 0.75$ and $Q_{ptie} = 3.5$ result in an MPC controller that responds effectively to disturbances within its own area and responds minimally to disturbances outside of the area. This weighting was arrived at heuristically, by setting an appropriate value of Q_f, given static S_f and R_f weights, and then sweeping over Q_{ptie} values and examining the frequency set point movements.

The dependence of the ratio Q_f/Q_{ptie} to the controller performance may be compared to the dependence of the frequency bias term B to conventional AGC's integral control. With conventional AGC, an area should not respond significantly to disturbances outside of its own area, which is ensured by weighting the deviation in frequency Δf term with B such that it matches the deviation in tie-line flow ΔP_{tie} for a disturbance outside of the area. This is reviewed in Appendix B IV, and the calculation of B is only dependent on the area's D and R values. However, a theoretical value for the ratio Q_f/Q_{ptie} that results in an anticipatory controller that only largely reacts to disturbances within its own area may be dependent on many factors within the system, and therefore the balanced ratio of Q_f/Q_{ptie} may be most effectively calculated heuristically by sweeping over various values.

To illustrate the complexity of calculating a theoretical balanced Q_f/Q_{ptie}, recall the MPC objective function

$$\min_{f_{set_1} \dots f_{set_m}} J = \left(\sum_{i=1}^{p} Q_f \left(\Delta f_{sys}\right)^2 + Q_{ptie} (\Delta P_{tie}^m)^2 \right) + \left(\sum_{j=1}^{m} R_f (f_{set})^2 + S_f \left(\dot{f}_{set}\right)^2\right) \quad \text{(B-1)}$$

which shows that f_{set} is manipulated such that the function J is minimized. The objective of creating a balanced ratio of Q_f/Q_{ptie} involves picking Q_f/Q_{ptie} such that $f_{set} = 0$ throughout the control horizon m in response to frequency and tie-line movements caused by a disturbance outside of the area. The integrated of tie-line flow deviations and the integrated system frequency deviation must be minimized with $f_{set} = 0$, which depends on the rate at which the tie-line flow returns to nominal and the rate at which the frequency returns to nominal. In addition, the behavior of the tie-line flow and frequency is affected throughout time by MPC controller frequency set point movements in other areas. Lastly, because MPC calculates the values $f_{set_1} \dots f_{set_m}$ at every time step based on updated measurements, an optimal $f_{set_1} = 0$ must be ensured for every step in the simulation, as at every time step, only f_{set_1} is implemented by the MPC controller.

However, the selection of a balanced Q_{ptie} value, given $Q_f = 0.75$, can be examined and shown to vary with the effective system droop value R with various sweeps. Figure B-17 and Figure B-18 show the MPC frequency set point values in Area 1 without inter-area communication, for a 0.01 pu step loss in generation in Area 2. To select a controller that acts minimally to disturbances outside the area, $Q_{ptie} = 2.7$ may work for a droop value of 0.03 for the 30% of responsive generation in the system while $Q_{ptie} = 3.5$ works well for a droop value of 0.04.

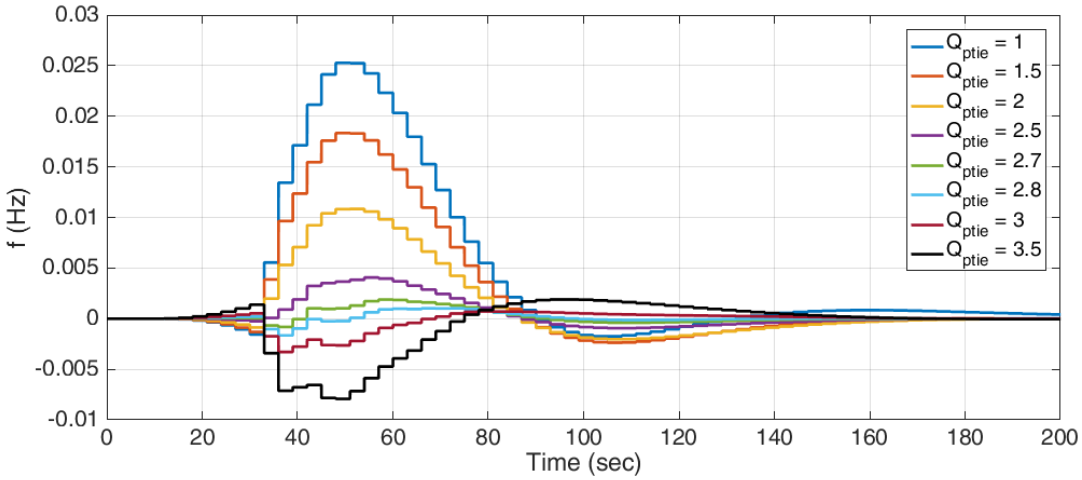


Figure B-17 Frequency set point of Area 1 with a disturbance in Area 2 with $R = 0.03$

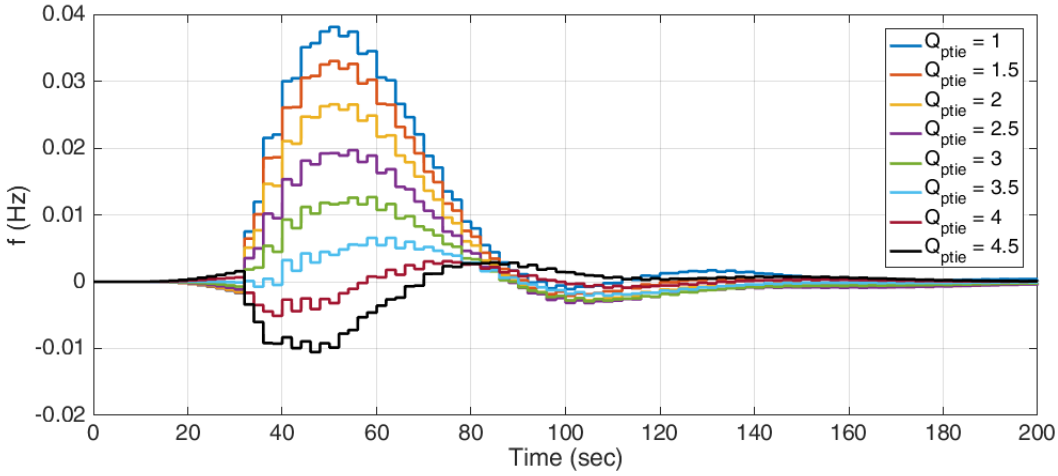


Figure B-18 Frequency set point of Area 1 with a disturbance in Area 2 with $R = 0.04$

APPENDIX C

SUPPLEMENTAL PLOTS FOR POWER SYSTEM SIMULATIONS

I. Supplementary Plots with White Noise in Load with MPC

MPC is known for its immunity to white noise, as the estimator within it (which usually uses some variant of the Kalman filter) can prevent excessive controller action in the presence of white noise. This is an important consideration for any control systems to placed in the actual power system, as the load is constantly fluctuating.

Figure C-1 and Figure C-2 compares the frequency and ACE results of a single-area system under MPC control and conventional AGC, where a zero-mean white-noise signal of a 0.0006 system pu (15 MW) variance was added to the load at every bus in the system, and the same white noise is applied to both control systems. MPC still outperforms conventional AGC, and the noise does not result in erratic movements of the MPC frequency set point.

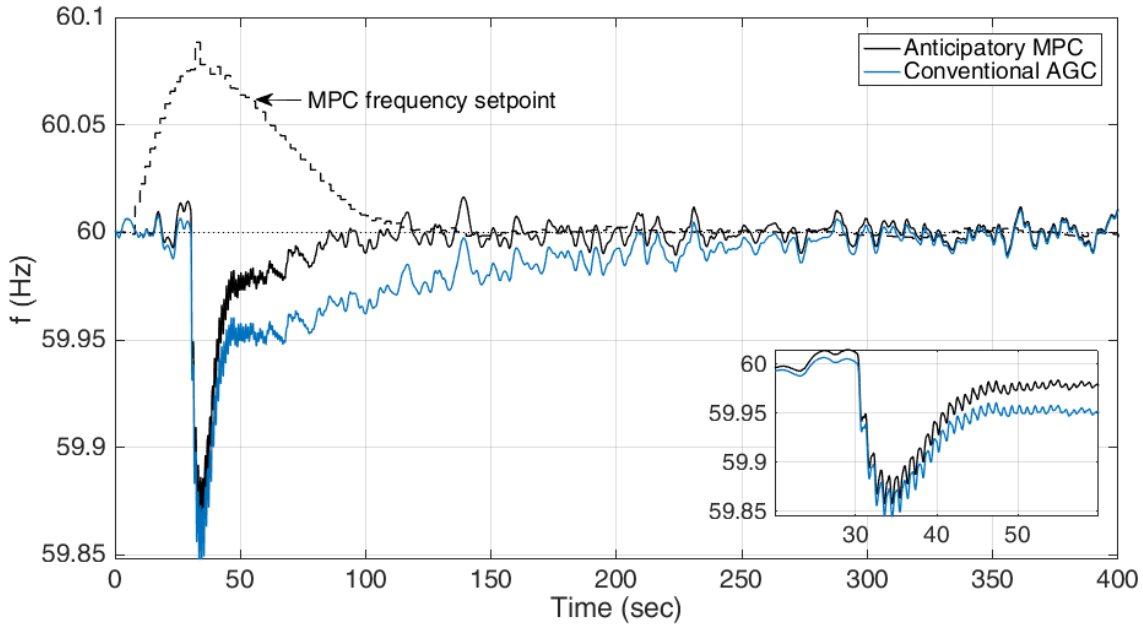


Figure C-1 Frequency in the Presence of White Noise with MPC and Conventional AGC

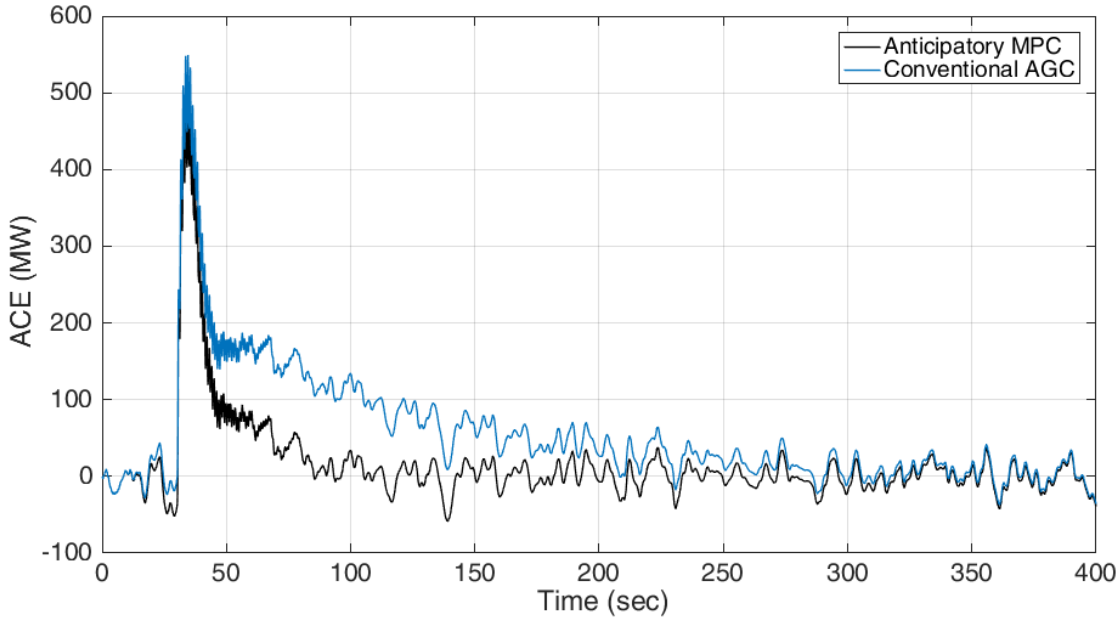


Figure C-2 ACE in the Presence of White Noise with MPC and Conventional AGC

II. Supplementary Plots for Preplanned Action

Responsive turbine power outputs and ACE plots are shown in this section for various preplanned, manual actions for the frequency set point paths shown in Figure 4-29. The plots for aggressive action are shown in Figure C-3 and Figure C-4, the plots for conservative action are shown in Figure C-5 and Figure C-6, and the plots for near optimal action are shown in Figure C-7 and Figure C-8.

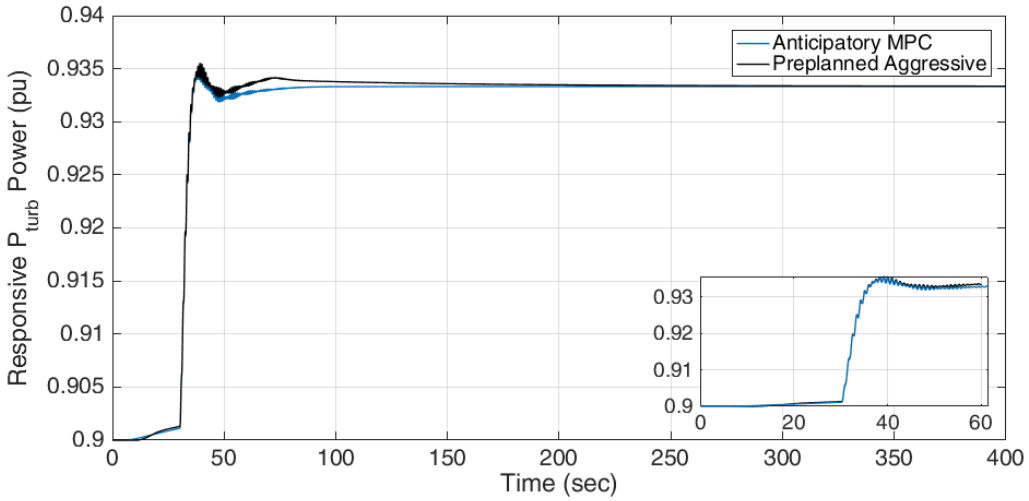


Figure C-3 Responsive Turbine Output of System with Preplanned Aggressive Action

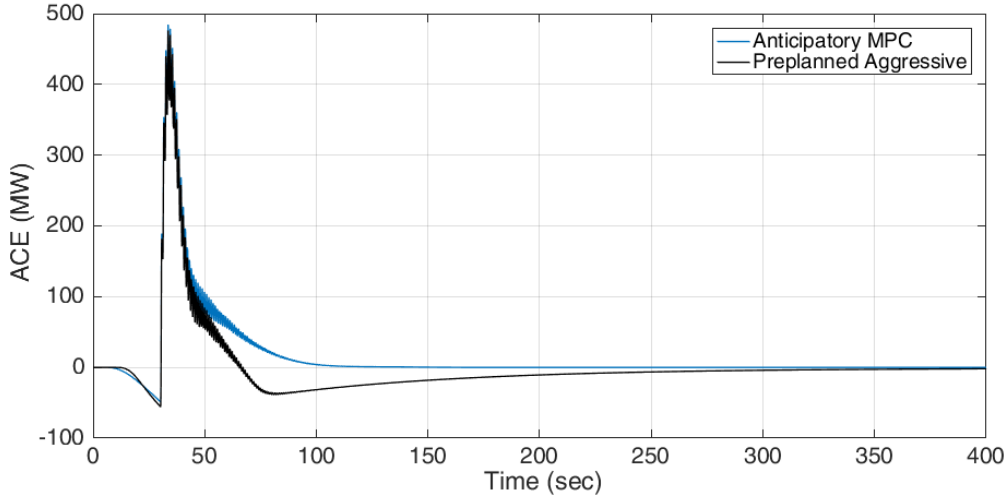


Figure C-4 ACE of System under Preplanned Aggressive Action

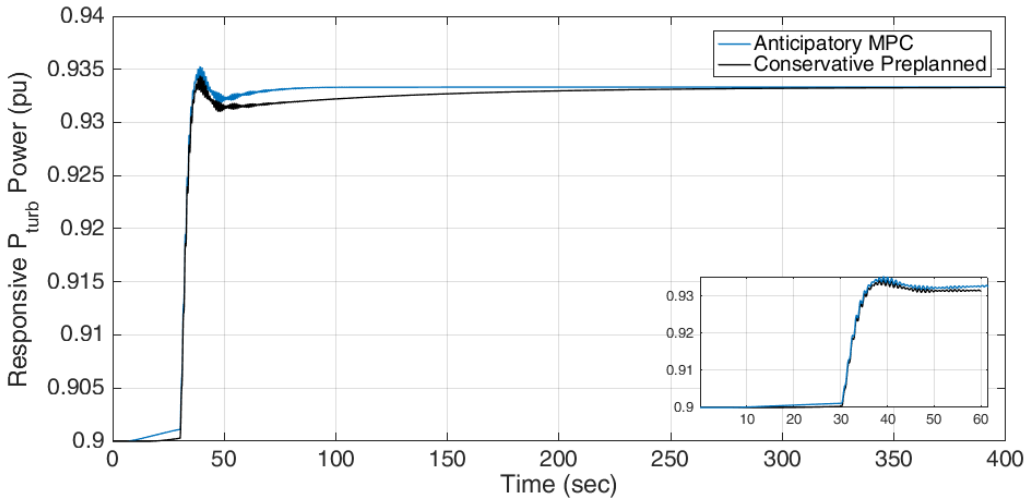


Figure C-5 Responsive Turbine Output of System with Preplanned Conservative Action

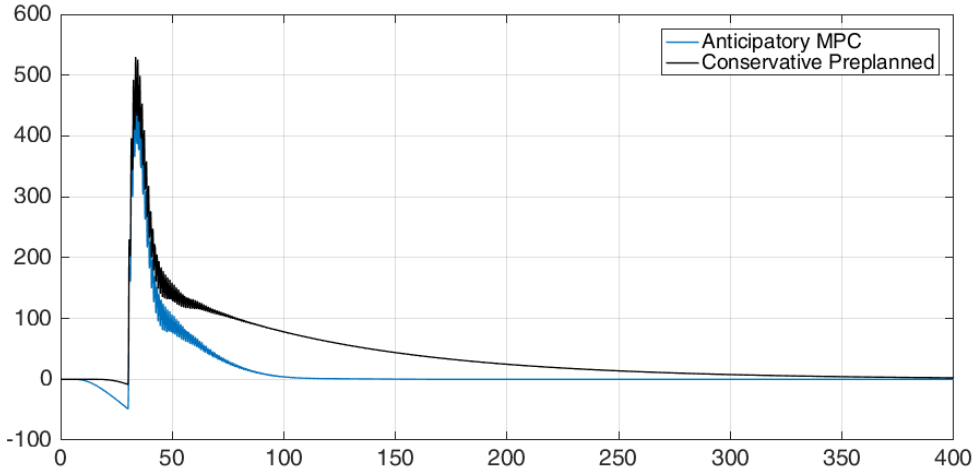


Figure C-6 ACE of System under Preplanned Conservative Action

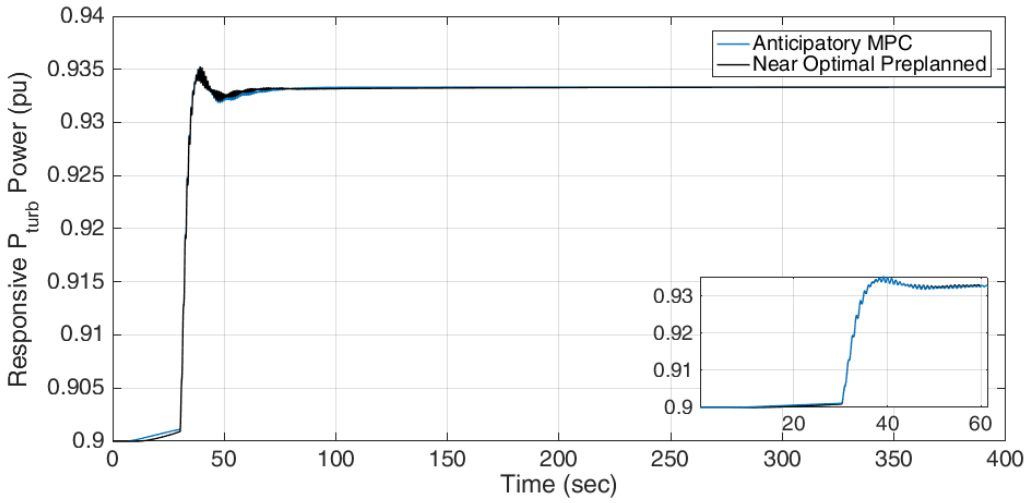


Figure C-7 Responsive Turbine Output of System with Preplanned Near Optimal Action

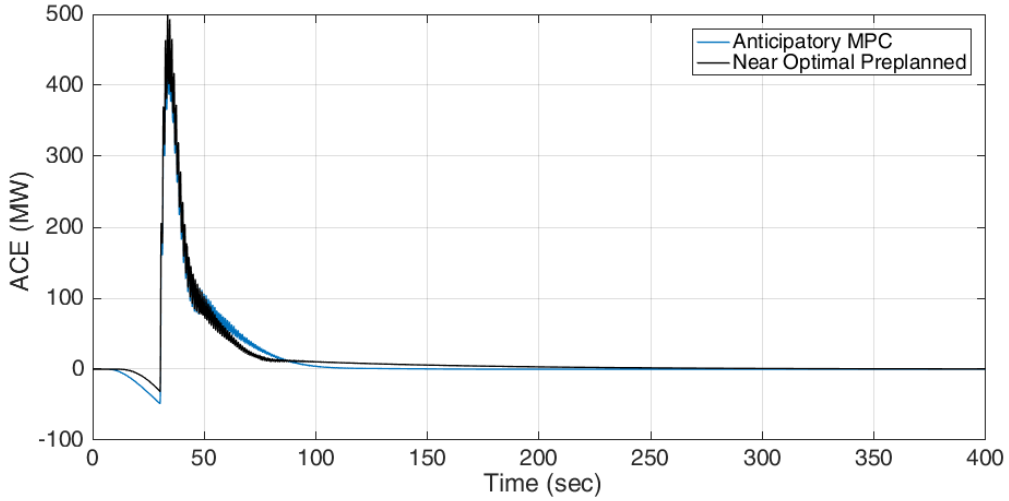


Figure C-8 ACE of System under Preplanned Near Optimal Action

III. Supplementary Plots for Single-area Disturbance with Forecast Errors

For the disturbances shown in Figure 4-30 which have errors in the magnitude estimate of the forecasted disturbance, plots for the responsive turbine power output and ACE are shown in Figure C-9, Figure C-10, Figure C-11, and Figure C-12. Timing error plots are shown in Figure C-13, Figure C-14, Figure C-15, and Figure C-16.

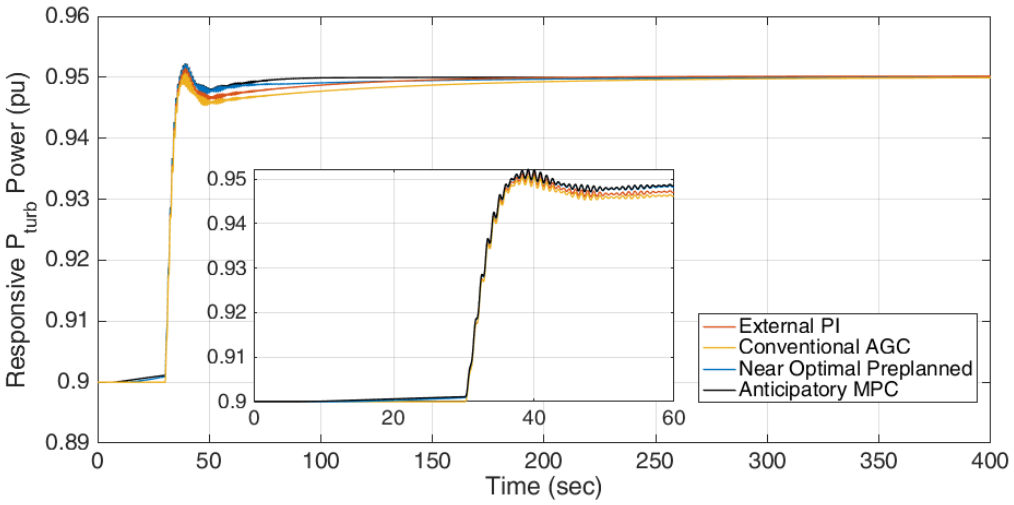


Figure C-9 Turbine Output of 1.5x Magnitude Error (Actual Drop of 0.015 pu)

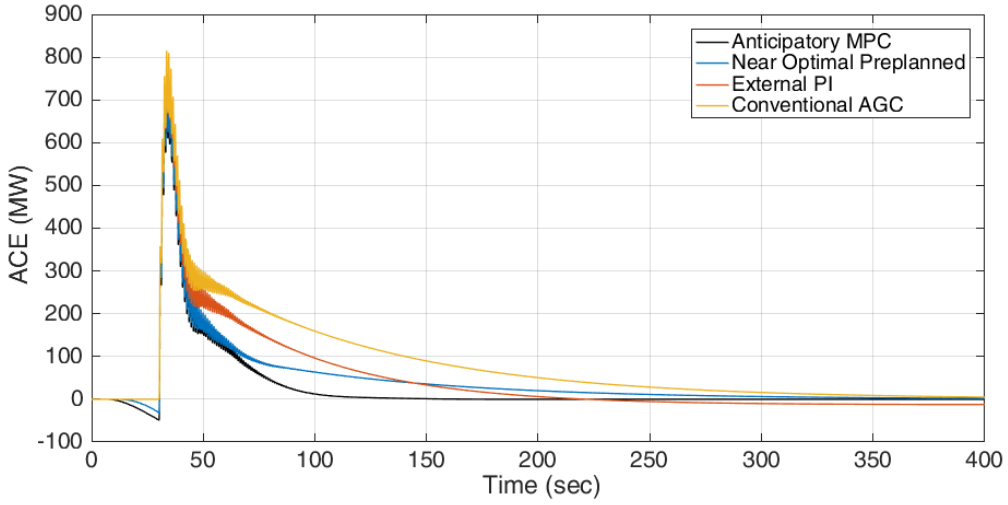


Figure C-10 ACE of 1.5x Magnitude Error (Actual Drop of 0.015 pu)

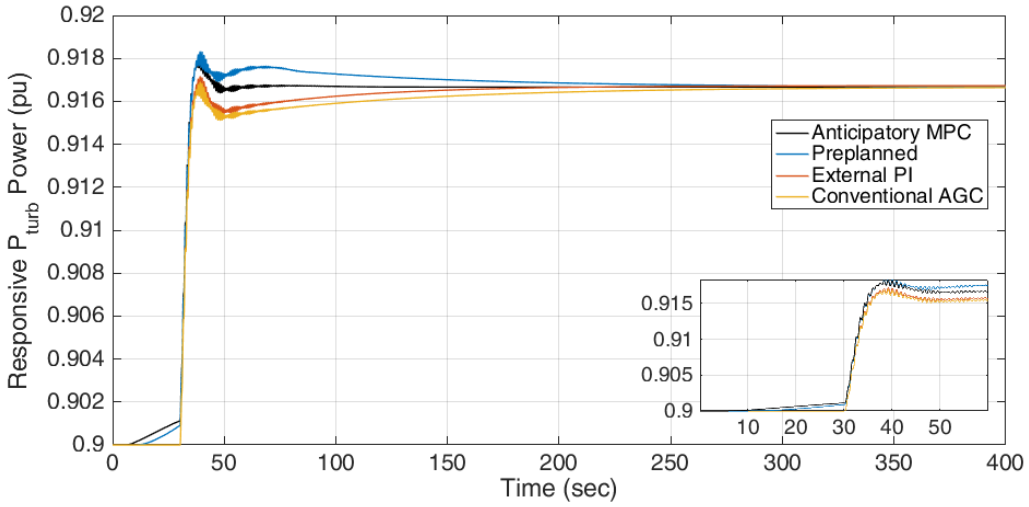


Figure C-11 Turbine Power of 0.5x Magnitude Error (Actual Drop of 0.005 pu)

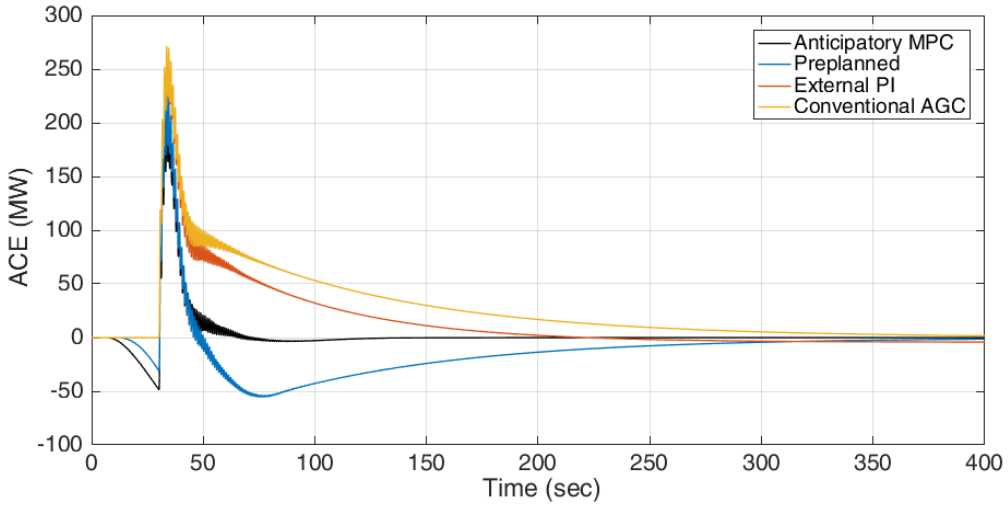


Figure C-12 ACE of 0.5x Magnitude Error (Actual Drop of 0.005 pu)

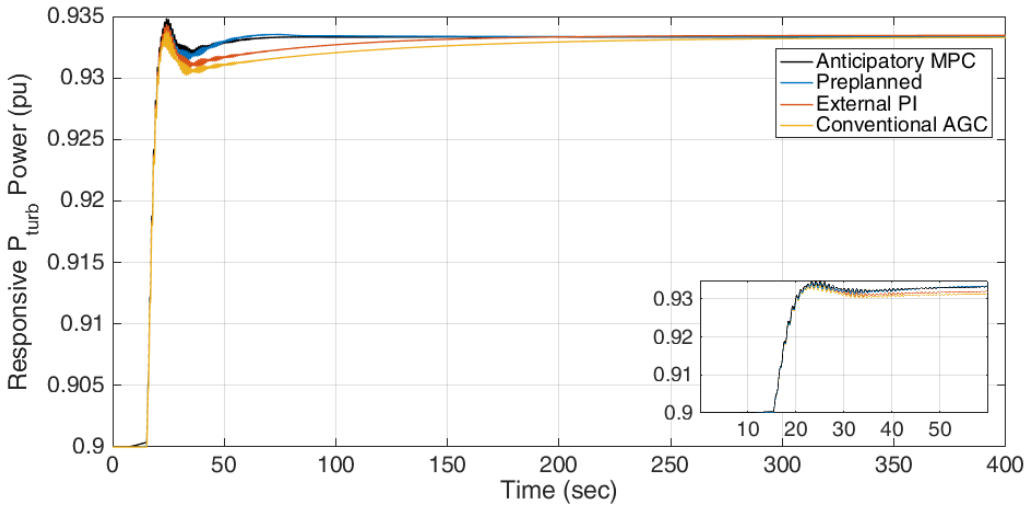


Figure C-13 Turbine Power of a System with a Timing Error in the Forecast, where the Disturbance was 15 Seconds Before the Forecasted Time

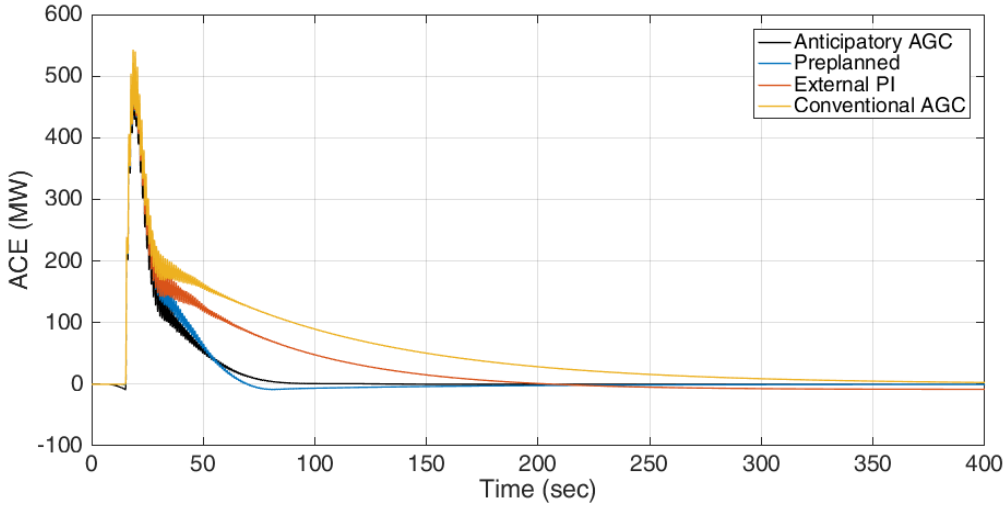


Figure C-14 ACE of a System with a Timing Error in the Forecast, where the Disturbance was 15 Seconds Before the Forecasted Time

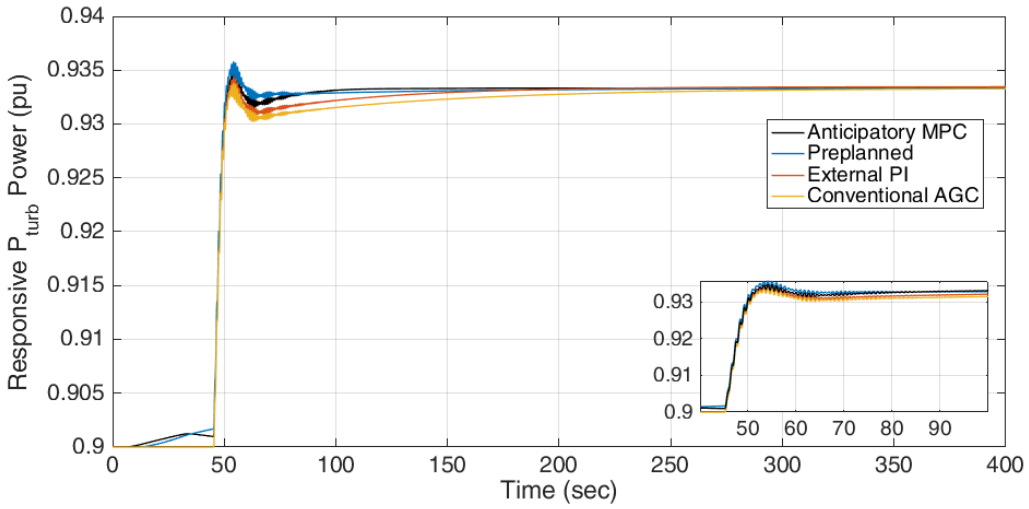


Figure C-15 Turbine Power Output of a System with a Timing Error in the Forecast, where the Disturbance was 15 seconds after the Forecasted Time

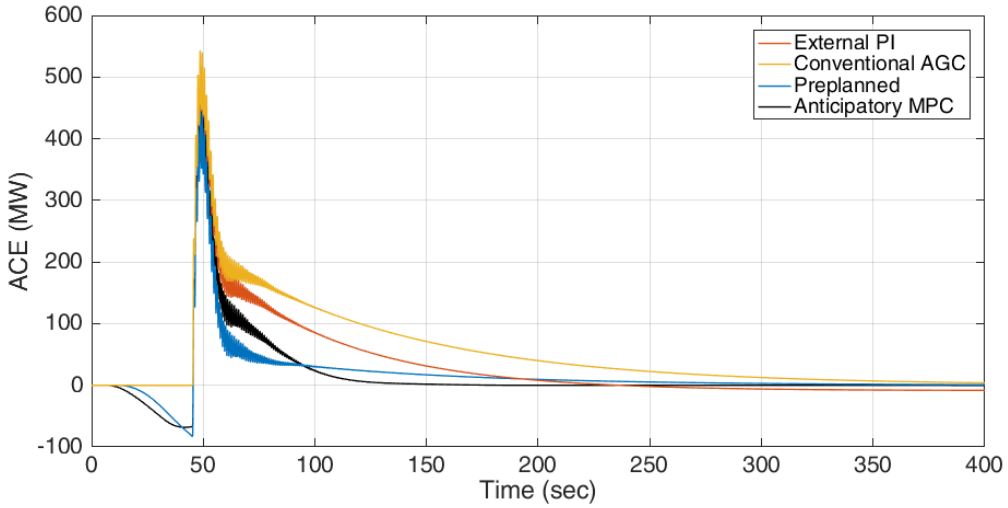


Figure C-16 ACE of a System with a Timing Error in the Forecast, where the Disturbance was 15 Seconds after the Forecasted Time

IV. Supplementary Plots for Sweeps over Percent Responsive Generation

In a single-area system, the amount of responsive generation in the system (both for primary response and secondary response) were swept over a range of 10% to 30%, as

seen in Figure 4-32. Responsive generation power output and the ACE of the system are plotted in Figure C-17 and Figure C-18.

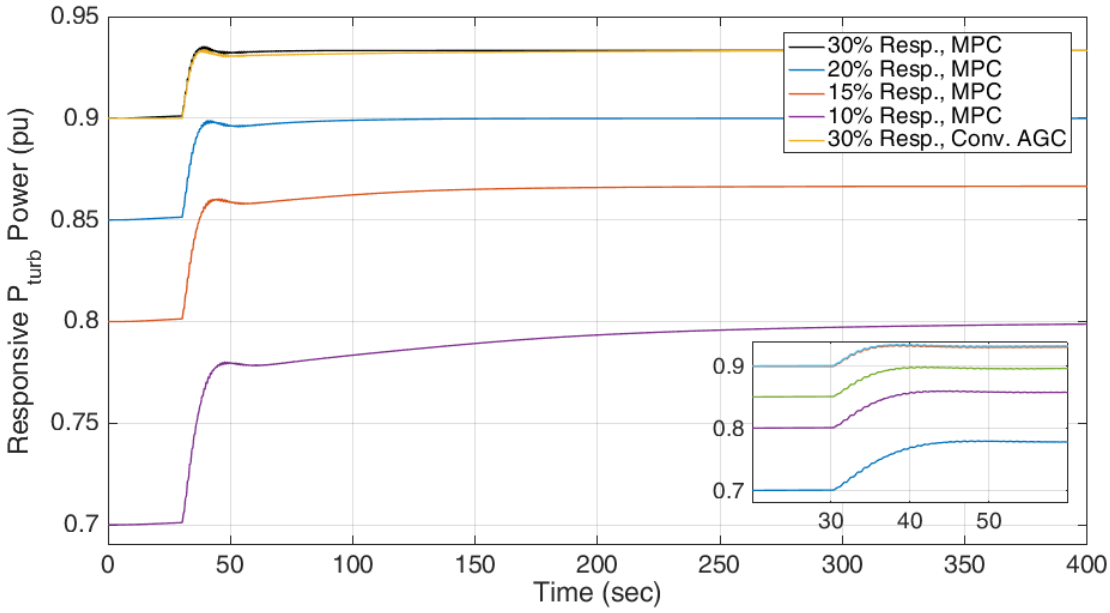


Figure C-17 Responsive Turbine Output of Responsive Generation Percentage Sweep

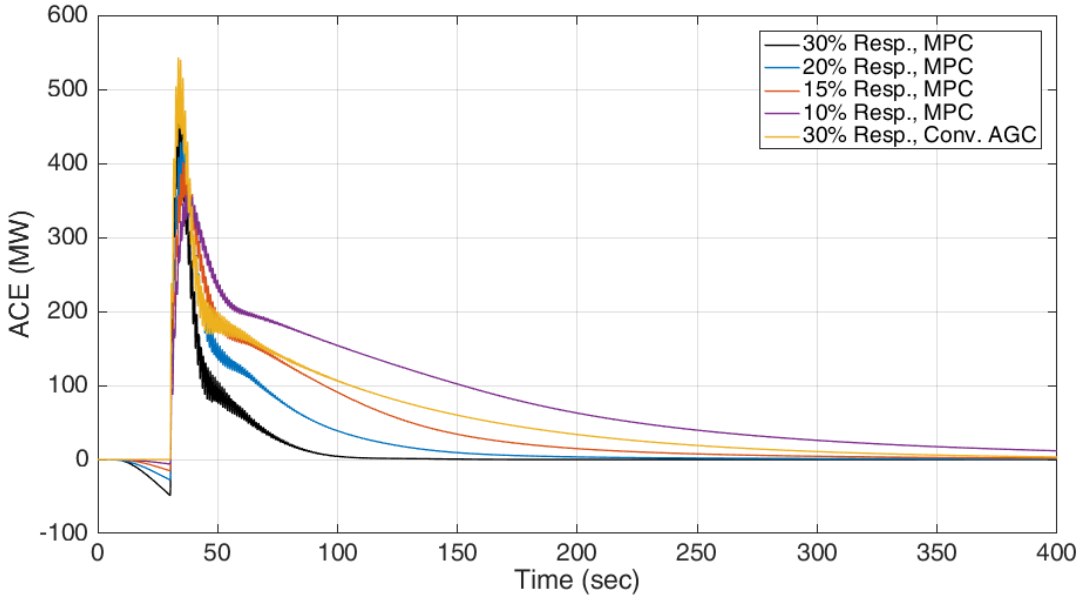


Figure C-18 ACE of Responsive Generation Percentage Sweep

V. Supplementary Plots for Multi-area Simulations

More results from the simulations in Section 4.4.3(i) are shown here. Plots of the responsive turbine power output and ACE for a multi-area system with a perfect forecast, with balanced weights in the objective function, are shown in Figure C-19, Figure C-20, and Figure C-21.

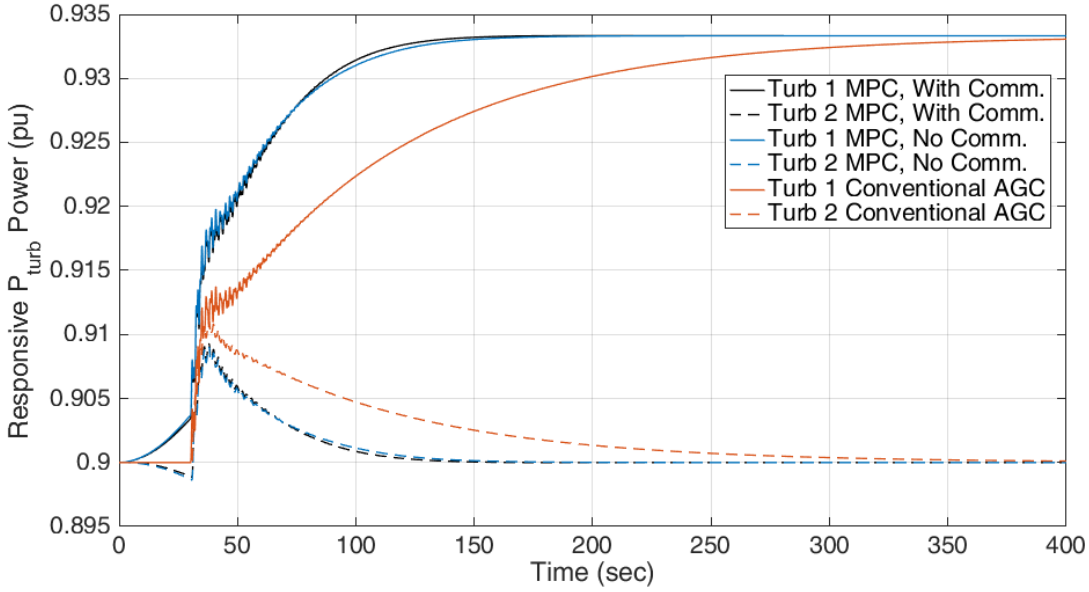


Figure C-19 Turbine Power Output of Multi-area MPC with Balanced Objective Function Weights

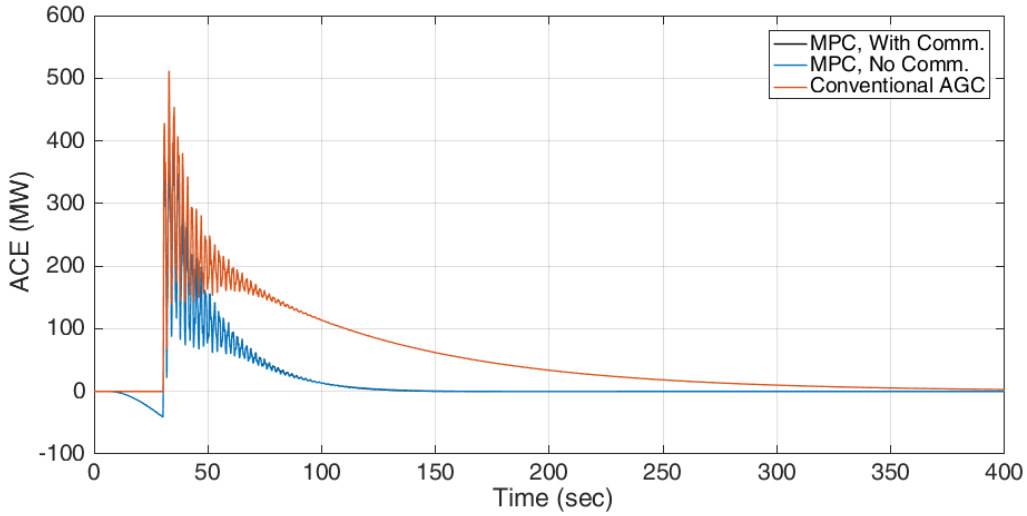


Figure C-20 ACE of Multi-area MPC with Balanced Objective Function Weights, Area 1

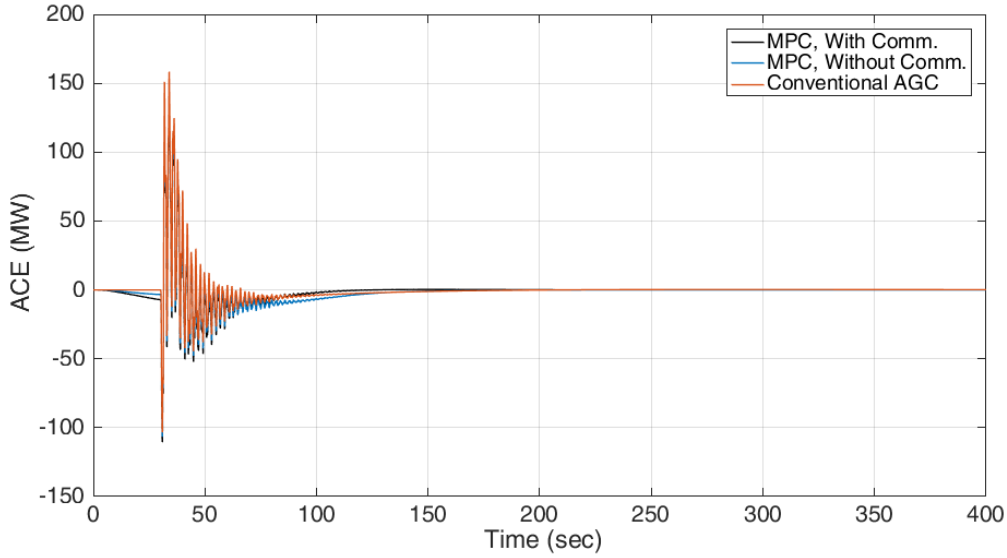


Figure C-21 ACE of Multi-area MPC with Balanced Objective Function Weights, Area 2

In addition, turbine output and ACE plots for a heavily weighted frequency deviation, with $Q_f = 1.5$, are shown in Figure C-22, Figure C-23, and Figure C-24.

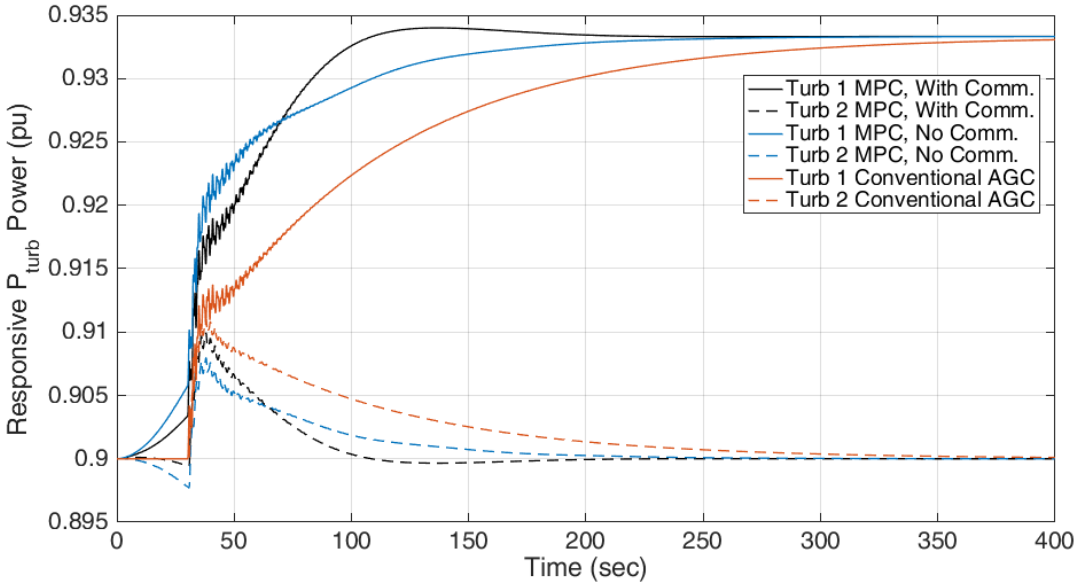


Figure C-22 Turbine Power of Multi-area MPC, with $Q_f = 1.5$

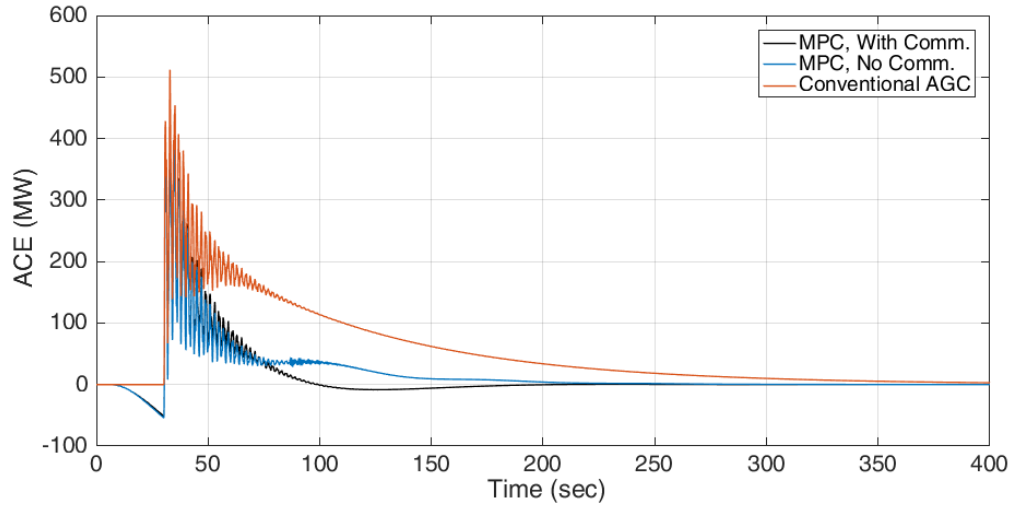


Figure C-23 ACE of Multi-area MPC, with $Q_f = 1.5$ in Area 1

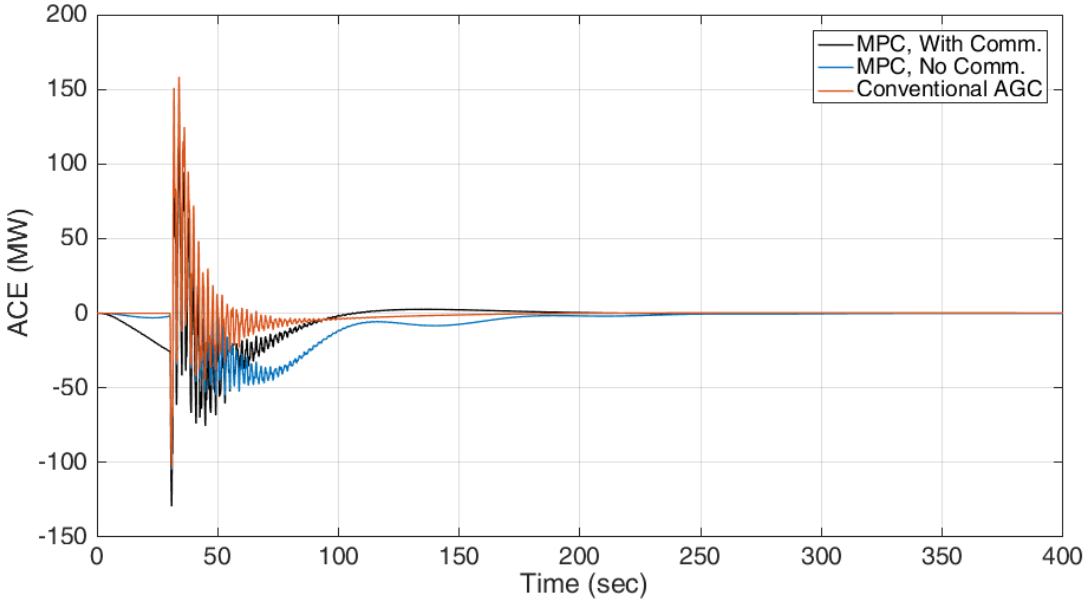


Figure C-24 ACE of Multi-area MPC, with $Q_f = 1.5$ in Area 2

Turbine output and ACE plots for a heavily weighted frequency deviation, with $Q_{ptie} = 7$, is shown in Figure C-25, Figure C-26, and Figure C-27.

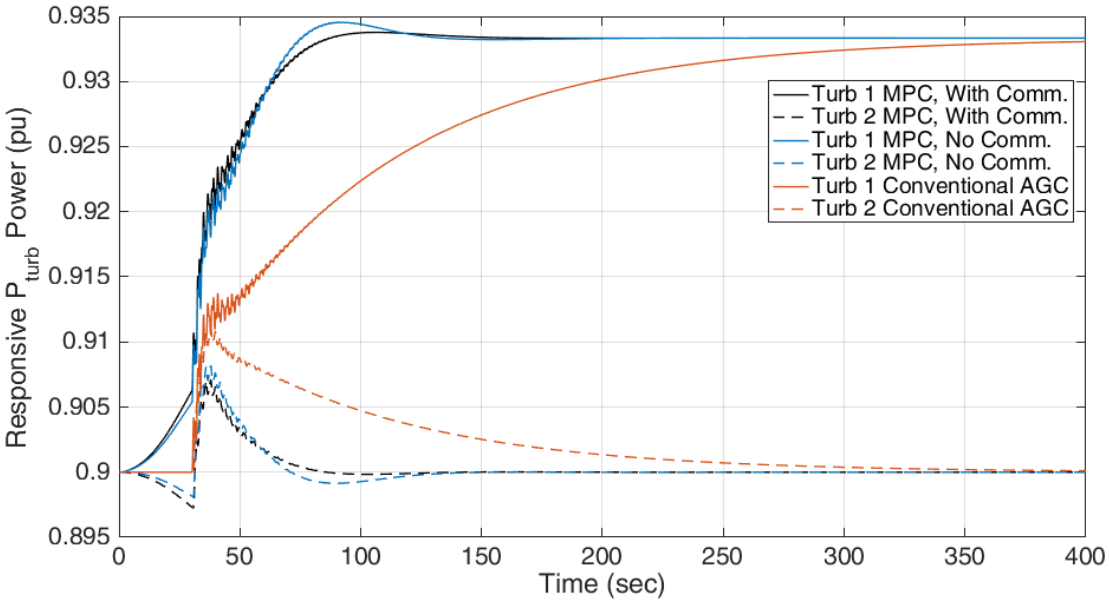


Figure C-25 Responsive Turbine Outputs of Multi-area MPC, with $Q_{ptie} = 7$

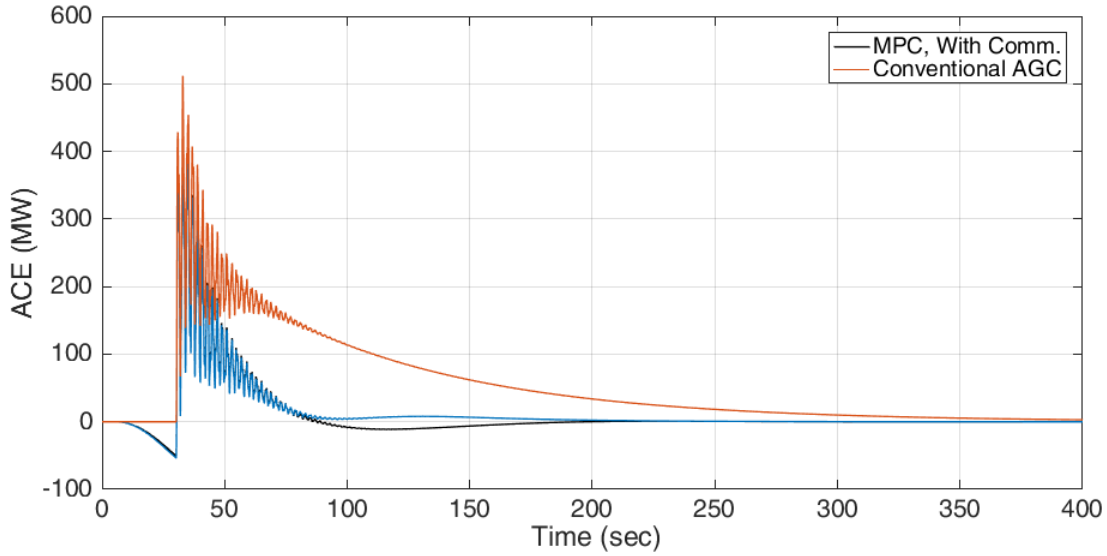


Figure C-26 ACE of Multi-area MPC, with $Q_{ptie} = 7$ in Area 1

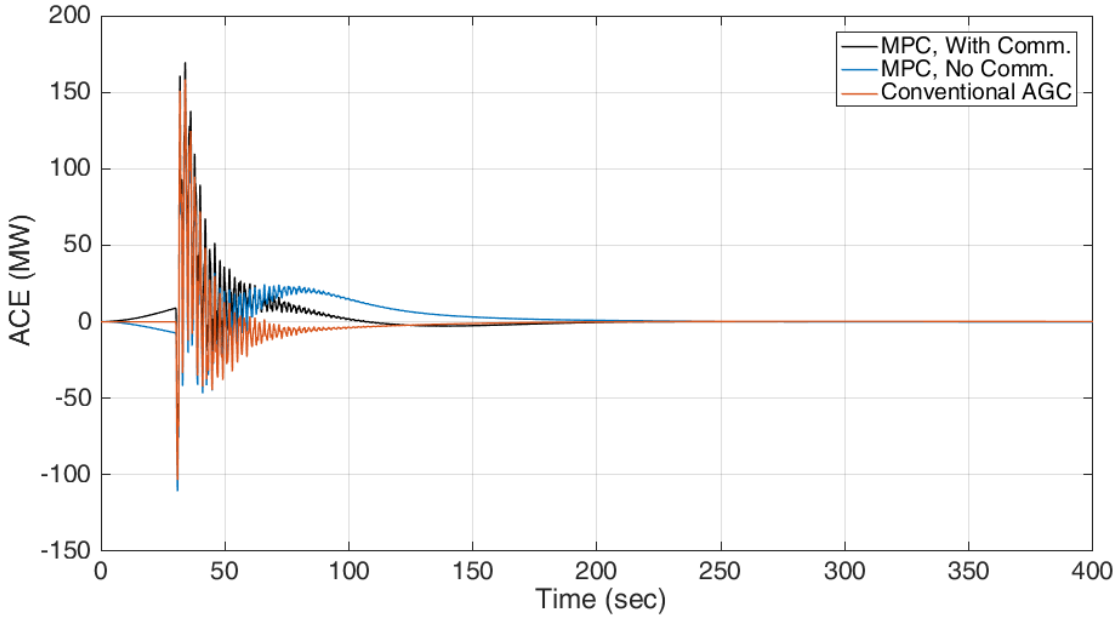


Figure C-27 ACE of Multi-area MPC, with $Q_{ptie} = 7$ in Area 2

VI. Supplementary Plots for 5-minute Ahead Forecast

This section presents some more results on the 5-minute ahead forecast simulations in Section 4.4.4. Conventional AGC turbine power output and ACE plots are shown in Figure C-28, Figure C-29, and Figure C-30, and comparisons between turbine outputs and ACE for different MPC paths are shown in Figure C-31, Figure C-32, and Figure C-33. Average frequency and tie-line deviation metrics are shown in Table C-1 and Table C-2.

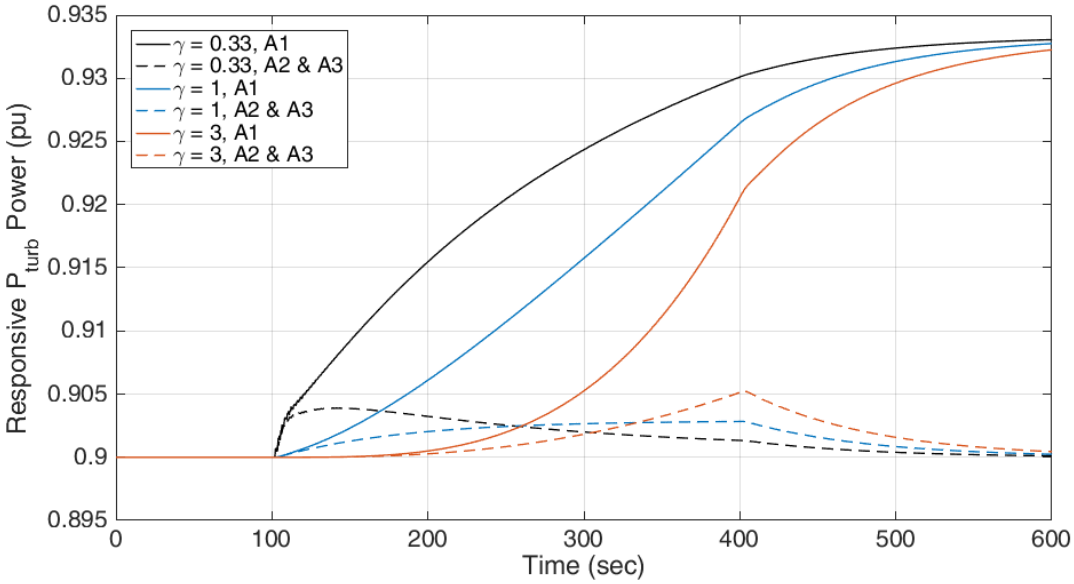


Figure C-28 Turbine Outputs of System under Conventional Control with Varied γ

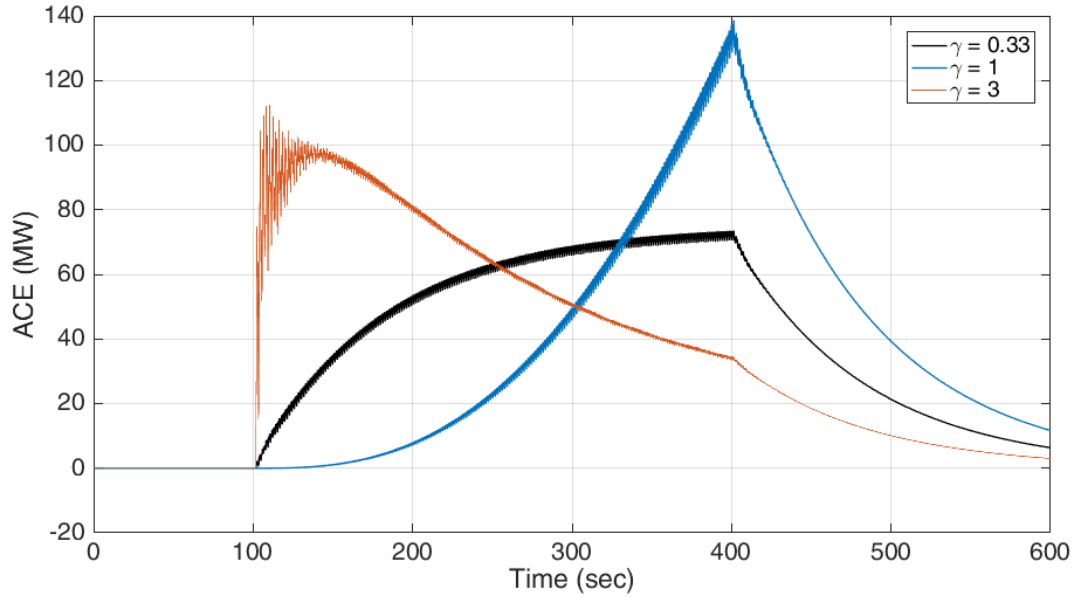


Figure C-29 ACE of System under Conventional Control with Varied γ for Area 1

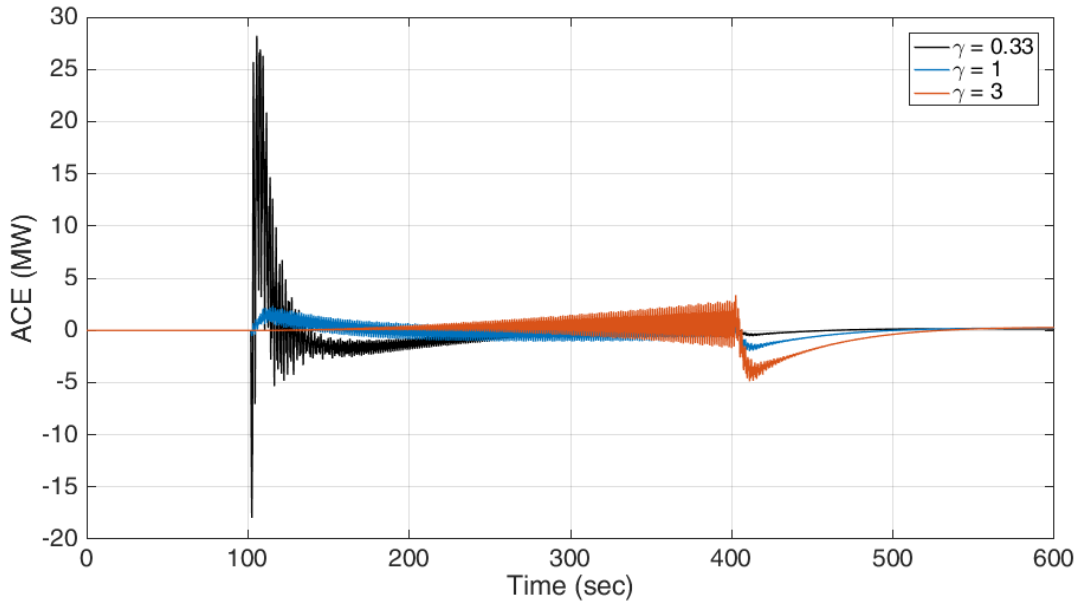


Figure C-30 ACE of System under Conventional Control with Varied γ for Area 2 and 3

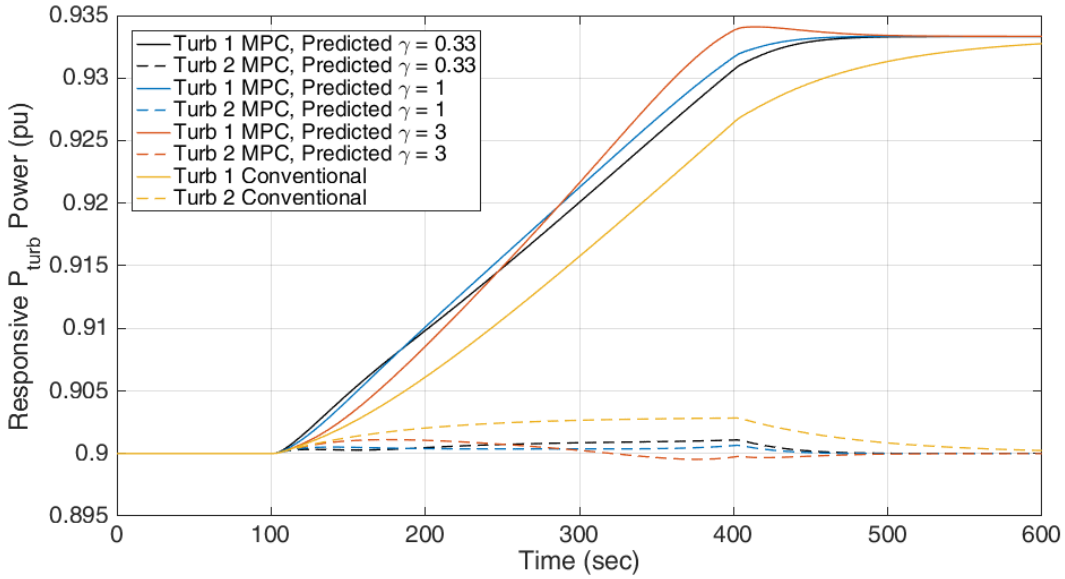


Figure C-31 Responsive Turbine Output under MPC Control with Varied $\hat{\gamma}$

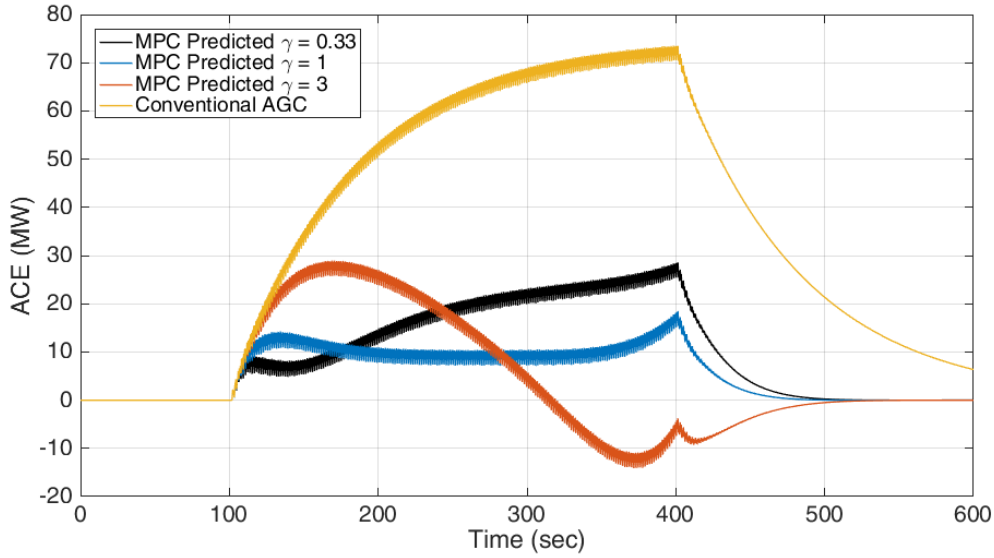


Figure C-32 ACE under MPC Control with Varied $\hat{\gamma}$ for Area 1

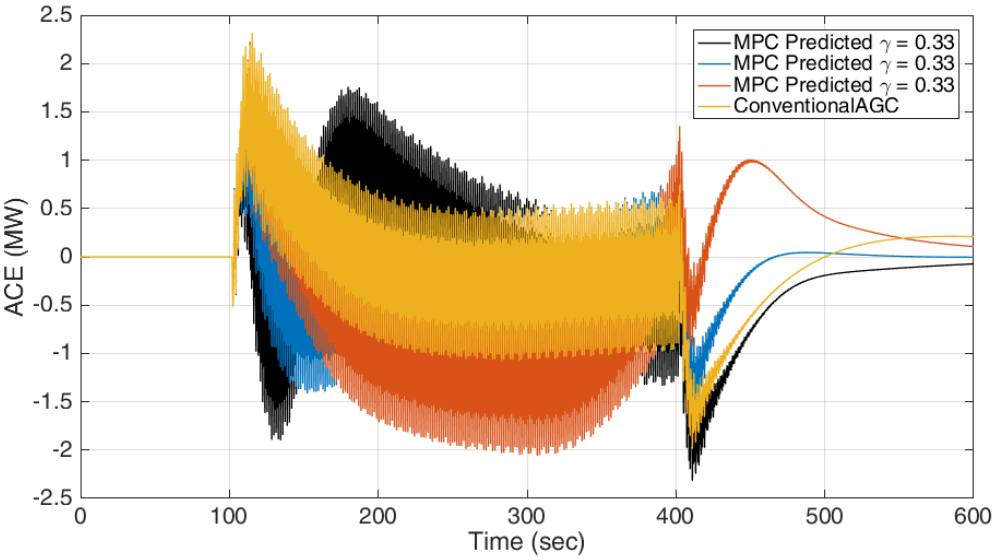


Figure C-33 ACE under MPC Control with Varied $\hat{\gamma}$ for Area 2 and 3

Table C-1 Average Frequency Deviation (mHz) for Varying Predicted vs. Actual Wind Ramps

		Actual γ		
		0.33	1	3
Anticipated $\hat{\gamma}$ (MPC)	0.33	0.878	0.882	0.878
	1	0.939	0.501	1.095
	3	1.582	0.754	0.415
Conventional		3.441	3.393	3.319
PI		2.135	2.110	2.114

Table C-2 Average Tie-line Deviation (MW) for Varying Predicted vs. Actual Wind Ramps

		Actual γ		
		0.33	1	3
Anticipated $\hat{\gamma}$ (MPC)	0.33	6.5	6.5	10.4
	1	7.1	3.9	8.1
	3	11.8	5.7	3.2
Conventional		24.4	24.1	23.7
PI		24.4	24.1	23.7

APPENDIX D

POWER SYSTEM STATE-SPACE MATRICES

State-space matrices for the multi-area power system with communication is shown below, in terms of the variables in Section 4.4.1. The matrices are divided into multiple sections to fit the entire matrix into the document.

A MATRIX:

	P_{tie}^{02}	P_{tie}^{01}	P_{tie}^{12}	$\Delta\omega^2$	V_p^2	P_m^2	P_{tot}^2	$\Delta\omega^1$
P_{tie}^{02}	0	0	0	-9.425	0	0	0	9.425
P_{tie}^{01}	0	0	0	0	0	0	0	9.425
P_{tie}^{12}	0	0	0	-9.425	0	0	0	0
$\Delta\omega^2$	1	0	1	-0.125	0.18	0.021	0	0
V_p^2	0	0	0	-3.125	-2	0	0.001	0
P_m^2	0	0	0	0	2	-0.1	0	0
P_{tot}^2	37.5	0	37.5	-39.844	0	0	0	0
$\Delta\omega^1$	-1	-1	0	0	0	0	0	-0.125
V_p^1	0	0	0	0	0	0	0	-3.125
P_m^1	0	0	0	0	0	0	0	0
P_{tot}^1	-37.5	-37.5	0	0	0	0	0	-39.844
$\Delta\omega_6^0$	0	1	-1	0	0	0	0	0
P_{tie}^m	0	-1	1	0	0	0	0	0
$V_{p,1}^0$	0	0	0	0	0	0	0	0
$P_{m,1}^0$	0	0	0	0	0	0	0	0
$V_{p,4}^0$	0	0	0	0	0	0	0	0
$P_{m,4}^0$	0	0	0	0	0	0	0	0
$V_{p,5}^0$	0	0	0	0	0	0	0	0
$P_{m,5}^0$	0	0	0	0	0	0	0	0
$V_{p,6}^0$	0	0	0	0	0	0	0	0
$P_{m,6}^0$	0	0	0	0	0	0	0	0
eACE	0	25000	-25000	0	0	0	0	0
$\Delta\omega_1^0$	0	0	0	0	0	0	0	0
$\Delta\omega_2^0$	0	0	0	0	0	0	0	0
$\Delta\omega_3^0$	0	0	0	0	0	0	0	0
$\Delta\omega_4^0$	0	0	0	0	0	0	0	0
$\Delta\omega_5^0$	0	0	0	0	0	0	0	0
P_{tot}^0	0	37.5	-37.5	0	0	0	0	0
P_{tie-13}^0	0	0	0	0	0	0	0	0
P_{tie-32}^0	0	0	0	0	0	0	0	0
P_{tie-35}^0	0	0	0	0	0	0	0	0
P_{tie-34}^0	0	0	0	0	0	0	0	0
P_{tie-45}^0	0	0	0	0	0	0	0	0

P_{tie-46}^{0}	0	0	0	0	0	0	0	0
P_{tie-56}^{0}	0	0	0	0	0	0	0	0
ε_{d}	0	0	0	0	0	0	0	0
ε_{ptie}	0	0	0	0	0	0	0	0

	V_p^1	P_m^1	P_{tot}^1	$\Delta\omega_6^0$	P_{tie}^m	$V_{p,1}^0$	$P_{m,1}^0$	$V_{p,4}^0$
P_{tie}^{02}	0	0	0	0	0	0	0	0
P_{tie}^{01}	0	0	0	-56.55	0	0	0	0
P_{tie}^{12}	0	0	0	56.55	0	0	0	0
$\Delta\omega^2$	0	0	0	0	0	0	0	0
V_p^2	0	0	0	0	0	0	0	0
P_m^2	0	0	0	0	0	0	0	0
P_{tot}^2	0	0	0	0	0	0	0	0
$\Delta\omega^1$	0.18	0.021	0	0	0	0	0	0
V_p^1	-2	0	0.001	0	0	0	0	0
P_m^1	2	-0.1	0	0	0	0	0	0
P_{tot}^1	0	0	0	0	0	0	0	0
$\Delta\omega_6^0$	0	0	0	-0.125	0	0	0	0
P_{tie}^m	0	0	0	0	-0.62832	0	0	0
$V_{p,1}^0$	0	0	0	0	0	-2	0	0
$P_{m,1}^0$	0	0	0	0	0	2	-0.1	0
$V_{p,4}^0$	0	0	0	0	0	0	0	-2
$P_{m,4}^0$	0	0	0	0	0	0	0	2
$V_{p,5}^0$	0	0	0	0	0	0	0	0
$P_{m,5}^0$	0	0	0	0	0	0	0	0
$V_{p,6}^0$	0	0	0	-18.75	0	0	0	0
$P_{m,6}^0$	0	0	0	0	0	0	0	0
eACE	0	0	0	-1.59E+05	0	0	0	0
$\Delta\omega_1^0$	0	0	0	0	0	0.0216	0.00252	0
$\Delta\omega_2^0$	0	0	0	0	0	0	0	0
$\Delta\omega_3^0$	0	0	0	0	0	0	0	0
$\Delta\omega_4^0$	0	0	0	0	0	0	0	0.036
$\Delta\omega_5^0$	0	0	0	0	0	0	0	0
P_{tot}^0	0	0	0	-239.06	0	0	0	0
P_{tie-13}^0	0	0	0	0	0	0	0	0
P_{tie-32}^0	0	0	0	0	0	0	0	0
P_{tie-35}^0	0	0	0	0	0	0	0	0
P_{tie-34}^0	0	0	0	0	0	0	0	0
P_{tie-45}^0	0	0	0	0	0	0	0	0
P_{tie-46}^0	0	0	0	-56.55	0	0	0	0

P^0_{tie-56}	0	0	0	-56.55	0	0	0	0
ε_d	0	0	0	0	0	0	0	0
ε_{ptie}	0	0	0	0	0	0	0	0

	$P^0_{m,4}$	$V^0_{p,5}$	$P^0_{m,5}$	$V^0_{p,6}$	$P^0_{m,6}$	eACE	$\Delta\omega^0_1$	$\Delta\omega^0_2$
P^{02}_{tie}	0	0	0	0	0	0	0	0
P^{01}_{tie}	0	0	0	0	0	0	0	0
P^{12}_{tie}	0	0	0	0	0	0	0	0
$\Delta\omega^2$	0	0	0	0	0	0	0	0
V^2_p	0	0	0	0	0	0	0	0
P^2_m	0	0	0	0	0	0	0	0
P^2_{tot}	0	0	0	0	0	0	0	0
$\Delta\omega^1$	0	0	0	0	0	0	0	0
V^1_p	0	0	0	0	0	0	0	0
P^1_m	0	0	0	0	0	0	0	0
P^1_{tot}	0	0	0	0	0	0	0	0
$\Delta\omega^0_6$	0	0	0	0.0696	0.00812	0	0	0
P^m_{tie}	0	0	0	0	0	0	0	0
$V^0_{p,1}$	0	0	0	0	0	0	-18.75	0
$P^0_{m,1}$	0	0	0	0	0	0	0	0
$V^0_{p,4}$	0	0	0	0	0	0	0	0
$P^0_{m,4}$	-0.1	0	0	0	0	0	0	0
$V^0_{p,5}$	0	-2	0	0	0	0	0	0
$P^0_{m,5}$	0	2	-0.1	0	0	0	0	0
$V^0_{p,6}$	0	0	0	-2	0	0	0	0
$P^0_{m,6}$	0	0	0	2	-0.1	0	0	0
eACE	0	0	0	0	0	-0.62832	0	0
$\Delta\omega^0_1$	0	0	0	0	0	0	-0.125	0
$\Delta\omega^0_2$	0	0	0	0	0	0	0	-0.125
$\Delta\omega^0_3$	0	0	0	0	0	0	0	0
$\Delta\omega^0_4$	0.0042	0	0	0	0	0	0	0
$\Delta\omega^0_5$	0	0.0528	0.00616	0	0	0	0	0
P^0_{tot}	0	0	0	0	0	0	0	0
P^0_{tie-13}	0	0	0	0	0	0	56.55	0
P^0_{tie-32}	0	0	0	0	0	0	0	-56.55
P^0_{tie-35}	0	0	0	0	0	0	0	0
P^0_{tie-34}	0	0	0	0	0	0	0	0
P^0_{tie-45}	0	0	0	0	0	0	0	0
P^0_{tie-46}	0	0	0	0	0	0	0	0
P^0_{tie-56}	0	0	0	0	0	0	0	0

ε_d	0	0	0	0	0	0	0	0
ε_{ptie}	0	0	0	0	0	0	0	0

	$\Delta\omega_3^0$	$\Delta\omega_4^0$	$\Delta\omega_5^0$	P_{tot}^0	P_{tie-13}^0	P_{tie-32}^0	P_{tie-35}^0	P_{tie-34}^0
P_{tie}^{02}	0	0	0	0	0	0	0	0
P_{tie}^{01}	0	0	0	0	0	0	0	0
P_{tie}^{12}	0	0	0	0	0	0	0	0
$\Delta\omega^2$	0	0	0	0	0	0	0	0
V_p^2	0	0	0	0	0	0	0	0
P_m^2	0	0	0	0	0	0	0	0
P_{tot}^2	0	0	0	0	0	0	0	0
$\Delta\omega^1$	0	0	0	0	0	0	0	0
V_p^1	0	0	0	0	0	0	0	0
P_m^1	0	0	0	0	0	0	0	0
P_{tot}^1	0	0	0	0	0	0	0	0
$\Delta\omega_6^0$	0	0	0	0	0	0	0	0
P_{tie}^m	0	0	0	0	0	0	0	0
$V_{p,1}^0$	0	0	0	0.001	0	0	0	0
$P_{m,1}^0$	0	0	0	0	0	0	0	0
$V_{p,4}^0$	0	-18.75	0	0.001	0	0	0	0
$P_{m,4}^0$	0	0	0	0	0	0	0	0
$V_{p,5}^0$	0	0	-18.75	0.001	0	0	0	0
$P_{m,5}^0$	0	0	0	0	0	0	0	0
$V_{p,6}^0$	0	0	0	0.001	0	0	0	0
$P_{m,6}^0$	0	0	0	0	0	0	0	0
eACE	0	0	0	0	0	0	0	0
$\Delta\omega_1^0$	0	0	0	0	-1	0	0	0
$\Delta\omega_2^0$	0	0	0	0	0	1	0	0
$\Delta\omega_3^0$	-0.125	0	0	0	1	-1	-1	-1
$\Delta\omega_4^0$	0	-0.125	0	0	0	0	0	1
$\Delta\omega_5^0$	0	0	-0.125	0	0	0	1	0
P_{tot}^0	0	0	0	0	0	0	0	0
P_{tie-13}^0	-56.55	0	0	0	0	0	0	0
P_{tie-32}^0	56.55	0	0	0	0	0	0	0
P_{tie-35}^0	56.55	0	-56.55	0	0	0	0	0
P_{tie-34}^0	56.55	-56.55	0	0	0	0	0	0
P_{tie-45}^0	0	56.55	-56.55	0	0	0	0	0
P_{tie-46}^0	0	56.55	0	0	0	0	0	0
P_{tie-56}^0	0	0	56.55	0	0	0	0	0
ε_d	0	0	0	0	0	0	0	0

ε_{ptie}	0	0	0	0	0	0	0	0

	P^0_{tie-45}	P^0_{tie-46}	P^0_{tie-56}	ε_d	ε_{ptie}
P^{02}_{tie}	0	0	0	0	0
P^{01}_{tie}	0	0	0	0	0
P^{12}_{tie}	0	0	0	0	0
$\Delta\omega^2$	0	0	0	0	0
V_p^2	0	0	0	0	0
P_m^2	0	0	0	0	0
P^2_{tot}	0	0	0	0	0
$\Delta\omega^1$	0	0	0	0	0
V_p^1	0	0	0	0	0
P_m^1	0	0	0	0	0
P^1_{tot}	0	0	0	0	0
$\Delta\omega^0_6$	0	1	1	0	-1
P^m_{tie}	0	0	0	0	1
$V^0_{p,1}$	0	0	0	0	0
$P^0_{m,1}$	0	0	0	0	0
$V^0_{p,4}$	0	0	0	0	0
$P^0_{m,4}$	0	0	0	0	0
$V^0_{p,5}$	0	0	0	0	0
$P^0_{m,5}$	0	0	0	0	0
$V^0_{p,6}$	0	0	0	0	0
$P^0_{m,6}$	0	0	0	0	0
eACE	0	0	0	0	-25000
$\Delta\omega^0_1$	0	0	0	0	0
$\Delta\omega^0_2$	0	0	0	1	0
$\Delta\omega^0_3$	0	0	0	0	0
$\Delta\omega^0_4$	-1	-1	0	0	0
$\Delta\omega^0_5$	1	0	-1	0	0
P^0_{tot}	0	0	0	0	-37.5
P^0_{tie-13}	0	0	0	0	0
P^0_{tie-32}	0	0	0	0	0
P^0_{tie-35}	0	0	0	0	0
P^0_{tie-34}	0	0	0	0	0
P^0_{tie-45}	0	0	0	0	0
P^0_{tie-46}	0	0	0	0	0
P^0_{tie-56}	0	0	0	0	0
ε_d	0	0	0	1	0
ε_{ptie}	0	0	0	0	1

B MATRIX:

	f_{set}	w_{ptie}	$\hat{d}$	w_d	$\hat{d}^2$	$\hat{d}^1$	f_{set}^2
P_{tie}^{02}	0	0	0	0	0	0	0
P_{tie}^{01}	0	0	0	0	0	0	0
P_{tie}^{12}	0	0	0	0	0	0	0
$\Delta\omega^2$	0	0	0	0	0	1	0
V_p^2	0	0	0	0	0	0	0
P_m^2	0	0	0	0	0	0	0
P_{tot}^2	0	0	0	0	0	0	0
$\Delta\omega^1$	0	0	0	0	1	0	0
V_p^1	0	0	0	0	0	0	0
P_m^1	0	0	0	0	0	0	0
P_{tot}^1	0	0	0	0	0	0	5.3125
$\Delta\omega_6^0$	0	-1	0	0	0	0	0
P_{tie}^m	0	1	0	0	0	0	0
$V_{p,1}^0$	0	0	0	0	0	0	0
$P_{m,1}^0$	0	0	0	0	0	0	0
$V_{p,4}^0$	0	0	0	0	0	0	0
$P_{m,4}^0$	0	0	0	0	0	0	0
$V_{p,5}^0$	0	0	0	0	0	0	0
$P_{m,5}^0$	0	0	0	0	0	0	0
$V_{p,6}^0$	0	0	0	0	0	0	0
$P_{m,6}^0$	0	0	0	0	0	0	0
eACE	3541.7	-25000	0	0	0	0	0
$\Delta\omega_1^0$	0	0	0	0	0	0	0
$\Delta\omega_2^0$	0	0	1	1	0	0	0
$\Delta\omega_3^0$	0	0	0	0	0	0	0
$\Delta\omega_4^0$	0	0	0	0	0	0	0
$\Delta\omega_5^0$	0	0	0	0	0	0	0
P_{tot}^0	5.3125	-37.5	0	0	0	0	0
P_{tie-13}^0	0	0	0	0	0	0	0
P_{tie-32}^0	0	0	0	0	0	0	0
P_{tie-35}^0	0	0	0	0	0	0	0
P_{tie-34}^0	0	0	0	0	0	0	0
P_{tie-45}^0	0	0	0	0	0	0	0
P_{tie-46}^0	0	0	0	0	0	0	0
P_{tie-56}^0	0	0	0	0	0	0	0
ε_d	0	0	0	0	0	0	0
ε_{ptie}	0	0	0	0	0	0	0

	f_{set}^{1}
P_{tie}^{02}	0
P_{tie}^{01}	0
P_{tie}^{12}	0
$\Delta\omega^{2}$	0
V_{p}^{2}	0
P_{m}^{2}	0
P_{tot}^{2}	5.3125
$\Delta\omega^{1}$	0
V_{p}^{1}	0
P_{m}^{1}	0
P_{tot}^{1}	0
$\Delta\omega_{6}^{0}$	0
P_{tie}^{m}	0
$V_{p,1}^{0}$	0
$P_{m,1}^{0}$	0
$V_{p,4}^{0}$	0
$P_{m,4}^{0}$	0
$V_{p,5}^{0}$	0
$P_{m,5}^{0}$	0
$V_{p,6}^{0}$	0
$P_{m,6}^{0}$	0
eACE	0
$\Delta\omega_{1}^{0}$	0
$\Delta\omega_{2}^{0}$	0
$\Delta\omega_{3}^{0}$	0
$\Delta\omega_{4}^{0}$	0
$\Delta\omega_{5}^{0}$	0
P_{tot}^{0}	0
P_{tie-13}^{0}	0
P_{tie-32}^{0}	0
P_{tie-35}^{0}	0
P_{tie-34}^{0}	0
P_{tie-45}^{0}	0
P_{tie-46}^{0}	0
P_{tie-56}^{0}	0
ε_{d}	0
ε_{ptie}	0

C MATRIX (transposed, for clarity):

	Δf_{sys}	P_{turb}^{1}	P_{turb}^{2}	P_{turb}^{4}	P_{turb}^{5}	P_{turb}^{6}	Δp_{tie}	eACE
P_{tie}^{02}	0	0	0	0	0	0	0	0
P_{tie}^{01}	0	0	0	0	0	0	0	0
P_{tie}^{12}	0	0	0	0	0	0	0	0
$\Delta\omega^{2}$	0	0	0	0	0	0	0	0
V_{p}^{2}	0	0	0	0	0	0	0	0
P_{m}^{2}	0	0	0	0	0	0	0	0
P_{tot}^{2}	0	0	0	0	0	0	0	0
$\Delta\omega^{1}$	0	0	0	0	0	0	0	0
V_{p}^{1}	0	0	0	0	0	0	0	0
P_{m}^{1}	0	0	0	0	0	0	0	0
P_{tot}^{1}	0	0	0	0	0	0	0	0
$\Delta\omega_{6}^{0}$	45	0	0	0	0	0	0	0
P_{tie}^{m}	0	0	0	0	0	0	0.62832	0
$V_{p,1}^{0}$	0	0.0216	0	0	0	0	0	0
$P_{m,1}^{0}$	0	0.00252	0	0	0	0	0	0
$V_{p,4}^{0}$	0	0	0	0.036	0	0	0	0
$P_{m,4}^{0}$	0	0	0	0.0042	0	0	0	0
$V_{p,5}^{0}$	0	0	0	0	0.0528	0	0	0
$P_{m,5}^{0}$	0	0	0	0	0.00616	0	0	0
$V_{p,6}^{0}$	0	0	0	0	0	0.0696	0	0
$P_{m,6}^{0}$	0	0	0	0	0	0.00812	0	0
eACE	0	0	0	0	0	0	0	0.62832
$\Delta\omega_{1}^{0}$	0	0	0	0	0	0	0	0
$\Delta\omega_{2}^{0}$	0	0	0	0	0	0	0	0
$\Delta\omega_{3}^{0}$	0	0	0	0	0	0	0	0
$\Delta\omega_{4}^{0}$	0	0	0	0	0	0	0	0
$\Delta\omega_{5}^{0}$	0	0	0	0	0	0	0	0
P_{tot}^{0}	0	0	0	0	0	0	0	0
P_{tie-13}^{0}	0	0	0	0	0	0	0	0
P_{tie-32}^{0}	0	0	0	0	0	0	0	0
P_{tie-35}^{0}	0	0	0	0	0	0	0	0
P_{tie-34}^{0}	0	0	0	0	0	0	0	0
P_{tie-45}^{0}	0	0	0	0	0	0	0	0
P_{tie-46}^{0}	0	0	0	0	0	0	0	0
P_{tie-56}^{0}	0	0	0	0	0	0	0	0
ε_{d}	0	0	0	0	0	0	0	0
ε_{ptie}	0	0	0	0	0	0	0	0

D MATRIX:

	f_{set}	w_{ptie}	$\hat{d}$	w_d	$\hat{d}^2$	$\hat{d}^1$	f_{set}^2	f_{set}
Δf_{sys}	0	0	0	0	0	0	0	0
P_{turb}^1	0	0	0	0	0	0	0	0
P_{turb}^2	0	0	1	1	0	0	0	0
P_{turb}^4	0	0	0	0	0	0	0	0
P_{turb}^5	0	0	0	0	0	0	0	0
P_{turb}^6	0	0	0	0	0	0	0	0
Δp_{tie}	0	0	0	0	0	0	0	0
eACE	0	0	0	0	0	0	0	0