PNNL-23680

U.S. DEPARTMENT OF
ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Stochastic Operations and Planning

YV Makarov
PV Etingov
Z Huang
LE Miller
Y Sun
B Vyakaranam
D Wu
R Diao
Z Hou
D Meng
N Samaan
M Vallem
S Wang
Y Zhang

April 2015

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.
(8/2010)

PNNL-23680

Stochastic Operations and Planning

YV Makarov
PV Etingov
Z Huang
LE Miller
Y Sun
B Vyakaranam
D Wu
R Diao
Z Hou
D Meng
N Samaan
M Vallem
S Wang
Y Zhang

April 2015

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Abstract

The increasing impacts of randomness and variability on modern power systems are changing and will dramatically continue to change patterns of system behavior, how systems are planned, how systems are dispatched, and how energy is exchanged. The existing deterministic approaches primarily used by utilities are based on deterministically established dispatch and flow patterns, a few "typical" stresses, and known congested paths. Consequently, these approaches are becoming increasingly inadequate for dealing with the problem of uncertainty. A new generation of probabilistic methods, reliability and control performance criteria, tools, and business practices is necessary to address these challenges.

This project is funded by the Office of Electricity Delivery and Energy Reliability (OE) at the U.S. Department of Energy (DOE). It directly addresses these needs and the DOE OE goal "…to develop the next-generation of system planning and operations tools and controls to visualize system dynamics, identify areas of reliability concern, prevent and mitigate cascading events, and support post-event analyses".

The key objectives of this project are as follows:

- Develop a concept and methodologies for real-time and offline power system analytics, representing a power system as a stochastic machine.
- Develop and test new uncertainty quantification methods and tools.
- Support the transition from deterministic to probabilistic methods in planning and operations.

The main results of this first phase project include the following:

1. A zonal geographically distributed model was developed for the Western Electricity Coordinating Council (WECC) system reflecting various sources of uncertainty in their interaction (load, wind and solar generation). The model forms a basis for developing and testing probabilistic methods.
2. Our approach based on separation of slow and fast power system motions (represented as net load variations) and associated uncertainties using probabilistic methods helps to improve predictability and reduce uncertainty in the system. In this project we attempted to predict the forecast error based on state-of-the-art prediction methods. The ability to partially predict forecast errors would help to reduce the forecast error itself as well as ultimately reduce the prediction interval around the errors. Such a task is very challenging. Our results show that this, in fact, is feasible. We successful demonstrate that that is possible to reduce the forecast error if the signal has autocorrelation or cross-correlation with other signals.
3. The statistical separation of fast and slow motions helps to better distribute the balancing job between slow and fast balancing resources; so that they are more efficiently used and the cost is reduced (the fast-responsive resources are usually more expensive). In this project, advanced statistical methods were applied to separate slow, fast and very fast system motions caused by variation of the net system load.

4. We developed and demonstrated new methods for quantification and adjustment of transmission limits using probabilistic methods (in the WECC system). We demonstrated how the transmission limits in the WECC system can be quantified and adjusted to minimize the risk of violations caused by uncertainty. This adjustment can also help to increase the utilization of transmission facilities within capacity and reliability limits.

5. A new probabilistic methodology and new analytical software called the Transmission Uncertainty and Prediction Tool (TUT) was developed at Pacific Northwest National Laboratory using funding from DOE OE, DOE Office of Energy Efficiency and Renewable Energy (EERE), and the California Energy Commission (CEC). The developed methodology predicts the impact of the uncertainties on congested paths, worst-case voltage drops, and reactive power margins for several hours ahead of time and, if needed, proposes control actions to mitigate the problem if necessary.

6. We demonstrated how TUT can benefit from parallelization.

The following significant impacts are expected to be achieved by this phase of the project:

- Risk-based adaptive constraints for system dispatch, unit commitment and energy imbalance markets will result in a better utilization of transmission within reliability requirements.

- Predictive/preventive control based on probability and risk will reduce the need for last-minute operator actions such as load shed or disconnection of wind generation.

- Improved uncertainty quantification and forecasting will reduce the balancing reserve needs and associated cost.

- The advanced prototype level Transmission Uncertainty and Prediction Tool (TUT) will predict the impact of uncertainties on the transmission system.

Summary

This project is funded by the Office of Electricity Delivery and Energy Reliability (OE) at the U.S. Department of Energy (DOE).

Motivation of this project

The increasing impacts of randomness and variability on modern power systems are changing and will dramatically continue to change patterns of system behavior, how systems are planned, how systems are dispatched, and how energy is exchanged. The existing deterministic approaches primarily used by utilities are based on deterministically established dispatch and flow patterns, a few "typical" stresses, and known congested paths. Consequently, these approaches are becoming increasingly inadequate for dealing with the problem of uncertainty. A new generation of probabilistic methods, reliability and control performance criteria, tools, and business practices is necessary to address these challenges.

Benefits of probabilistic methods

1. Adapting system planning and operations methods to increasingly stochastic system behavior.
2. Improving reliability by addressing uncertainty and variability.
3. Better and robust decision making for a wider range of possible system conditions.
4. Optimizing capital investments in system development.
5. More economical system operation due to (1) uncertainty quantification and proactive minimization; and (2) optimizing the "mainstream" system performance within the foreseeable range of uncertainty and variability, e.g., covering 99% of uncertainty.
6. Optimizing reserves to address system uncertainties and variability within the main uncertainty range.
7. Developing the ability to scientifically plan and prepare the system for extreme events outside of the mainstream operational range.

What will happen without the transition from deterministic to probabilistic methods?

1. Increasing risk of system failures, blackouts and near-misses.
2. More congestion problems in real time; over-conservative operational limits; underutilization of transmission assets.
3. Excessive or insufficient operating reserves.
4. Less economical system operation while addressing unexpected situations.

What are the barriers to transitioning from deterministic to probabilistic methods?

There is a widespread understanding of the need for such transition. But there are several significant challenges:

- Lack of primary statistical information accumulated by the industry.
- Lack of the methods and means for storing, processing, consolidation and validation of this information.

- Lack of the models reflecting all sources of uncertainty.
- Absence of trustworthy probabilistic reliability criteria and business practices.
- Insufficient familiarity and experience with state-of-the-art probability theory and methods among engineers.
- Insufficient range and number of production-grade probabilistic tools.
- Investments and effort needed for the transition.

Proposed overall plan

The purpose of this effort is to attract interest from industry organizations and vendors to make a subsequent implementation project feasible by laying the foundation of methodologies for incorporating uncertainties into power system planning and operation. Ultimately, this project could lead to an industry-wide effort (a major nation-wide initiative) to transform grid planning and operation from the existing deterministic platform to the new probabilistic platform. Over the course of future phases, we envision the following major steps:

- Develop a comprehensive framework and roadmap for probabilistic planning and operations based on state-of-the art criteria, methodologies, software tools, and technologies.
- Cooperate closely with the industry, regulators, universities, government, national laboratories, software vendors, and other interested organizations to make sure that the best ideas, know-how and skills are reflected in the framework and the roadmap.
- Initiate and lead a nation-wide effort for implementing this framework within the next 5 years as a standard US practice in planning and operations.
- Facilitate continuing education, dissemination, and technology transfer in the area of probabilistic methods and applications.
- Create and lead a Probabilistic Technology Interest Group (PTIG) as a tool to implement the Roadmap and forum for organization interested in the area. The Group could be based on (Utility Variable-Generation Integration Group) UVIG (membership fees) or North American Synchrophasor Initiative (NASPI) models (funded by DOE).

Specific objectives of Phase 1 of the project

- Develop a concept and methodologies for real-time and offline power system analytics, representing a power system as a stochastic machine.
- Develop and test new uncertainty quantification methods and tools with the following capabilities:
    - Multi-source, multi-variant geographically distributed uncertainty model.
    - Characterize prevailing system motions by a limited number of principal components.
    - Quantify probabilistic limits along principal component analysis (PCA) coordinates.
    - Explore opportunities for statistical separation of fast and slow system motions and their prediction.
    - High performance computing (HPC) applications.
    - Transmission uncertainty and prediction tool.
    - Contribute to the transition from deterministic to probabilistic methods in planning and operations.

Phase I results

1. A zonal geographically distributed model was developed for the Western Electricity Coordinating Council (WECC) system reflecting various sources of uncertainty in their interaction (load, wind and solar generation). The model forms a basis for developing and testing probabilistic methods.

2. Our approach based on separation of slow and fast power system motions (represented as net load variations = load variations – wind variations – solar variations) and associated uncertainties using probabilistic methods helps to improve predictability and reduce uncertainty in the system. Slower motions of stochastic processes are more predictable due to more autocorrelation between the subsequent points. By employing this autocorrelation, it is possible to improve the forecast of parameters subject to uncertainty. By quantifying uncertainty around the improved forecast, it is possible to reduce the range of uncertainty.

3. Connection of uncertainty with system motions. A power system maintains balance between its generation, changing load, variable generation, and interchange. Increasing uncertainty means more balancing work to be done by conventional generation, energy storage, and controllable load. In other words, these resources have to move more with increasing uncertainty and variability. Because the balancing service is a paid service, more system motion results in additional costs. The statistical separation of fast and slow motions helps to better distribute the balancing job between slow and fast balancing resources; so that they are more efficiently used and the cost is reduced (the fast-responsive resources are usually more expensive).

4. In this project, advanced statistical methods were applied to separate slow, fast, and very fast system motions caused by variation of the net system load. The very fast motion is actually white noise, which is a completely unpredictable component of a signal. If the forecast error is exactly white noise, this is a perfect forecast. The advanced statistical methods applied were:

    a. Wavelet transforms – autoregressive integrated moving average (ARIMA) method. Wavelet transforms were used to extract the slower component in the total load forecast error. The ARIMA model was used to predict this slower component. We demonstrated an uncertainty reduction of 10-12% by applying this combination of methods.

    b. We demonstrated, that the proposed decision tree regression analysis method helps to reduce the uncertainty (and associated balancing effort) up to 40%, and that the residual (very fast motion component) becomes close to white noise.

    c. Principal component analysis (PCA). By applying this method, we employ cross-correlation between the total load forecast errors in different zones of the WECC system. This is a new idea. The PCA helps to reduce the number of dimensions, where the predictable forecast components of multiple zonal errors can be located. This helps to improve the accuracy of the forecasts, and make the residual error close to white noise.

5. We developed and demonstrated new methods for quantification and adjustment of transmission limits using probabilistic methods (in the WECC system). Using the Transmission Uncertainty Prediction Tool (TUT), we demonstrated how the transmission limits in the WECC system can be quantified and adjusted to minimize the risk of violations caused by uncertainty. This adjustment

can also help to increase the utilization of transmission facilities within capacity and reliability limits.

6. PNNL Transmission Uncertainty and Prediction Tool (TUT). Interactions among wind, solar gen load forecast errors can have a random impact on transmission system. This event can lead to additional congestion and voltage stability problems. To mitigate this problem, Balancing Authorities (BAs) should be able to re-dispatch their conventional generators ahead of time, if such impact can be predicted. A new probabilistic methodology and new analytical software called the Transmission Uncertainty and Prediction Tool (TUT) was developed at Pacific Northwest National Laboratory using funding from DOE Office of Energy (OE), DOE Office of Energy Efficiency and Renewable Energy (EERE), and the California Energy Commission (CEC). The developed methodology predicts the impact of the uncertainties on congested paths, worst-case voltage drops, and reactive power margins for several hours ahead of time and, if needed, proposes control actions to mitigate the problem if necessary.

7. We demonstrated how TUT can benefit from parallelization.

Importance and significance of Phase 1 results

The following significant impacts are expected to be achieved by this project:

- Achieve better utilization of the transmission system capacity under reliability requirements through risk-based adaptive constraints for system dispatch, unit commitment, and energy imbalance markets.

- Reduce and eliminate the dependence on last-minute operator actions such as load shed or disconnection of wind generation though predictive/preventive control based on probability and risk.

- Reduce balancing reserve needs and the associated costs through improved uncertainty quantification and forecasting methods.
- The developed methodology and TUT tool have several very important advantages and opportunities for the balancing authorities (BAs) and other transmission system operators (TSOs):
    - Better quantification of available security margins.
    - Better reliability level.
    - Better utilization of transmission assets.
    - Better situation awareness and predictive system monitoring.
    - Preventive control.

Proposed Fiscal Year 2015 Tasks

- Move the Transmission Uncertainty and Prediction Tool to a near production level of development. The Tool will be transformed into a "fully" standalone application.

- Stochastic optimization and control. The project will reformulate key optimization and control problems to reflect the random nature of 21st-century power systems operations.
- Tail events quantification and assessment. The tail events are observed whenever the system state exceeds the limits suitable for the “mainstream” power system operations. The tail events are not very well studied or addressed in the modern system planning and operation. The proposed task will be one of the first efforts to quantify these events, predict them, and develop uncertainty-based controls to avoid them.

Acknowledgments

This project is funded by the Office of Electricity Delivery and Energy Reliability (OE) at the U.S. Department of Energy (DOE).

The project team wants to specially thank Mr. Gil Bindewald (DOE OE) for his continuing support, help and guidance.

The project team appreciates contributions of the following PNNL engineers and managers:

- Dr. Mark Morgan, Project Manager
- Dr. Feng Pan, PNNL Peer Reviewer
- Dr. Dale King, Product Line Manager
- Dr. Carl Imhoff, Electric Infrastructure Market Sector Manager
- Ms. Tamica Dickenson, Project Coordinator
- Ms. Sue Arey, Project Specialist

Acronyms and Abbreviations

AC	Alternating current
AGC	Automatic generation control
AIC	Akaike information criteria
ARIMA	Autoregressive integrated moving average
BA	Balancing authority
BIC	Bayesian information criteria
BR	Balancing requirement
CAISO	California Independent System Operator
CEC	California Energy Commission
COI	California-Oregon Intertie
CPS	Control performance standards
CSP	Concentrated solar power
DC	Direct current
DOE	Department of Energy
EERE	DOE Office of Energy Efficiency and Renewable Energy
EMS	Energy Management System
GIS	Geographical information systems
GUI	Graphical user interface
HILF	High-impact low-frequency
HPC	High performance computing
i.i.d.	Independent identically distributed random variables
IPC	Idaho Power Company
ISO	Independent system operator

KPSS	Kwiatkowski–Phillips–Schmidt–Shin test
MAE	Mean absolute error
MC	Monte Carlo
MGDM	Multivariate geographically distributed model
OE	DOE Office of Electricity Delivery and Energy Reliability
NASPI	North American Synchrophasor Initiative
NREL	National Renewable Energy Laboratory
PACE	PacifiCorp East
PCA	Principal component analysis
PDF	Probability density function
PG&E	Pacific Gas and Electric Company
PNNL	Pacific Northwest National Laboratory
PTDF	Power transfer function distribution factor
PTIG	Probabilistic Technology Interest Group
PV	Photovoltaic
QQ plot	A probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.
RJDBC	R package for Java Database Connectivity
RPS	Renewable portfolio standard
SCE	Southern California Edison
SDGE	San Diego Gas and Electric
SQL	Structured Query Language
SVD	Singular Value Decomposition
TEPPC	Transmission Expansion Planning Policy Committee
TSO	Transmission system operators

TUT	Transmission Uncertainty and Prediction Tool
UVIG	Utility Variable-Generation Integration Group
VGS	Variable Generation Subcommittee
WECC	Western Electricity Coordinating Council
WT	Wavelet transform

Contents

Abstract iii

Summary v

  Motivation of this project v

  Benefits of probabilistic methods v

  What will happen without the transition from deterministic to probabilistic methods? v

  What are the barriers to transitioning from deterministic to probabilistic methods? v

  Proposed overall plan vi

  Specific objectives of Phase 1 of the project vi

  Phase I results vii

  Importance and significance of Phase 1 results viii

  Proposed Fiscal Year 2015 Tasks viii

Acknowledgments xi

Acronyms and Abbreviations xiii

1.0 Introduction 1.1

  1.1. Motivation of This Project 1.1

  1.2 Current Work in the Area of Probabilistic Methods in Power 1.3

    1.2.1 WECC Planning Reference Book Review Summary 1.3

    1.2.2 Gaps Identified in the Existing Research and Applications 1.4

  1.3 The Content of This Phase of Work 1.5

  1.4 References 1.6

2.0 Project Objectives 2.1

3.0 Sources of Uncertainty and the Importance of Their Concurrent Consideration 3.1

  3.1 System Power Balance and Procedures to Achieve It 3.2

  3.2 Sources of Uncertainties and Their Characteristics 3.4

    3.2.1 Load Forecast Errors 3.4

    3.2.2 Wind Power Forecast Errors 3.5

    3.2.3 Solar Generation Forecast Errors 3.5

    3.2.4 Forced Generation Outages 3.8

    3.2.5 Uninstructed Deviation Errors 3.9

  3.3 Statistical Characteristics of Forecast Errors and Balancing Requirement 3.10

    3.3.1 Discretization Errors 3.10

    3.3.2 Non-parametric nature 3.10

    3.3.3 Autocorrelation of Wind and Load Forecast Errors 3.10

    3.3.4 Cross-correlation 3.11

    3.3.5 Non-stationary Nature 3.12

  3.4 References 3.13

4.0 Western Interconnection Uncertainty Model Development 4.1
4.1 Primary Dataset Description 4.1
4.1.1 WECC BA Structure 4.1
4.1.2 Load Data 4.3
4.1.3 Wind and Solar Data 4.6
4.2 References 4.8
5.0 Reduction of Uncertainty in Forecast Errors for Load and Wind Generation Using Integrated ARIMA Forecasting, Signal Decomposition, and Principal Component Analysis 5.1
5.1 Connection of Uncertainty with System Motions 5.1
5.2 Generation of Forecast Errors Using Seasonal ARIMA Models 5.1
5.2.1 Data Preprocessing 5.1
5.2.2 Fitting Seasonal ARIMA Models 5.2
5.3 Improved Representation and Reconstruction of Load and Wind Forecast Errors Using Signal Decomposition and Automated ARIMA 5.3
5.3.1 Signal Decomposition of Time Series 5.3
5.3.2 Forecast Error Reconstruction 5.6
5.4 Evaluating the Residuals/Errors Using Principal Components to Represent/Reconstruct System Behaviors 5.6
5.5 Simulation for Short-Term Forecast Errors 5.8
5.5.1 Sequential Gaussian Simulation 5.9
5.5.2 Error Reduction Using PCA 5.9
5.5.3 Tests on Statistical Properties 5.11
5.6 References 5.16
6.0 Improving System Predictability Using Statistical Separation of Slow, Fast and Very Fast System Motions 6.1
6.1 Methodology description 6.2
6.1.1 A Regression Tree Based Predictive Model 6.2
6.1.2 An Advanced Method for Analyzing Fast and Slow Motions 6.3
6.2 Case Studies 6.3
6.2.1 Load Forecast 6.3
6.2.2 Forecast Errors Prediction 6.5
6.3 Wavelet Analysis 6.8
6.4 Conclusion 6.12
6.5 References 6.12
7.0 Principal Stress Direction Extraction and Motion Trajectory Approximation 7.1
7.3 Conclusions 7.8
7.4 References 7.8
8.0 Probabilistic Quantification and Observation of WECC System Transmission Limits 8.1
8.1 Introduction 8.1
8.2 Methodology Description 8.1

    - 8.2.1 Mathematical formulation 8.3
  - 8.3 Demonstration 8.3
  - 8.4 Conclusion 8.10
- 9.0 Transmission Uncertainty and Prediction Tool (TUT) 9.1
- 10.0 Parallel Implementation of Transmission Uncertainty and Prediction Tool 10.1
  - 10.1 Computational Cost Analysis of Algorithms in TUT 10.1
    - 10.1.1 PTDF Calculation 10.1
    - 10.1.2 Interface Flow PDF Calculation 10.2
  - 10.2 Parallel Implementation 10.3
  - 10.3 Test Results 10.5
  - 10.4 Future Work 10.6
- 11.0 Conclusions and Future Work 1
  - 11.1 Conclusions 1
  - 11.2 Recommendations for Future Work 2
    - 11.2.1 Proposed Fiscal Year 2015 Work 2
    - 11.2.2 Long-term Recommendations – Big Picture 3
  - 11.3 Publications 3

Figures

Figure 1. Transition from deterministic to probabilistic technologies...1.2
Figure 2. Relationship between the scheduling, load following and regulation processes.............3.3
Figure 3. Overall uncertainty within a dispatch interval..3.3
Figure 4. Example of load forecast error in a real system..3.4
Figure 5. Impact of temperature forecast error on load 3.4
Figure 6. Distribution of the standard deviation of solar forecast errors depending on the clearness index...3.8
Figure 7. Distribution of solar forecast error in a very cloudy day and a very sunny day.............3.8
Figure 8. Balancing requirement caused by forced generation outage. 3.9
Figure 9. Hourly regulation requirement for a real BA 3.10
Figure 10. Autocorrelation of original and simulated load forecast errors 3.11
Figure 11. Auto-correlation of original and simulated wind forecast errors..............................3.11
Figure 12. Geographical locations of the wind farms..3.12
Figure 13. One-year hourly original and simulation results for load and wind forecast errors.....3.13
Figure 14. Current balancing authorities in WECC..4.2
Figure 15. TEPPC topology diagram for 2020 base case..4.3
Figure 16. Actual load, hourly average and interpolated load for 2009.......................................4.4
Figure 17. Error between interpolated load curve and actual load curve (normalized by peak load)..4.4
Figure 18. Imposing the load variance in 2009 to the interpolated load in 2020. ....4.5
Figure 19. Simulated load data for 2020. ....4.5
Figure 20. Peak load of different Balancing Authorities in the “WECC TEPPC 2020” case.........4.6
Figure 21. Installed wind capacities in 2020 for WECC BAs. 4.7
Figure 22. Installed solar capacities in 2020 for WECC BAs. 4.7
Figure 23.Time-of-day bias for the load forecast error data..5.2
Figure 24. Wind forecast error for bus 10116 (10 minute resolution) during February 2020. ....5.4
Figure 25. Trend component of wind forecast error for bus 10116 (10 minute resolution) during February 2020..5.4
Figure 26. Seasonal component of wind forecast error for bus 10116 (10 minute resolution) during February 2020. ....5.5
Figure 27. Random component of wind forecast error for bus 10116 (10 minute resolution) during February 2020..5.5
Figure 28. Generated random component from the fitted AR(3) model for wind forecast error at bus 10116 (10 minute resolution) during February 2020. ....5.6
Figure 29. Percentage of total variance in original data that is explained by each principal component. 5.7
Figure 30. Accumulated percentage of total variance in the original data that is explained by the principal components...5.7

Figure 31. Percentage of total variance in incremental data that is explained by each principal component. 5.8
Figure 32. Accumulated percentage of total variance in the incremental data that is explained by the principal components. 5.8
Figure 33. Randomly selected 10 sets of simulated load forecast errors for zone 1. The two curves are different, but their statistical characteristics are identical (mean, variance, autocorrelation pattern)..5.9
Figure 34. Percentage of total variance explained of actual load forecast errors within 4 hours prior to the given time instant (12/10/2020 10:00). 5.10
Figure 35. Percentage of total variance explained of actual wind forecast errors within 4 hours prior to the given time instant (12/10/2020 10:00). 5.10
Figure 36. Randomly selected 10 sets of simulated load forecast errors for zone 1 using PCA....5.11
Figure 37. Normality test on the 1st set of simulation data for all zones....................................5.12
Figure 38. Stationarity test on the 1st set of simulation data for all zones..................................5.12
Figure 39. Stationarity test on the residuals of the 1st set of simulation data for all zones. 5.13
Figure 40. Normality test on the 1st set of simulation data for all zones using the PCA approach.5.13
Figure 41. Normality test on the residuals of the 1st set of simulation data for all zones using the PCA approach...5.14
Figure 42. Stationarity test on the 1st set of simulation data for all zones using the PCA approach5.14
Figure 43. Correlation matrices for simulated load forecast errors, without and with using PCA 5.15
Figure 44. Example for a regression tree..6.2
Figure 45. Day-ahead load forecast..6.4
Figure 46. Hour-ahead load forecast...6.4
Figure 47. Load forecasted forecast error..6.6
Figure 48. Net load original forecast error and preducted predicted forecast error........................6.7
Figure 49. Error distribution and absolute error distribution (MAE: mean absolute error)............6.7
Figure 50. Wavelet analyses of net load forecast errors:...6.8
Figure 51. Wavelet analyses of predictions of new load forecast errors..6.9
Figure 52. Wavelet analyses of residuals with net load forecast error prediction.........................6.10
Figure 53. QQ plot and PDF of the residuals..6.10
Figure 54. Wavelet analyses of residuals with load forecast error predictions............................6.11
Figure 55. Plot of v_8 (representing increments in the interval between 8 and 9 a.m.) for different days in January in the parameter space of 39 zones..7.2
Figure 56. Singular values for different components...7.4
Figure 57. The first three principal components. 7.4
Figure 58. Approximation using first three principal components...7.5
Figure 59. Approximation using the first two principal components. 7.6
Figure 60. Principal directions for 24 intervals of January. 7.7
Figure 61. Principal directions for 24 intervals of July...7.8
Figure 62. Power flow probability density function for the COI interface of WECC...................8.2
Figure 63. Probability density function for the COI interface with generation re-dispatch. 8.2

Figure 64: Improvement in congestion in the ALLSTON-KEELER interface. 8.7
Figure 65: Improvement in congestion in the CA INDEPENDENT - MEXICO (CFE) interface. .8.7
Figure 66: Improvement in congestion in the IDAHO - SIERRA interface. 8.8
Figure 67: Improvement in congestion in the INYO - CONTROL 115 KV TIE interface............8.9
Figure 68: Improvement in congestion in the MP-SL interface..8.9
Figure 69: Improvement in congestion in the PATH 26 interface. 8.10
Figure 70. An illustration of transmission congestion caused by load, wind and solar forecast errors..9.1
Figure 71. A concept of probabilistic transmission uncertainty analysis.....................................9.2
Figure 72. Probabilistic analysis of transmission flows in WECC system 9.3
Figure 73. Flow chart of PTDF calculation calling PowerWorld®...10.2
Figure 74. Flow chart of PDF calculation: PFD of interface flow in each contingency 10.3
Figure 75. Flow chart of PDF calculation: Combine PDF over all contingencies 10.3
Figure 76. Parallel computation of PTDF through SimAuto®, with distributed computing add-on10.4
Figure 77. Parallel computation of PTDF through SimAuto®, without distributed computing add-on 10.4
Figure 78. Control items to parallel PTDF computation..10.5
Figure 79. Performance of parallel PTDF computation 10.6

Tables

Table 1. Cross-correlation of wind forecast errors....3.12
Table 2. 32 Balancing Authorities in WECC....4.6
Table 3. Name of zones....5.15
Table 4. Interface base power flows and limits of flow changes....8.4
Table 5. Calculated generation changes to eliminate probabilistic interface congestions....8.6
Table 6. Computation time of PTDF and Monte Carlo....10.5

1.0 Introduction

The increasing impacts of randomness and variability on modern power systems are changing and will dramatically continue to change patterns of system behavior, how systems are planned, how systems are dispatched, and how energy is exchanged. The existing deterministic approaches primarily used by utilities are based on deterministically established dispatch and flow patterns, a few “typical” stresses, and known congested paths. Consequently, these approaches are becoming increasingly inadequate for dealing with the problem of uncertainty. A new generation of probabilistic methods, reliability and control performance criteria, tools, and business practices is necessary to address these challenges.

1.1. Motivation of This Project

There is a significant need to develop key elements of a new grid methodology based on probabilistic system models and performance/reliability criteria. These methodology elements will address the fundamental new ways of power system modeling, prediction, stressing, analysis, and control.

Figure 1 shows one conceptual view of the transition from deterministic to probabilistic technologies used in power system operations and planning. The root problem in this development is building uncertainty and variability models for all contributing factors. In this figure, some additional sources of uncertainty are shown, such market impacts, loop flows, micro grids and demand-side controls, and, of course contingencies. In many instances, this effort requires well-organized, coordinated, systematic, and continuous collection and processing of primary information, such as transmission lines and generator outage rates. The next step is building uncertainty models for each source, as well as the overall large-scale uncertainty model (statistical system model), which also becomes a geographically distributed model when the transmission aspect of uncertainty is analyzed. Such models can reproduce statistical characteristics of the primary historical information and serve as scenario generators for numerous Monte Carlo simulations, which are frequently needed for probabilistic analyses. The high-performance computational framework becomes essentially important due to the enormous computational challenge of probabilistic methods. Branches of the “tree” represent some other tasks that can be implemented using the new probabilistic framework. Together, with the addition of many other tasks, they will produce a new paradigm for power system planning and operation and decision support tools for grid operators. New approaches will be developed, for instance, statistical linearization or probabilistic definition of system stresses. It is important to notice that the global uncertainty quantification is closely linked with better prediction of system parameters and behavior.

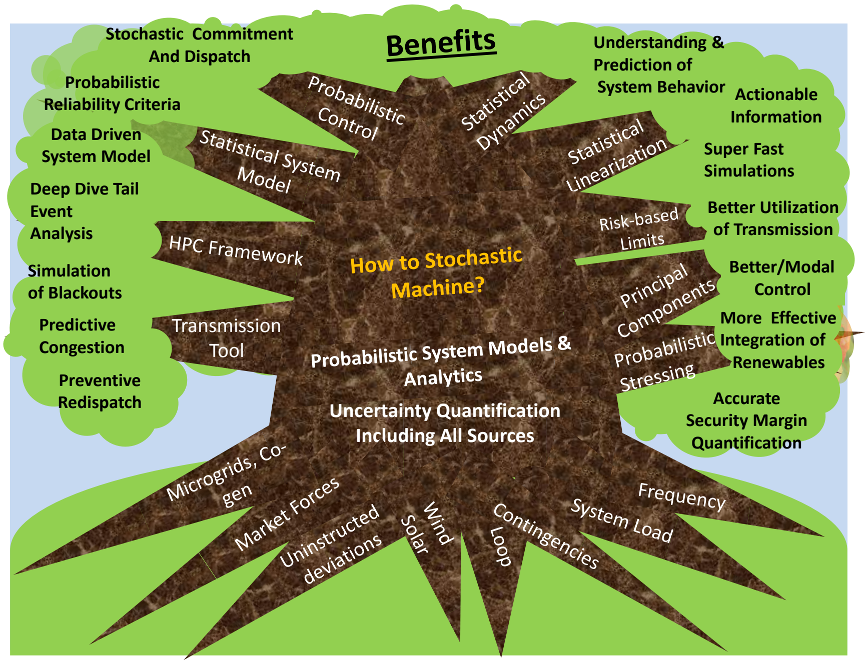

Benefits
Stochastic Commitment And Dispatch
Probabilistic Reliability Criteria
Data Driven System Model
Deep Dive Tail Event Analysis
Simulation of Blackouts
Predictive Congestion
Preventive Redispatch
Understanding & Prediction of System Behavior
Actionable Information
Super Fast Simulations
Better Utilization of Transmission
Better/Modal Control
More Effective Integration of Renewables
Accurate Security Margin Quantification
Probabilistic Control
Statistical Dynamics
Statistical System Model
Statistical Linearization
HPC Framework
Risk-based Limits
Transmission Tool
Principal Components
Probabilistic Stressing
How to Stochastic Machine?
Probabilistic System Models & Analytics
Uncertainty Quantification Including All Sources
Microgrids, Co-gen
Market Forces
Uninstructed deviations
Wind
Solar
Contingencies
Loop
System Load
Frequency

Figure 1. Transition from deterministic to probabilistic technologies.

1.2 Current Work in the Area of Probabilistic Methods in Power

The area of probabilistic methods applied to power systems attracts a very significant interest from researchers, and there is a very extensive literature devoted to this subject. Some of the most developed topics are as follows:

General methods and approaches:

- Probabilistic power flow and contingency analysis
- Probabilistic optimization
- Probabilistic stability analyses
- Advanced Monte Carlo methods for multidimensional discrete/continuous models
- High Performance Computing (HPC) applications

Planning:

- Probabilistic composite reliability assessment
- Grid expansion studies
- Capacity planning
- Renewables integration
- Capacity credit

Operations:

- Probabilistic forecasting
- Probabilistic operational limits and risk-based security analysis
- Probabilistic state estimation
- Probabilistic assessment of balancing reserves requirement
- Probabilistic unit commitment and dispatch

A partial review of probabilistic methods in power was conducted by WECC Variable Generation Subcommittee in its report "WECC VGS Planning Reference Book" [1, 2]. This project was funded by DOE Office of Electricity Delivery and Energy Reliability and led by PNNL. The next section is an extraction of findings provided in [1, 2]. Those references also contain a bibliography.

1.2.1 WECC Planning Reference Book Review Summary

Report [1, 2] outlines the needs, opportunities, existing approaches, and some of the research in the area of probabilistic planning. The need for probabilistic approaches is dictated by the increasing variability and uncertainty in the modern power grids. The variability component appears whenever an assumption is made that the system parameters are constant, whereas they are changing over time, so that the difference between the assumed values and the actual values becomes evident. Uncertainty reflects our inability to accurately predict the future. Actual values slightly deviate from forecasted values. Variable generation is not the only source of variability and uncertainty in planning. Generation expansion uncertainty, uncertainty of market-driven dispatches, load uncertainty, random generator and transmission system outages, and other random factors contribute significantly to overall variability and uncertainty. It is important to build an integrated uncertainty model to correctly reflect all the impacts in their interaction. Working collectively, the sources of uncertainty lead to an increasing gap between the existing deterministic planning methods and the realities of the observed and expected system probabilistic behavior. To be able to address these challenges, the need in development and accepting probabilistic planning methods becomes evident.

A frequently used method of solving this problem is Monte Carlo simulation. This technique involves repeated simulation with random variables considered. Planning models using the Monte Carlo methodology have become increasingly useful when planning for integrating variable generation.

Well-being analysis combines the deterministic perception with probability concepts. This new framework reduces the gap between deterministic (e.g., the loss of the largest available unit) and probabilistic approaches. Well-being analysis has been applied in the last decade to areas such as generating systems, operating reserve assessment, and composite generation and transmission systems.

The necessities of probabilistic planning approaches require development of probabilistic planning criteria. It is virtually impossible to compare the statistically distributed results of applying probabilistic methods with the existing deterministic criteria. For example, a power flow distribution in transmission lines, obtained by running a probabilistic load flow algorithm, can be reduced to simply check for an overload condition in N–1 contingencies. Such an approach would result in an over-conservative assessment of available transfer capability on the grid, and ultimately to underuse of the available transmission facilities. To eliminate this possibility, the simple fact of appearance of a system problem in a limited number of planning scenarios could be replaced by the probability, risk (probability times the impact of the violation), severity, expected duration, frequency of occurrence, and other similar probability-based criteria. The WECC report [1, 2] contains a list of some possible probabilistic criteria, as well as relevant approaches for their utilization.

High-impact low-frequency (HILF) events have the potential to cause catastrophic impacts on the electric power system, but either rarely occur or never occur. Examples of HILF risks include coordinated cyber, physical, and blended attacks, the high-altitude detonation of a nuclear weapon, and major natural disasters like earthquakes, tsunamis, large hurricanes, pandemics, and geomagnetic disturbances caused by severe weather. It is not possible to consider all HILF events and weight their probability. A scenario-based analysis can be recommended for screening such events. A successful risk management plan will identify the threats and measure the protection decisions against the costs associated with proposed mitigation.

1.2.2 Gaps Identified in the Existing Research and Applications

The work in this project is focused on the existing gaps in the development of probabilistic methods and their application in the industry. Our objective is to contribute to addressing some of these gaps and helping to facilitate much-needed collective effort among academia, research organizations, vendors and industry organizations.

The following gaps have been identified:

- Lack of primary statistical information accumulated by industry.
- Lack of the methods and means for storing, processing, consolidation and validation of this information.

- Lack of the models reflecting all sources of uncertainty.
- Absence of trustworthy probabilistic reliability criteria and business practices.
- Insufficient familiarity and experience with state-of-the-art probability theory and methods among engineers.
- Insufficient range and number of production-grade probabilistic tools.
- Investments and effort needed for the transition.

1.3 The Content of This Phase of Work

- A zonal geographically distributed model was developed for the Western Electricity Coordinating Council (WECC) system reflecting various sources of uncertainty in their interaction (load, wind and solar generation). The model forms a basis for developing and testing probabilistic methods.

- Our approach based on separation of slow and fast power system motions (represented as net load variations) and associated uncertainties using probabilistic methods helps to improve predictability and reduce uncertainty in the system. Slower motions of stochastic processes are more predictable due to more autocorrelation between the subsequent points. By employing this autocorrelation, it is possible to improve the forecast of parameters subject to uncertainty. By quantifying uncertainty around the improved forecast, it is possible to reduce the range of uncertainty.

- Connection of uncertainty with system motions. A power system maintains balance between its generation, changing load, variable generation, and interchange. Increasing uncertainty means more balancing work to be done by conventional generation, energy storage, and controllable load. In other words, these resources have to move more with increasing uncertainty and variability. Because the balancing service is a paid service, more system motion results in additional costs. The statistical separation of fast and slow motions helps to better distribute the balancing job between slow and fast balancing resources; so that they are more efficiently used and the cost is reduced (the fast-responsive resources are usually more expensive).

- In this project, advanced statistical methods were applied to separate slow, fast and very fast system motions caused by variation of the net system load. The very fast motion is actually white noise, which is a completely unpredictable component of a signal. If the forecast error is exactly white noise, this is a perfect forecast. The advanced statistical methods applied were:

    a. Wavelet transforms – ARIMA method. Wavelet transforms were used to extract the slower component in the total load forecast error. The ARIMA model was used to predict this slower component. We demonstrated an uncertainty reduction of 10-12% by applying this combination of methods.

    b. We demonstrated, that the proposed decision tree regression analysis method helps to reduce the uncertainty (and associated balancing effort) up to 40%, and that the residual (very fast motion component) becomes close to white noise.

c. Principal component analysis (PCA). By applying this method, we employ cross-correlation between the total load forecast errors in different zones of the WECC system. Thai is a new idea. The PCA helps to reduce the number of dimensions, where the predictable forecast components of multiple zonal errors can be located. This helps to improve the accuracy of the forecasts, and make the residual error close to white noise.

- We developed and demonstrated new methods for quantification and adjustment of transmission limits using probabilistic methods (in the WECC system). Using the Transmission Uncertainty and Prediction Tool (TUT), we demonstrated how the transmission limits in the WECC system can be quantified and adjusted to minimize the risk of violations caused by uncertainty. This adjustment can also help to increase the utilization of transmission facilities within capacity and reliability limits.

- PNNL Transmission Uncertainty and Prediction Tool (TUT). Interactions among wind, solar gen load forecast errors can have a random impact on transmission system. This event can lead to additional congestion and voltage stability problems. To mitigate this problem, balancing authorities (BAs) should be able to re-dispatch their conventional generators ahead of time, if such impact can be predicted. A new probabilistic methodology and new analytical software called the Transmission Uncertainty and Prediction Tool (TUT) was developed at Pacific Northwest National Laboratory using funding from DOE Office of Energy (OE), DOE Office of Energy Efficiency and Renewable Energy (EERE), and the California Energy Commission (CEC). The developed methodology predicts the impact of the uncertainties on congested paths, worst-case voltage drops, and reactive power margins for several hours ahead of time and, if needed, proposes control actions to mitigate the problem if necessary.

- We demonstrated how TUT can benefit from parallelization.

1.4 References

[1] Y.V. Makarov, M. Hunsaker, A. Diaz-Gonzalez, R. T. Guttromson, P. Du, P.V. Etingov, H. Ghoudjehbaklou, J. Ma, D. Tovar, V.V. Viswanathan, and B. Vyakaranam, 2013. “WECC Variable Generation Planning Reference Book: A Guidebook for Including Variable Generation in the Planning Process.” Salt Lake City, Utah. Volume 1: Main Document, Version 1, May 14, 2013. [Online.] Available: http://www.wecc.biz/committees/StandingCommittees/JGC/VGS/PWG/ActivityP8/Final%20Version/WECC_Variable_Generation_Planning_Reference_Book_Main.docx. Accessed 25 September 2014.

[2] Y.V. Makarov, M. Hunsaker, A. Diaz-Gonzalez, R. T. Guttromson, P. Du, P.V. Etingov, H. Ghoudjehbaklou, J. Ma, D. Tovar, V.V. Viswanathan, and B. Vyakaranam, 2013. “WECC Variable Generation Planning Reference Book: A Guidebook for Including Variable Generation in the Planning Process.” Salt Lake City, Utah. Volume 2: Appendices, Version 1, May 14, 2013. [Online.] Available: http://www.wecc.biz/committees/StandingCommittees/JGC/VGS/PWG/ActivityP8/Final%20Version/WECC_Variable_Generation_Planning_Reference_Book_Appendices.docx. Accessed 25 September 2014.

2.0 Project Objectives

This project develops a new paradigm and new methodologies for real-time and offline power system analytics, representing a power system as a stochastic machine that behaves randomly and exhibits different characteristics compared with the largely deterministic systems that we dealt with in the past.

The tasks included in this first phase of the work are as follows:

- Develop multivariate geographically distributed model (MGDM) including all contributing uncertainty factors in their interaction.
- Characterize prevailing system motions (changing net load requirements) and associated uncertainty around them by a limited number of principal directions.
- Quantify and observe transmission uncertainty and security limits based on the developed MGDM.
- Develop methods for statistical separation of fast and slow changing components in the overall uncertainty model, their detection and prediction.
- Develop a prototype Transmission Uncertainty and Prediction Tool (TUT) and a framework for its high performance computing (HPC) applications.

The project will also develop and test new uncertainty quantification methods and tools for direct integration of uncertainty and variability information into grid operations on the transmission and distribution levels. The tools will be based on state-of-the-art probabilistic power production forecasts (including confidence intervals, inputs from distributed energy resources, and ramps), collective consideration of all sources of uncertainty (solar and wind generation, load, uninstructed deviations, and forced outages), as well as on prediction of MW uncertainty ranges of the total system load, critical path/line flows and interchange that the system operator will have to address or be aware of.

3.0 Sources of Uncertainty and the Importance of Their Concurrent Consideration

Uncertainties in forecasting the output of intermittent resources such as wind and solar generation, as well as system loads, are not reflected in existing tools used for generation commitment, dispatch and market operation. The same is true for other sources of uncertainties, such as uninstructed deviations of conventional generators from their set points, generator forced outages and failures to start, losses of major transmission lines, and frequency variations. These uncertainties can cause deviations in the system balance that require inefficient and costly last-minute solutions in the near real-time conditions. Major unexpected variations in wind power, unfavorably combined with load forecast errors and forced generator outages could cause significant power mismatches, which could be essentially unmanageable without knowing these variations in advance. In extreme cases, dispatch decisions could not be feasible because of the generators' start up, ramping, and capacity constraints. With the growing penetration of intermittent resources, the uncertainties could pose serious risks to control performance and desired operational characteristics, as well as the reliability of a power grid. Without knowing the risks caused by uncertainties, i.e. the probability, timing and magnitude of potential system imbalances, system operators have only very limited ability to evaluate the potential problems and find solutions to mitigate their adverse impacts. Some important questions need to be addressed in counteracting the impact of uncertainties. For instance, when should one start more units to balance against possible fast ramps in the future over a given time horizon? Would the available online capacity be sufficient to balance against variations of uncertain parameters on the intra-hour and minute-to-minute basis? The need to evaluate uncertainties associated with wind and solar generation and to incorporate the knowledge into the algorithms and operating practices is well understood already. Some wind forecast service providers offer uncertainty information for their forecasts. For instance, AWS Truepower [1] and 3TIER [2] companies developed wind generation forecasting tools with built-in capability to assess wind generation uncertainty. Similar tools have been developed in Europe. In the context of the European Union project, ANEMOS, a tool for on-line wind generation uncertainty estimation based on adaptive re-sampling or quantile regression has been developed [3]. A German company, Energy and Meteo Systems, developed a tool for forecasting wind generation, assessing the uncertainty ranges associated with wind forecast, and predicting extreme ramping events [4]. Reference [5] discusses a wind generation interval forecast approach using the quantile method. Reference [6] used statistical analysis based on standard deviation to predict wind generation forecast errors. Work is underway to incorporate these uncertainties into power system operations [7], [8]. Unfortunately, in many cases these efforts are limited to wind generation uncertainties only; they ignore additional sources mentioned at the beginning of this paragraph. Moreover, these approaches, while considering the megawatt imbalances, do not address essential characteristics such as ramp rate (megawatts per minute) and ramp duration uncertainties (minutes), required by the generators participating in the balancing process.

This section gives an overview of the sources of uncertainties in power systems, their important characteristics, models, and approaches for integrating information into system operations. The key discussion points are as follows:

- Variable generation and system load are far from being the only sources of uncertainties. Additional uncertainty is introduced by uninstructed deviations of conventional generators from their set points, forced outages of conventional generation, accidental load drops, major intermittent loads, unscheduled loss of major transmission lines, frequency variations, and other sources influencing power balance in a control area. All these sources must be accounted for.

- All sources of uncertainty interact in a complicated statistical way, so that most of the time their combined impact is reduced when compared to the sum of impacts of individual sources. At the same time, from time to time, collective impact of uncertainties can add up in an unfavorable way creating extreme system imbalances (so-called “tail events”).

- The statistical model of the uncertainties could be more complicated than it appears. It includes continuous unpredicted parameter variations and discrete sudden events (such as forced generation outages), non-parametric statistical distributions, non-stationary time dependent processes, tail events, autocorrelation and cross-correlation moments, and other external factors (such as temperature forecast errors or wind ramp prediction errors).

- Besides the capacity in megawatts, uncertainty prediction should include additional dimensions such as ramps (megawatts per minute), ramp duration (minutes), and cycling characteristics of conventional generators and energy storage facilities. Thus, the uncertainty analysis becomes multidimensional. These characteristics form a performance envelope necessary to successfully balance the system in view of uncertainties.

- Three modes of uncertainty integration are proposed: “passive,” “active,” and “proactive.” “Passive” integration is the first level of integration, which brings awareness of uncertainties into control center software tools through information visualization and alarming. “Active” integration uses the uncertainty information to modify existing grid operation functions such as unit commitment. “Proactive” integration develops new grid operation functions enabled by the uncertainty information.

3.1 System Power Balance and Procedures to Achieve It

An interconnected power system usually consists of one or multiple Balancing Authorities (BAs). Each BA must maintain a balance between its generation, load, interchange, and losses. The system’s conventional generation is committed and dispatched to meet the total balancing requirement, *BR* [MW]. BAs’ performance is judged based on Control Performance Standards (CPS) requiring a certain degree of success in keeping the system imbalance within certain (sometimes statistically defined) bounds. The system balancing requirement (BR), which is the same as net load, is the balancing job required from conventional generators, power exports, and energy storage facilities, is expressed by the formula:

$$BR = \Delta L - \Delta WG - \Delta SG - FO - UD + VDE - 10 \cdot B \cdot \Delta F - TE \tag{1}$$

where Δ – denotes the difference between the actual and the pre-scheduled values for dispatch intervals; L – is the system load; WG – is wind generation; SG – is solar generation; FO – reflects imbalances caused by forced generation outages; UD – is the total uninstructed deviation of conventional generation units (including failure to start up); VDE – is the variability errors within a dispatch interval (or difference between the block-energy schedules and the continuous actual variation of BA, e.g., the difference between the hour ahead schedule and actual generation in Figure 2); $10 \cdot B \cdot \Delta F$ – is a frequency dependent term; and TE – is the time error correction term. Depending on the context and information available, the meaning and the actual presence of the terms in (1) can change. For instance, in the scheduling and real-time dispatch procedures, the uninstructed deviation term, frequency dependent term, and time error term are omitted and the actual values are replaced by their forecasts.

The balance in a BA is achieved through several processes including scheduling (day-ahead and hour-ahead block hour scheduling), real-time dispatch (load following or real-time scheduling), and regulation. Figure 2 shows the relationship between these processes.

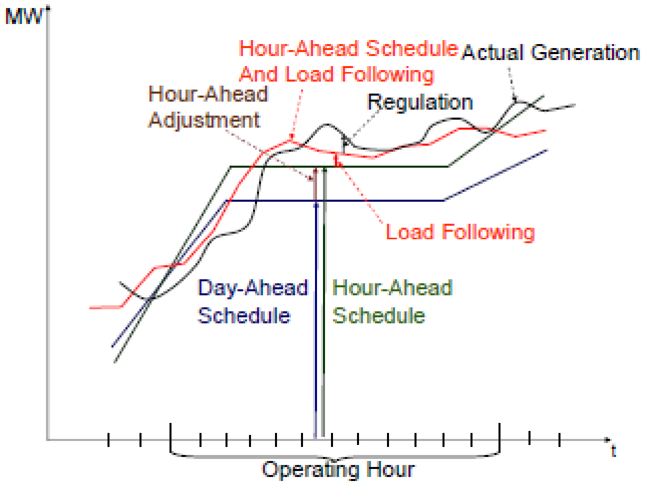

MW
Hour-Ahead Schedule
And Load Following
Actual Generation
Hour-Ahead
Adjustment
Regulation
Load Following
Day-Ahead
Schedule
Hour-Ahead
Schedule
t
Operating Hour

Figure 2. Relationship between the scheduling, load following and regulation processes

The scheduling and real-time dispatch processes deal with certain dispatch intervals (e.g., 1 clock hour, 30-minute or 5-minute intervals), where the power output is assumed to be equal to the average power requirement within the interval. The overall variability around these values is caused by the forecasted values for these average requirements as well as shown in Figure 3 [11], [12].

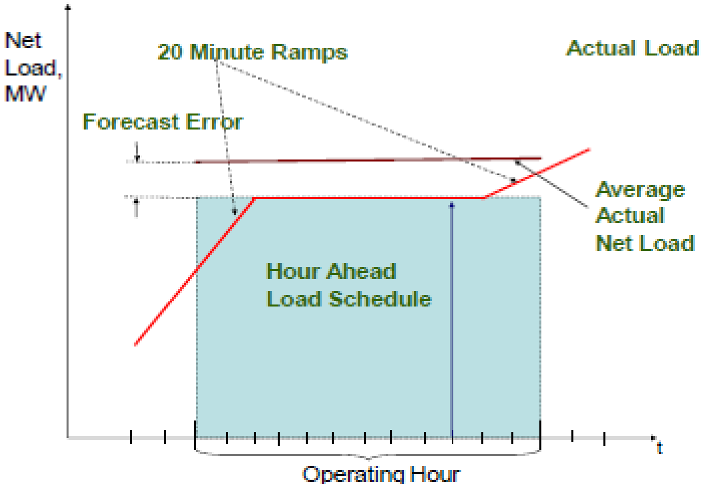

Net
Load,
MW
20 Minute Ramps
Actual Load
Forecast Error
Average
Actual
Net Load
Hour Ahead
Load Schedule
t
Operating Hour

Figure 3. Overall uncertainty within a dispatch interval

The transitional ramps shown between the dispatch intervals in Figure 2 and Figure 3 are used in some systems, e.g., the Western Interconnection in the U.S., to minimize discretization errors as well as reduce generator ramping requirements. These 20-minute ramps are also applied to balancing areas' interchange schedules.

3.2 Sources of Uncertainties and Their Characteristics

Multiple sources of uncertainty contribute to the overall uncertainty facing the system. There are uncertainties contributing to the system balancing requirement *BR* (right hand side of equation (1)), and uncertainties associated with the market-driven generation dispatch and caused by limited knowledge of future generation bids. This report concentrates on uncertainties of the first kind.

3.2.1 Load Forecast Errors

Load forecast errors, ΔL in equation (1), contribute significantly to the overall uncertainty of the system's balancing requirement *BR*. In operations, load forecasts are usually provided for the next operating day (hourly block energy schedules), next operating hour (hourly block energy schedules or schedules for average load for smaller dispatch intervals), and in real-time forecasts (average load for within-hour dispatch intervals, e.g., 5-10-15 minutes). The day-ahead mean absolute percent error (MAPE) usually stays within 1-3% of the maximum load – see Figure 4. It is important to mention that instantaneous error values can significantly exceed MAPE from time to time contributing to so-called tail events [15].

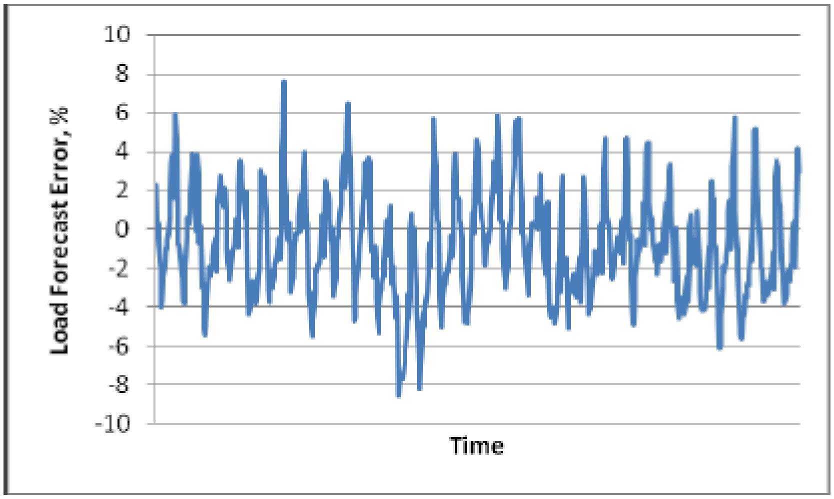


Figure 4. Example of load forecast error in a real system [15]

Load forecast errors depend upon multiple factors including temperature and humidity forecast errors. This sensitivity is changing with the observed air temperature – Figure 5 [16].

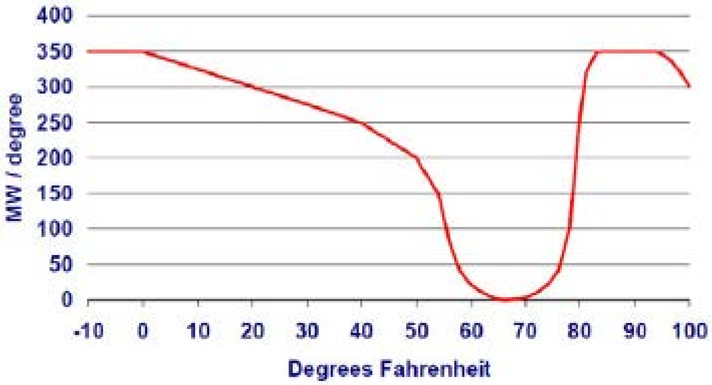


Figure 5. Impact of temperature forecast error on load

Another important issue with forecast is the forecast bias (systematic overestimation and underestimation of the system load).

3.2.2 Wind Power Forecast Errors

The issues related to the statistical characteristics of wind forecast error attract significant interest from researchers and practicing engineers. A good review of the state-of-the-art in this area is given in [18].

The wind power forecasting errors are sometimes simulated using the truncated normal distribution whose characteristics are determined by curve fitting. The truncation process is applied to reflect the natural limits posed by minimum (zero) and maximum (installed capacity minus the capacity of offline units) wind generation. This model cannot be accepted without caution because of indications that the wind power forecast error does not actually follow the truncated normal distribution. There are efforts in place to propose better approximations for the error. For instance, the Beta distribution has been used in [17]. In [14], a different approach based on experimental probability density functions (PDFs) is proposed instead. The approach is suitable for handling non-parametric distributions, so that no hypothesis is needed regarding the wind generation error distribution law.

There is a continuing effort in place to improve the accuracy of wind generation forecasting algorithms. For example, in Germany the day-ahead wind generation forecast error has been reduced to 4.5% [19]. Nevertheless, significant challenges remain with the very short term forecasts (the naïve persistence forecast model frequently demonstrated better performance than more scientific approaches) and prediction of wind generation ramps.

Wind generation forecast errors are sensitive to multiple external factors. Accurate modeling of these errors requires models that reflect these factors. In [20], a Bayesian network model is developed that reflects the influence of external factors on the wind and load forecast errors.

3.2.3 Solar Generation Forecast Errors

Characteristics of the solar power forecast errors have not been sufficiently well studied yet. Nonetheless, practical needs require these models. In [21], a new model has been proposed and has been used in the California Independent System Operator (ISO) 20% and 33% penetration of renewable generation studies.

3.2.3.1 Variation of Solar Generation

Solar radiation (or solar irradiance) determines the level of solar energy production at any specific location and is neither completely random nor completely deterministic. Extraterrestrial (above the clouds) solar radiation can be confidently predicted for any place and time interval. Solar radiation shows both yearly and daily variation. The area’s atmospheric conditions (clouds, dust storms, etc.) determine the randomness of solar radiation at the ground level (also called global solar radiation).

The ranges of yearly variation can be described by monthly maximum solar radiation for a sunny day, and the minimum solar radiation for a total cloudy day. The maximum and minimum solar radiation levels can be used for solar power generation forecasting as well as for forecast error simulations. The effect of clouds and other factors on solar generation varies. Minute-to-minute global solar radiation measurements R_m show that radiation varies within a one-hour period around an hourly mean value R_h. A statistically varying term (ε_R) can be used to simulate this process. This statistical term was set to have the same distribution as short duration variations seen in actual measurements. Therefore, it can be evaluated by using actual measurements,

$$\mathrm{R_m} = \mathrm{R_h} + \varepsilon_\mathrm{R} \tag{2}$$

3.2.3.2 Clearness Index (CI).

If the sky is clear, solar radiation and solar power production are predictable based on the annual and daily extraterrestrial pattern. Thus, solar forecast errors are small. Solar radiation and generation forecast errors are mostly caused by clouds and other factors. These factors include: clouds (depth, water or ice concentration, and types of water particles or ice crystals), water vapor amount, and aerosol type and amount (column).

The clearness index is an index indicating what percentage of the sky is clear. High CI could mean higher global solar radiation (i.e., global solar radiation levels being closer to their extraterrestrial values) and lower forecast errors. CI is used for solar power generation forecasting. Clearness index for a given period is obtained by dividing the observed global radiation R_g by extraterrestrial global irradiation R:

$$k = \frac{R_g}{R} \tag{3}$$

where R_g is horizontal global solar radiation and R is horizontal extraterrestrial solar radiation.

3.2.3.3 Simulating the Solar Generation Forecast Errors

Statistical characteristics of hour-ahead and real-time solar generation forecast errors are complex and depend on various factors including the extraterrestrial solar radiation annual and daily patterns, hour-to-hour clearness index, dynamic patterns of the cloud systems, types of solar generators (photovoltaic (PV), concentrated thermal, etc.), geographical location and spatial distribution of solar power plants, and other factors.

3.2.3.4 Upper and Lower Limits for Solar Generation Forecast Errors

Unlike wind generation, solar generation is limited by extraterrestrial solar irradiance levels, which change over a day. The maximum possible generation can be achieved at CI = 1 and this maximum value Pmax (t) also changes over a day following a similar mostly deterministic pattern (note that there is also an annual component in this process). Variances of the generation under these conditions can be only caused by diffuse solar irradiance and ambient temperature variations. Assuming that these variances are also included in $P_{max}(\ t)$, the maximum solar generation during day time can be described as a function of time, and is always less than the total capacity, i.e.,

$$SG_a(t) \leq P_{max}(t) \tag{4}$$

where $P_{max}(t)$ is the maximum solar generation capacity.

Solar power generation forecast $f(t)$ can be modeled by the relation:

$$P_{min}(t) \leq f(t) = SG_a(t) - \varepsilon(t) \leq P_{max}(t) \quad (5)$$

where the minimum solar power capacity $P_{min}(t)$ could be assumed to be zero; the maximum capacity of solar farm generation $P_{max}(t)$ is a function of time. During the night time,

$$f(t) = SG_a(t) = \varepsilon(t) = 0 \quad (6)$$

From Equation (5), we have,

$$SG_a(t) - P_{max}(t) \leq \varepsilon(t) \leq SG_a(t) \quad (7)$$

where $SG_a(t) \leq P_{max}(t)$ may be negative or zero.

3.2.3.5 Standard Deviation of Solar Forecast Error Evaluated Using CI

Different solar generation patterns in day time and night time need to be taken into account when evaluating solar forecast errors. At night time, solar irradiance and solar generation are zero, thus solar forecast errors are zero. Sunrise and sunset time are different in different seasons at different regions. Then daily patterns of the clearness index are different. Previous years' information regarding this matter can be categorized and used for solar forecast error evaluation.

Depending on the time period and weather conditions, solar forecast errors can show different patterns, such as:

a. Forecast error is zero, $\varepsilon = 0$, at night time;

b. Forecast error is small or close to 0, $\varepsilon \rightarrow 0$, on sunny days, that is when $CI \rightarrow 1$;

c. Forecast error is limited or close to zero under heavily clouding conditions, that is when $CI \rightarrow 0$; and

d. Forecast error varies in a wide range for the intermediate values of *CI*.

Thus, the standard distribution of solar forecast errors can be described as a function of *CI*, e.g., $0 \leq std_{min} \leq std(\varepsilon) = f(CI) \leq std_{max}$. Figure 6 shows a possible distribution of the standard deviation of solar forecast errors depending on CI.

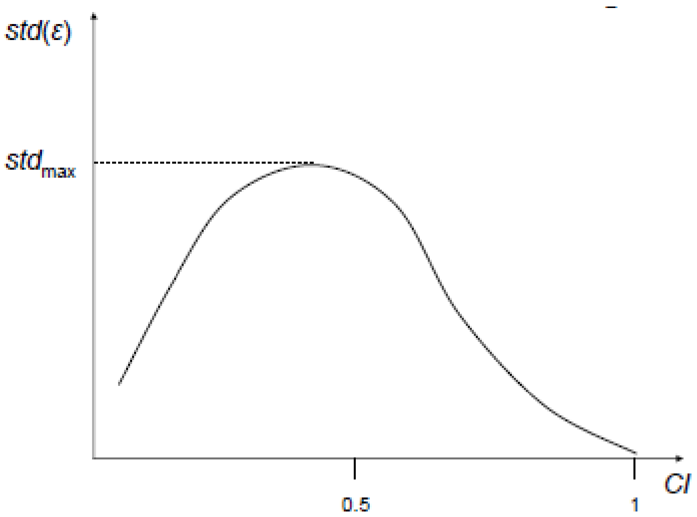


Figure 6. Distribution of the standard deviation of solar forecast errors depending on the clearness index

For a sunny day, the variation of solar forecast errors is shown in Figure 7(a). The forecast error can be predominantly negative. If the sky is completely covered with clouds, the distribution of solar forecast error could be similar to Figure 7 (b). The forecast error can be predominantly positive.

3.2.3.6 Real-time Solar Forecast Model

A persistence model is used for the real-time wind forecast, but for solar generation, there are obvious incremental patterns during morning hours just after sunrise, and decremental patterns during evening hours just before sunset. Solar generation could increase or decrease dramatically in a very short time in the sunrise or sunset hours. This will cause significant ramp rate increase during these hours. The persistence model cannot address this concern. Therefore, a new model based on CI is proposed in [21] to simulate real-time solar forecast.

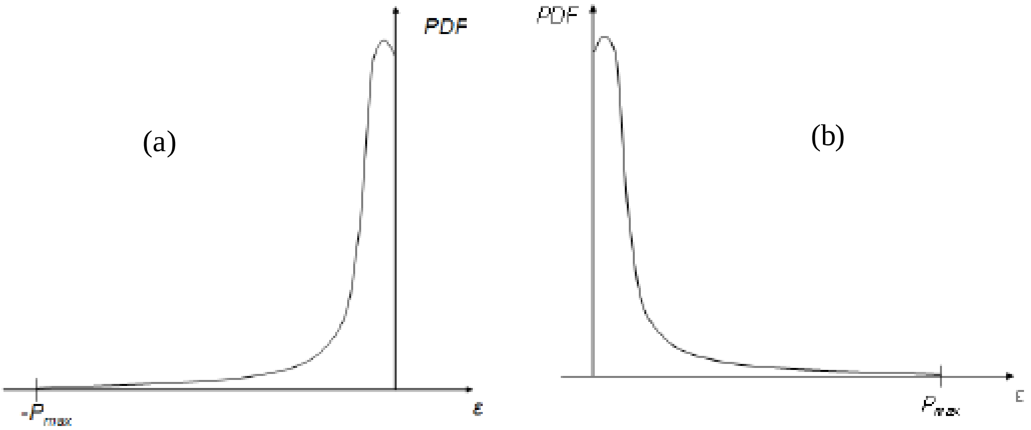


Figure 7. Distribution of solar forecast error in a very cloudy day and a very sunny day

3.2.4 Forced Generation Outages

Generator outages are addressed by specially procured contingency reserves. The imbalances caused by forced generator outages are initially mitigated by the system governor response and the automatic generation control (AGC) system, and then either by (1) committing and dispatching generation resources suitable for the intra-hour balancing purposes, or (2) applying load reduction schemes. It could take 5-15 minutes or even more to activate them. As a result, the system imbalances caused by forced generation outages could last for about 5-15 minutes. A schematic model for the balancing requirement is shown in Figure 8.

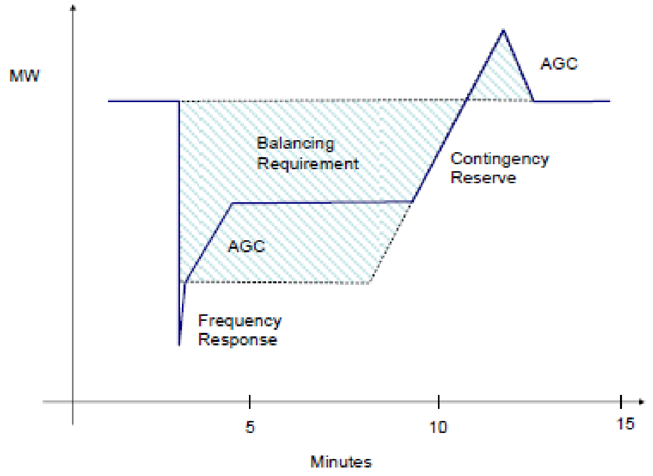


Figure 8. Balancing requirement caused by forced generation outage.

After a forced outage, the corresponding control area is subjected to a sudden imbalance which depends on the level of generation on the tripped unit, system inertia, and available frequency (governor) response. The part of the initial imbalance addressed by governor response does not contribute to the balancing requirement because of the fast recovery process involved (seconds). The AGC system in the affected control area starts to move regulating units to cover another part of the system imbalance (minutes). The job done by regulating units is addressing a part of the overall balancing requirement. Frequently the available regulating reserve is not sufficient to completely restore the system balance. The imbalance stays in the system for about 10 minutes until the contingency reserve units are started, synchronized with the system (non-spinning reserve), and dispatched to the desired level (both non-spinning and spinning reserves). Then the AGC units are moved back to completely restore the balance. A model for forced generation outage and contingency reserve activation has been developed by Washington University under a contract with PNNL [15]. The model is based on the following elements:

a. Forced outage rates and dispatch level of online units.

b. Probability of having any online unit forced out any minute.

c. Contingency reserve activation model.

3.2.5 Uninstructed Deviation Errors

The impacts of uncertainties caused by forced generation outages, failures to start up, and uninstructed deviations of conventional generators on the system balancing requirements are frequently neglected. Nevertheless, the total uninstructed deviations resulting from their inability to follow the set points precisely could reach several hundred megawatts and may have a profound impact on the system balancing requirements.

3.3 Statistical Characteristics of Forecast Errors and Balancing Requirement

Statistical characteristics of uncertainties are complicated (e.g., they can form non-parametric distributions or represent non-stationary processes) and depend upon (1) the forecast horizon; (2) other forecast errors (cross-correlation); (3) time of the day; (4) external factors such as air temperature (load and wind forecast errors) and humidity (wind forecast errors); (5) clearness index (solar forecast error); (6) net load ramps; (7) forced generation outages (uninstructed deviations); and (8) other factors.

3.3.1 Discretization Errors

Discretization errors are the difference between the scheduled values within a dispatch interval and the actual minute-by-minute variations of the balancing requirement (discretization errors) – see Figure 2. Unlike the forecast errors, the discretization errors are functions of the variability of balancing requirement and the size of dispatch intervals.

3.3.2 Non-parametric nature

Nonparametric distributions cannot be described using a standard probability distribution, e.g., the normal distribution, although sometimes they can be approximated using a standard distribution or a combination of distributions with certain limited accuracy. For instance, in [11], [12], the truncated normal distribution was used to approximate distributions of wind and load forecast errors. The balancing requirement distribution could be a more difficult case. Sometimes its shape becomes essentially non-parametric. Figure 9 contains an example of histograms of a non-parametric distribution of the regulation requirement in a large BA with wind and solar variable energy resources.

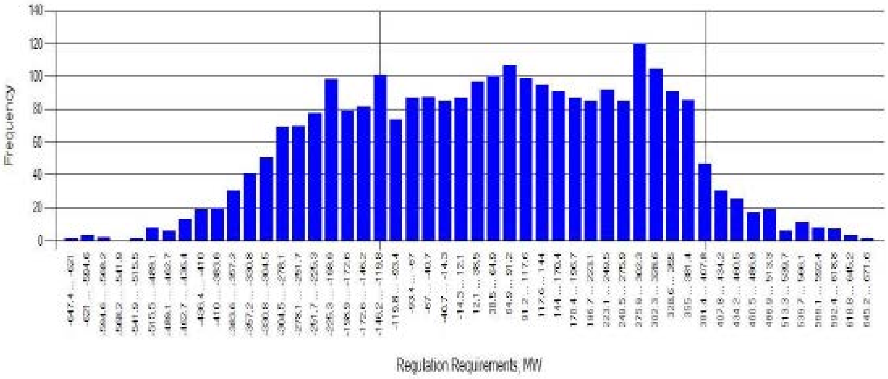


Figure 9. Hourly regulation requirement for a real BA

Frequently, the *BR* distribution exhibits long "heavy" tails indicating certain limited probability of major imbalances caused by unfortunate combinations of random factors contributing to *BR* in Equation (1). Normally, the central part of the *BR* distribution could be or should be balanced using the existing balancing reserves, whereas the tail part could create infrequent but significant problems because it requires balancing reserves that are not normally procured in the system.

3.3.3 Autocorrelation of Wind and Load Forecast Errors

The load and wind forecast errors usually exhibit strong autocorrelation between the subsequent forecasts. Autocorrelation means that, for example, if for certain operating hour, a large positive forecast error is observed, it is likely that a similar error would be observed for the next hour. The autocorrelation can be observed not only between the subsequent errors (*i*-1, *i*), but also between errors (*i*-2, *i*), (*i*-3, *i*), … Figure 10 and Figure 11 illustrate autocorrelation coefficients of real load and wind generation forecast errors (as well as those generated from error models) [14].

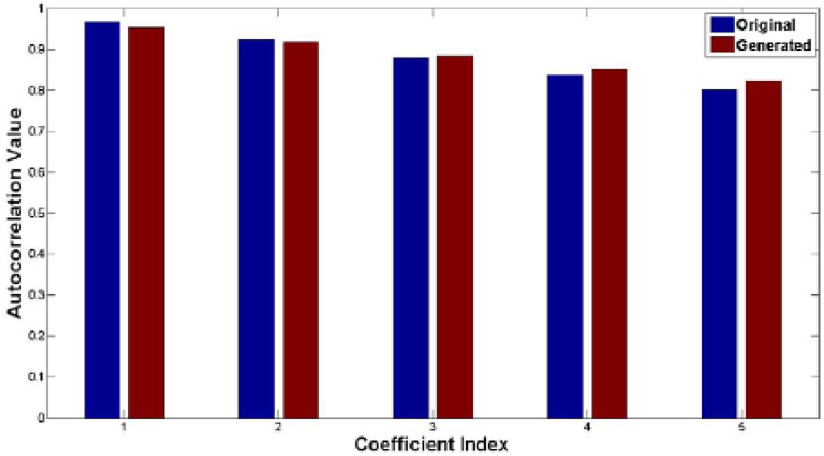


Figure 10. Autocorrelation of original and simulated load forecast errors

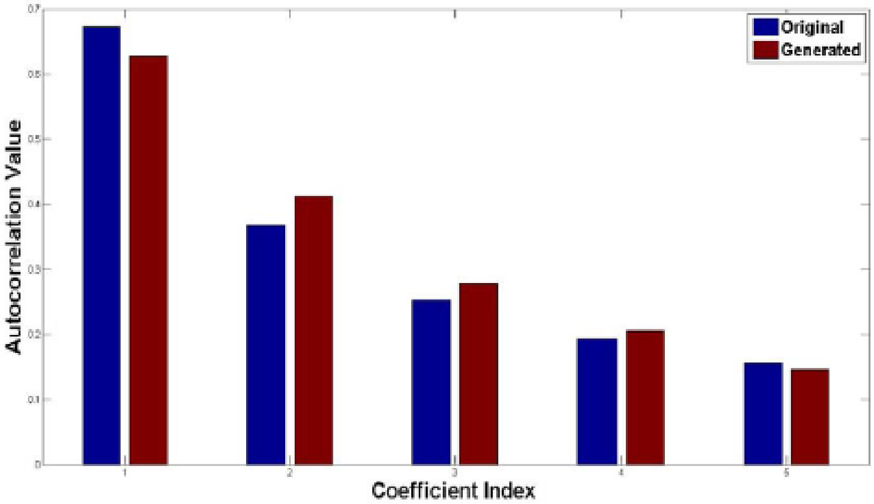


Figure 11. Auto-correlation of original and simulated wind forecast errors

3.3.4 Cross-correlation

Wind forecast errors also show correlations between wind farms at different locations. A study was performed to evaluate the performance of three wind farms that are geographically distant as shown in Figure 12 [14]. The comparison of the statistics is shown in Table I.

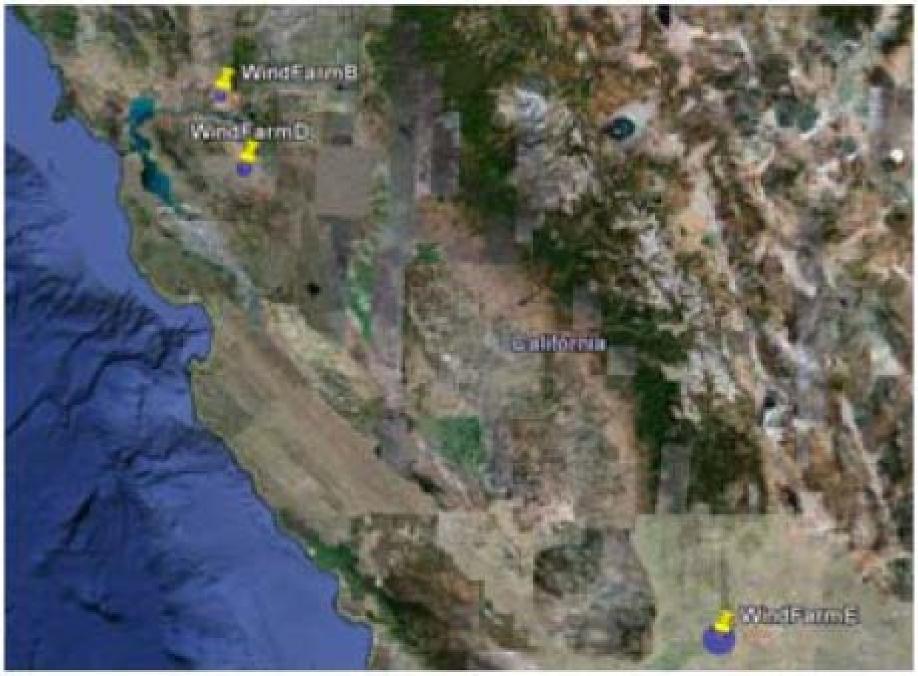


Figure 12. Geographical locations of the wind farms.

Table 1. Cross-correlation of wind forecast errors.

Cross-Correlation	B	D	E
B	1	0.6047	0.1194
D	0.6047	1	0.2121
E	0.1194	0.2121	1

3.3.5 Non-stationary Nature

In non-stationary processes, statistical characteristics of the process change over time. Figure 13 shows real-life examples of such processes for wind and load forecast errors [14]. The original distribution is the actual distribution of the load and wind forecast errors. It demonstrates non-stationary patterns in terms of its moving average as well as the variance. The “new sequence” results represent an effort to simulate the same distribution using a stationary model and a specially designed random number generator. The model was successfully reproducing the mean, standard deviation, and even autocorrelation of the original series, but fails to reproduce the time-varying components.

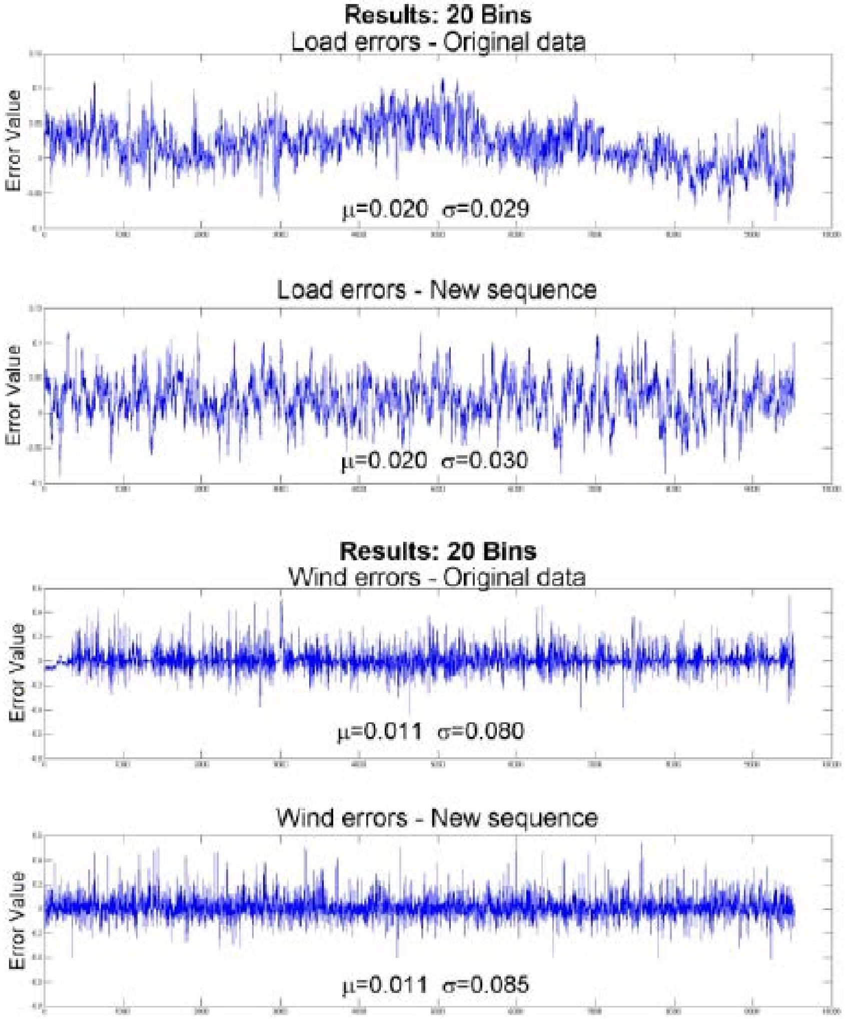


Figure 13. One-year hourly original and simulation results for load and wind forecast errors.

3.4 References

[1] J. Zack, "An analysis of the errors and uncertainty in wind power production forecasts," in Proc. WINDPOWER 2006, Pittsburgh, June 4-7, 2006.

[2] J. Lerner, M. Grundmeyer and M. Garvert, "The role of wind forecasting in the successful integration and management of an intermittent energy source," Energy Central, Wind Power, vol. 3, no. 8, July 2009.

[3] G. Kariniotakis, "ANEMOS, leading European Union research on wind power forecasting," in Proc. International Wind Forecast Techniques and Methodologies Workshop, July 24 -25, 2008. [Available online]: http://www.bpa.gov/corporate/business/innovation/docs/2008/BPA_California%20ISO%20ANEMOS%20Presentation.pdf.

[4] "Wind power prediction Previento," Energy & Meteo Systems. [Available online]: http://energymeteo.de/de/media/e_m_Broschuere.pdf.

[5] P. Pinson, G. Kariniotakis, H. Aa. Nielsen, T. S. Nielsen, and H. Madsen, "Properties of quantile and interval forecasts of wind generation and their evaluation," in Proc. European Wind Energy Conference & Exhibition, Greece, 2006.

[6] A. Luig, S. Bofinger and H. G. Beyer, "Analysis of confidence intervals for the prediction of the

regional wind power output," in Proc. 2001 European Wind Energy Conference, Denmark, June 2001.

[7] J. Kehler; M. Hu, M. McMullen, J. Blatchford "ISO Perspective and Experience with Integrating Wind Power Forecasts Into Operations," in Proc. IEEE General Meeting, Minneapolis, 2010.

[8] D. Maggio, C. D'Annunzio, S.-H. Huang, C. Thompson "Utilization of Forecasts for Wind-Powered Generation Resources in ERCOT Operations," in Proc. IEEE General Meeting, Minneapolis, 2010.

[9] Red Eléctrica de España (REE), "Wind Development, Integration Issues and Solutions – TSO Spain", presentation in The Northwest Wind Integration Forum, Portland, Oregon, July 29 and 30, 2010. [Online.] Available: http://www.nwcouncil.org/energy/wind/meetings/2010/07/

[10] Y. V. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational Impacts of Wind Generation in California", IEEE Transactions on Power Systems, Vol. 24, No. 2, May 2009.

[11] S. Lu, Y.V. Makarov, B. McManus and J. Pease, "Methodologies to Establish Load Following and Regulation Requirements for Large-Scale Wind Integration," WindPower 2010 Conference and Exhibition, Dallas, Texas, May 23-26, 2010.

[12] Y. V. Makarov, S. Lu, B. McManus, and J. Pease, "The Future Impact of Wind on BPA Power System Ancillary Services", Proc. WindPower 2008, Paper #315, Houston, TX, June 1-4, 2008.

[13] Y. V. Makarov, Z. Huang, P. V. Etingov, J. Ma, R. T. Guttromson, K. Subbarao, and B. B. Chakrabarti, "Incorporating Wind Generation and Load Forecast Uncertainties into Power Grid Operations," PNNL-19189, Pacific Northwest National Laboratory, Richland, WA, January 2010.

[14] Y.V. Makarov, J.F. Reyes-Spindola, N. Samaan, R. Diao, and R.P. Hafen, "Wind and Load Forecast Error Model for Multiple Geographically Distributed Forecasts," Proc. 9th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants. Quebec City, Quebec, Canada, October 18-19, 2010.

[15] Y. V. Makarov, Z. Huang, P. V. Etingov, J. Ma, R.T. Guttromson, K. Subbarao, and B. B. Chakrabarti, "Wind Energy Management System EMS Integration Project: Incorporating Wind Generation and Load Forecast Uncertainties into Power Grid Operations," PNNL-19189, Pacific Northwest National Laboratory, Richland, WA, 2010. [Online.] Available: http://www.pnl.gov/main/publications/external/technical_reports/PNNL-19189.pdf.

[16] P. Walshe, "Temperature and Demand Forecasting in a Large Utility," TVA Presentation. [Online.] Available: http://www.isse.ucar.edu/electricity/workshop/Presentations/pdf/Walshe.pdf.

[17] H. Bludszuweit, J.A. Dominguez-Navarro, and A. Llombart, "Statistical Analysis of Wind Power Forecast Error," IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 983–991, Aug. 2008.

[18] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, "Wind Power Forecasting: State-of-the-Art 2009," Report ANL/DIS-10-1, Argonne National Laboratory, Illinois, November 6, 2009. [Online.] Available: http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=968212.

[19] Ernst B, U Schreirer, F Berster, C Scholz, HP Erbring, S Schlunke, J Pease, and YV Makarov, "Large Scale Wind and Solar Integration in German," PNNL-19225, Pacific Northwest National Laboratory, Richland, WA. [Online.] Available: http://www.pnl.gov/main/publications/external/technical_reports/PNNL-19225.pdf.

[20] S. Lu, Y. V. Makarov, A. J. Brothers, C. A. McKinstry, S. Jin, and J. Pease, "Prediction of Power System Balancing Requirement and Tail Event," IEEE PES Transmission and Distribution Conference 2010, pp. 1-7.

[21] Y.V. Makarov, J. Ma, C. Loutan, and G. Rosenblum, "Solar Forecast Error Simulation Methodology for CAISO 33% Renewables Study," PNNL Technical Report, Prepared for CAISO, Jan.2010

4.0 Western Interconnection Uncertainty Model Development

Task 1 in our DOE OE "Probabilistic methods for planning and operation" project includes the development of a state-of-the-art model for geographically distributed uncertainty to incorporate multiple uncertainties from load, wind, and solar forecast and their interaction. This model development is based on a PNNL-Northwest Power Pool (NWPP) report [1]. The model will be able to handle multi-variable, distributed, non-parametric, non-stationary processes with cross-correlation and auto-correlation. This model will produce multiple combinations of node-specific uncertain and random parameters needed for multi-variant runs (e.g., Monte Carlo runs) of probabilistic tools for planning and operations (e.g., probabilistic power flow, probabilistic contingency analysis, probabilistic stability analysis, stochastic unit commitment and optimal dispatch, etc.)

4.1 Primary Dataset Description

Load, wind and solar values are from the "WECC TEPPC 2020 PC0" case [2]. The model is based on the "WECC 2020 HS1A" power flow base case.

4.1.1 WECC BA Structure

WECC currently consists of 38 BAs with six of these being generation-only BAs (that do not operate transmission capacity). The 38 BAs are shown in Figure 14. In the model for the TEPPC 2020 case, the BAs in WECC are grouped into seven sub-regions. The model used for the TEPPC case has 39 load areas, as shown in Figure 15.

In this study, 32 BAs are modeled; the six generation-only BAs are not modeled, but instead their generation resources are modeled as belonging to other BAs. Each BA has its own load profile with the following exceptions:

- The California Independent System Operator (CAISO) is divided into four load areas (Pacific Gas and Electric Company [PG&E] Valley, PG&E Bay, Southern California Edison [SCE] and San Diego Gas and Electric [SDGE]),
- Idaho Power Company (IPC) is divided into three load areas (Treasure Valley, Magic Valley and Far East), and
- PacifiCorp East (PACE) is divided into three load areas (PACE ID, PACE WY, and PACE UT).

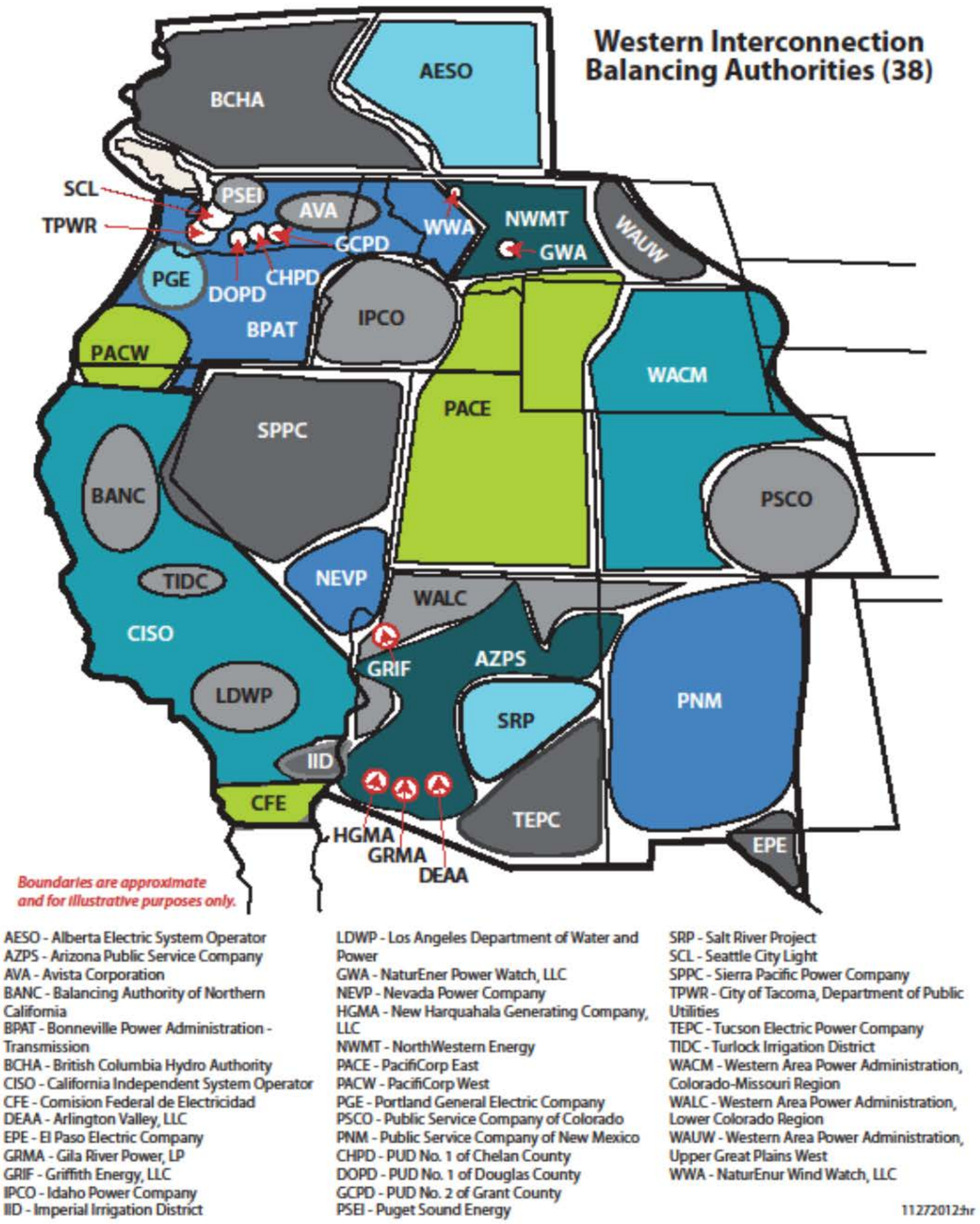

Western Interconnection Balancing Authorities (38)
BCHA
AESO
SCL
TPWR
PSEI
AVA
GCPD
WWA
NWMT
GWA
WAUW
PGE
DOPD
CHPD
BPAT
IPCO
PACW
PACE
WACM
SPPC
BANC
PSCO
TIDC
NEVP
WALC
CISO
GRIF
AZPS
LDWP
SRP
PNM
IID
CFE
TEPC
EPE
HGMA
GRMA
DEAA
Boundaries are approximate and for illustrative purposes only.
AESO - Alberta Electric System Operator
AZPS - Arizona Public Service Company
AVA - Avista Corporation
BANC - Balancing Authority of Northern California
BPAT - Bonneville Power Administration - Transmission
BCHA - British Columbia Hydro Authority
CISO - California Independent System Operator
CFE - Comision Federal de Electricidad
DEAA - Arlington Valley, LLC
EPE - El Paso Electric Company
GRMA - Gila River Power, LP
GRIF - Griffith Energy, LLC
IPCO - Idaho Power Company
IID - Imperial Irrigation District
LDWP - Los Angeles Department of Water and Power
GWA - NaturEner Power Watch, LLC
NEVP - Nevada Power Company
HGMA - New Harquahala Generating Company, LLC
NWMT - NorthWestern Energy
PACE - PacifiCorp East
PACW - PacifiCorp West
PGE - Portland General Electric Company
PSCO - Public Service Company of Colorado
PNM - Public Service Company of New Mexico
CHPD - PUD No. 1 of Chelan County
DOPD - PUD No. 1 of Douglas County
GCPD - PUD No. 2 of Grant County
PSEI - Puget Sound Energy
SRP - Salt River Project
SCL - Seattle City Light
SPPC - Sierra Pacific Power Company
TPWR - City of Tacoma, Department of Public Utilities
TEPC - Tucson Electric Power Company
TIDC - Turlock Irrigation District
WACM - Western Area Power Administration, Colorado-Missouri Region
WALC - Western Area Power Administration, Lower Colorado Region
WAUW - Western Area Power Administration, Upper Great Plains West
WWA - NaturEnur Wind Watch, LLC
11272012:hr

Figure 14. Current balancing authorities in WECC.[1]

---

[1] Source: http://www.ferc.gov/market-oversight/mkt-electric/wecc-balancing-authorities.pdf

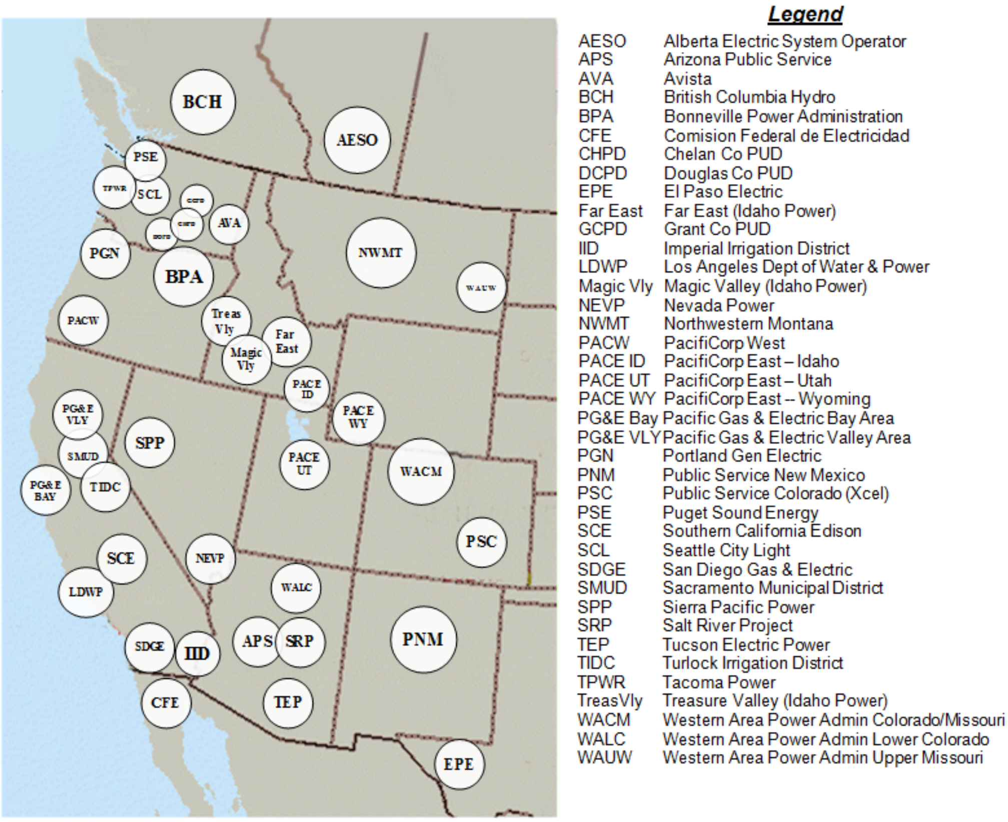

BCH
AESO
PSE
TPWR
SCL
AVA
PGN
BPA
NWMT
WAUW
PACW
Treas Vly
Far East
Magic Vly
PACE ID
PG&E VLY
SPP
PACE WY
SMUD
PACE UT
WACM
PG&E BAY
TIDC
SCE
NEVP
PSC
LDWP
WALC
SDGE
IID
APS
SRP
PNM
CFE
TEP
EPE
Legend
AESO Alberta Electric System Operator
APS Arizona Public Service
AVA Avista
BCH British Columbia Hydro
BPA Bonneville Power Administration
CFE Comision Federal de Electricidad
CHPD Chelan Co PUD
DCPD Douglas Co PUD
EPE El Paso Electric
Far East Far East (Idaho Power)
GCPD Grant Co PUD
IID Imperial Irrigation District
LDWP Los Angeles Dept of Water & Power
Magic Vly Magic Valley (Idaho Power)
NEVP Nevada Power
NWMT Northwestern Montana
PACW PacifiCorp West
PACE ID PacifiCorp East – Idaho
PACE UT PacifiCorp East – Utah
PACE WY PacifiCorp East -- Wyoming
PG&E Bay Pacific Gas & Electric Bay Area
PG&E VLY Pacific Gas & Electric Valley Area
PGN Portland Gen Electric
PNM Public Service New Mexico
PSC Public Service Colorado (Xcel)
PSE Puget Sound Energy
SCE Southern California Edison
SCL Seattle City Light
SDGE San Diego Gas & Electric
SMUD Sacramento Municipal District
SPP Sierra Pacific Power
SRP Salt River Project
TEP Tucson Electric Power
TIDC Turlock Irrigation District
TPWR Tacoma Power
TreasVly Treasure Valley (Idaho Power)
WACM Western Area Power Admin Colorado/Missouri
WALC Western Area Power Admin Lower Colorado
WAUW Western Area Power Admin Upper Missouri

Figure 15. TEPPC topology diagram for 2020 base case.

4.1.2 Load Data

Because no load data with 1-minute resolution are provided for the study year 2020, these data need to be generated based on selecting reasonable assumptions. The available load data for the study were the following:

1. Hourly loads for the year 2020 for the 39 load areas from the TEPPC 2020 PC0 case.
2. Actual minute-by-minute load data for the year 2009 for the 32 BAs that are not generation-only BAs.

With this information, the approach used was to impose the minute-to-minute variability of the 2009 load data on to the 2020 hourly load data. The procedures applied to generate the required 1-minute load data for the study year 2020 are the following [3]:

1. Compute a time series of hourly average load data for all 32 BAs in 2009, *Load_1h_avg_2009*, with 1-minute resolution.
2. Apply the nonlinear interpolation method in MATLAB® to *Load_1h_avg_2009* and obtain a new interpolated load series, *Load_2009_interpolated*, shown in Figure 16.
3. Calculate the *error* between *Load_actual_2009* and *Load_2009_interpolated*, indicating the differences between the actual load and the interpolated load, shown in Figure 17.

4. Normalize the *error* based on the peak load in 2009 for each BA individually and scale the *error* by multiplying the peak load in 2020, to obtain *Error_2020_load*. The error for the day 2/29/2020 is taken directly from the error data on the previous day, 2/28/2020.
5. Take the provided hourly load data in 2020 and interpolate the 1-hour resolution data to obtain interpolated load data, *Load_2020_interpolated*, with 1-minute resolution.
6. Apply the *Error_2020_load* to *Load_2020_interpolated* and obtain the desired load curves in 2020, with 1-minute resolution, shown in Figure 18.

This procedure was applied to each of the 32 BAs, which generated the 1-minute load curves for the entire year 2020 shown in Figure 16.

It should be noted that for the BAs having more than one load profile, the same minute-to-minute variability were applied to all load profiles because the 2009 data were available only at the BA level. The peak load in each BA is plotted in Figure 20. The names of the BAs corresponding to the BA numbers in the different figures are reported in Table 2.

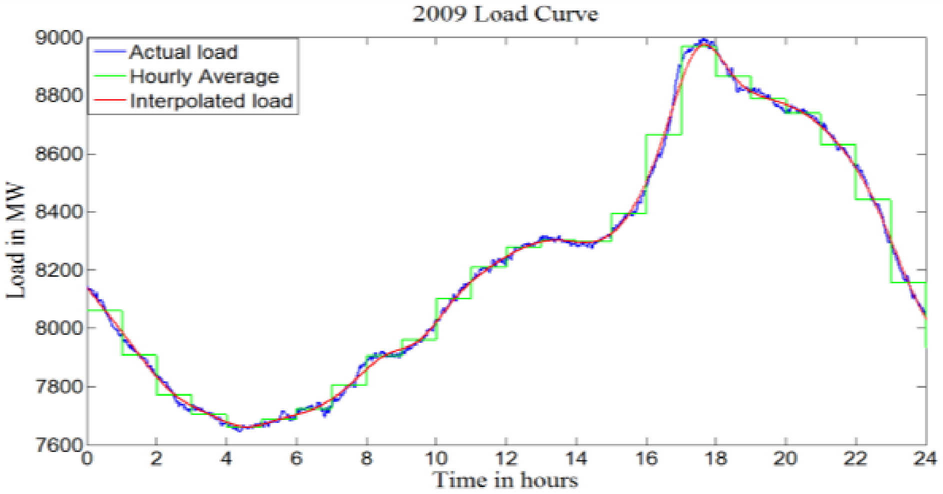


Figure 16. Actual load, hourly average and interpolated load for 2009.

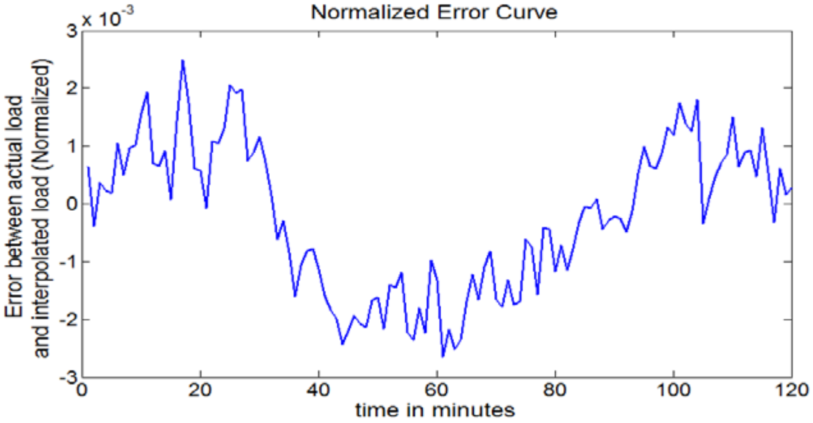


Figure 17. Error between interpolated load curve and actual load curve (normalized by peak load).

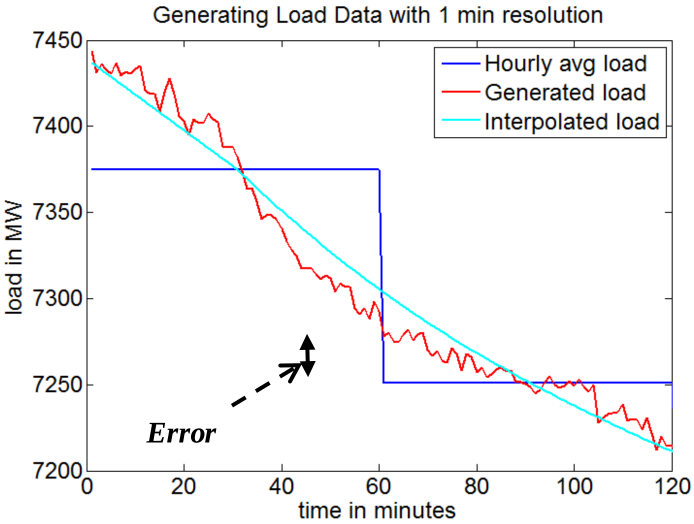


Figure 18. Imposing the load variance in 2009 to the interpolated load in 2020.

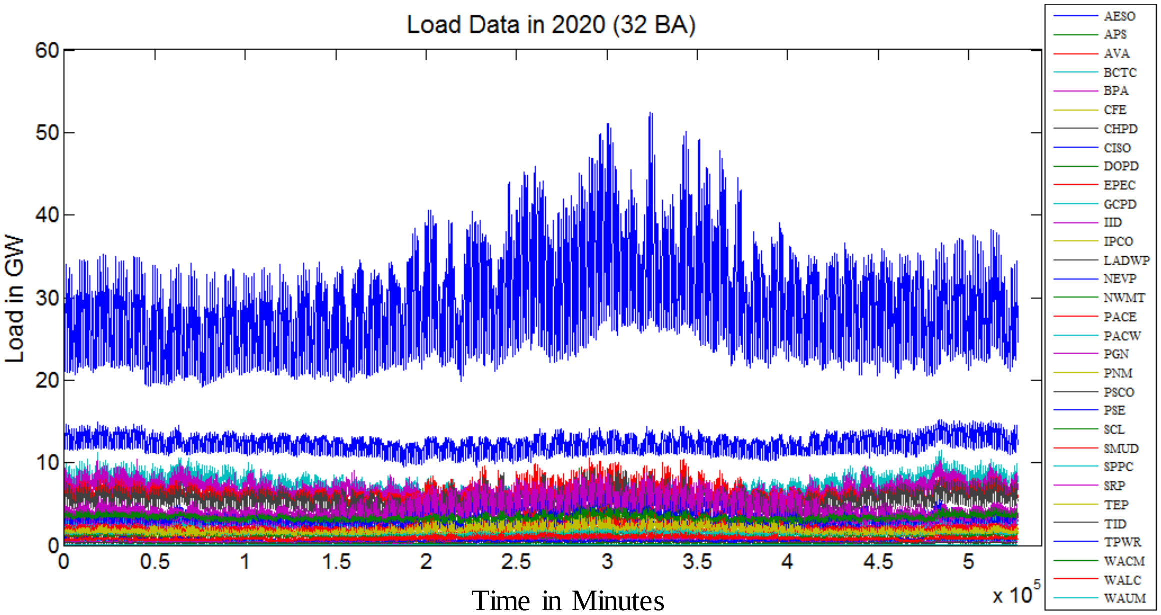


Figure 19. Simulated load data for 2020.

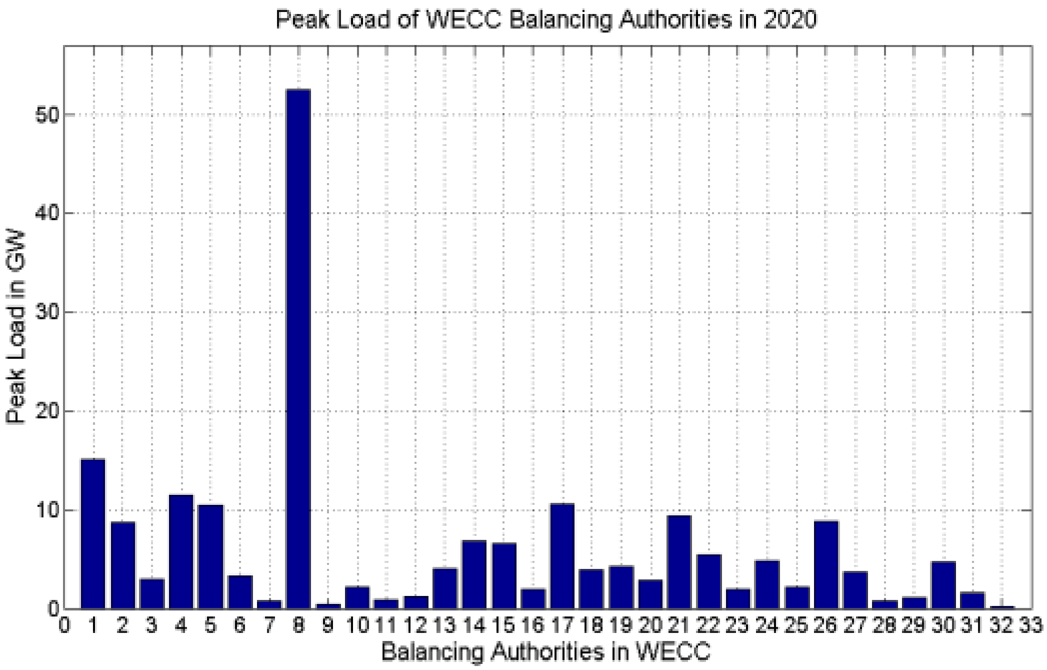


Figure 20. Peak load of different Balancing Authorities in the “WECC TEPPC 2020” case.

Table 2. 32 Balancing Authorities in WECC

No.	BA Name	No.	BA Name	No.	BA Name	No.	BA Name
1	'AESO'	9	'DOPD'	17	'PACE'	25	'SPPC'
2	'APS'	10	'EPEC'	18	'PACW'	26	'SRP'
3	'AVA'	11	'GCPD'	19	'PGN'	27	'TEP'
4	'BCTC'	12	'IID'	20	'PNM'	28	'TID'
5	'BPA'	13	'IPCO'	21	'PSCO'	29	'TPWR'
6	'CFE'	14	'LDWP'	22	'PSE'	30	'WACM'
7	'CHPD'	15	'NEVP'	23	'SCL'	31	'WALC'
8	'CAISO'	16	'NWMT'	24	'BANC'	32	'WAUM'

4.1.3 Wind and Solar Data

The projected wind and solar data for 2020 were taken from the 15% renewables penetration case a percentage of WECC energy demand as defined by the WECC Transmission Expansion Planning Policy Committee (TEPPC) [4]. Time series data for wind and solar production were generated based on 2006 weather models. The TEPPC 2020 PC0 case assumes that all Renewable Portfolio Standards (RPSs) in 2020 are met with the level of the RPSs in 2020 being derived based on assuming a linear progression for those RPSs that have target dates later than 2020. Using this approach yielded a WECC-wide (including BAs in Mexico and Canada) RPS of approximately a 15.5% renewables penetration. This 15.5% renewables penetration level was met by wind (8%), solar (3%), geothermal (2.4%), biomass (1.3%) and small hydropower (0.77%).

The installed wind capacity for each BA in the original TEPPC PC0 case is shown in Figure 21. The installed solar capacity for each BA, which includes both PV and concentrated solar Power (CSP) facilities (with 6-hours storage), is shown in Figure 22.

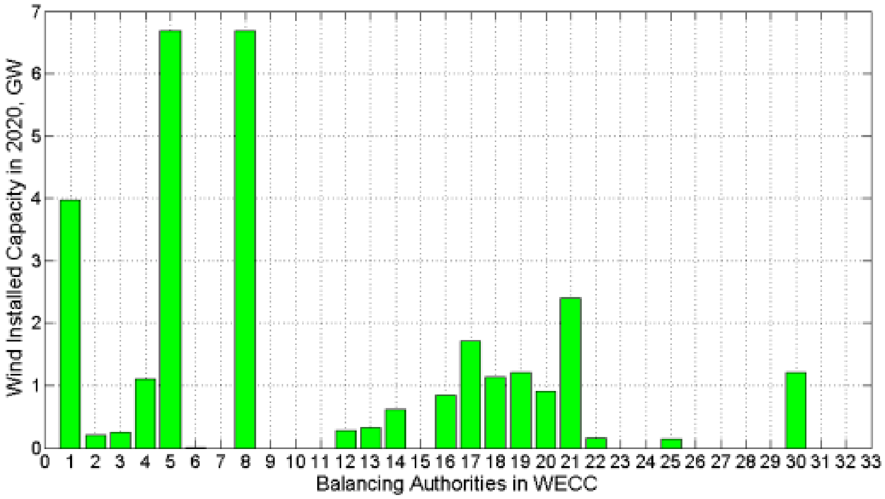


Figure 21. Installed wind capacities in 2020 for WECC BAs.

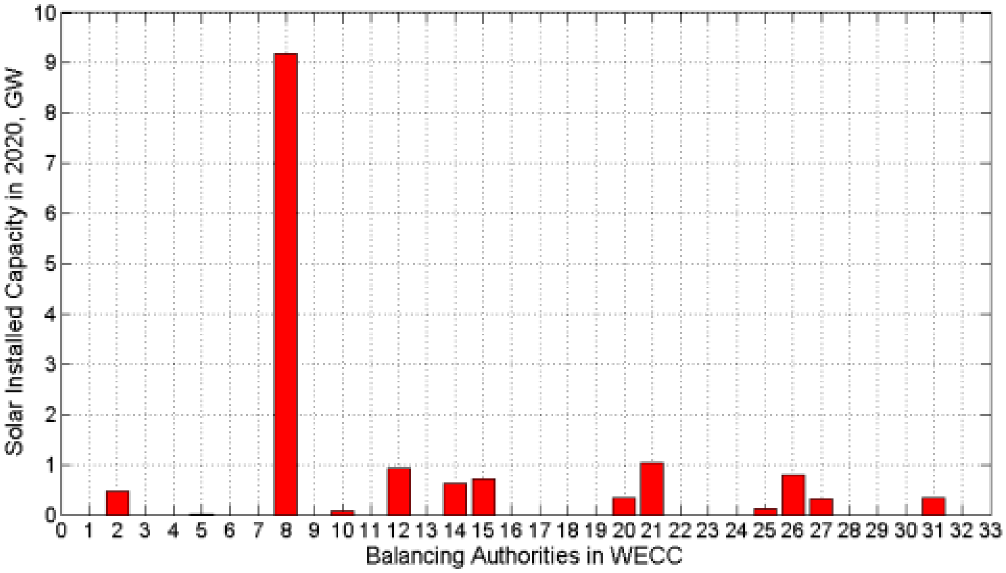


Figure 22. Installed solar capacities in 2020 for WECC BAs.

The wind generation data are based on western wind data sets that were developed by 3Tier [5] and used by the National Renewable Energy Laboratory (NREL). These wind data have a 10-minute resolution; 3Tier, Inc., provided a methodology for deriving 1-minute interpolations. Day-ahead forecasts for hourly wind generation are also available from the same data sets.

PNNL used and improved upon the wind generation data by disaggregating the wind profiles at the bus level. With this refined data, PNNL was able to generate day-ahead forecast and actual 10-min average for wind at the bus level and more realistically simulate the diversity of wind generation.

The solar generation data were obtained from NREL. These data are based on hourly, satellite-derived data from the State University of New York–Clean Power Research and a statistical model that synthesized the subhourly variations [6]. Power production data were developed for multiple solar technologies, including 50-MW fixed photovoltaics, 50-MW one-axis tracking photovoltaics, and 100-MW concentrating solar power plants with assumed 6-hour thermal energy storage. These data were developed for 1,488 grid locations that correspond to the Western Renewable Energy Zones.

Day-ahead and hour-ahead forecasts for solar generation were developed by PNNL using a methodology that quantifies forecast-error based on a Clearness Index [3], [7]. A simplified definition of a clearness index *CI* at any time T in a day is the ratio between actual solar energy production at time T divided by the maximum possible solar energy production at time T in the same day.

4.2 References

[1] N. A. Samaan, R. Bayless, M. Symonds, T. B. Nguyen, C. Jin, D. Wu, R. Diao, Y. V. Makarov, L. D. Kannberg, T. Guo, S. Dennison-Leonard , M. Goodenough, R. Schellberg, S. Conger, K. Harris, M. Rarity, S. Wallace, J. Austin, R. Noteboom, T. Van Blaricom , K. McRunnel, J. Apperson, M. Empey, P. V. Etingov, D. Warady, R. Brush, J. Newkirk, P. Williams, M. Landauer, H Owen, W. Morter, K. Haraguchi, J. Portouw, Downey, S. Sorey, S. Williams, T. Gossa, C. Kalich, P. Damiano, C. Macarthur, T. Martin, J. Hoerner, S. Knudsen, A. Johnson, R. Link, and D. Holcomb, ".Analysis of Benefits of an Energy Imbalance Market in the NWPP ." PNNL-22877, Pacific Northwest National Laboratory, Richland, WA, 2013. Available: http://www.nwpp.org/documents/MC-Public/NWPP_EIM_Final_Report_10_18_2013.pdf

[2] Western Electricity Coordinating Council. 2010. *Assumptions Matrix for the 2020 TEPPC Dataset.* [Online.] Accessed August 15, 2013, at http://www.wecc.biz/library/StudyReport/Documents/Assumptions%20Matrix%20for%20the%202020%20TEPPC%20Dataset.pdf

[3] Samaan N, M Milligan, YV Makarov, M Hunsaker, T Nguyen, C Jin, R Diao and R Hafen. *Balancing Authority Cooperation Concepts to Reduce Variable Generation Integration Costs in the Western Interconnection: Balancing Authorities Consolidation and Sharing of Balancing Reserves*. Final Project Report, Pacific Northwest National Laboratory, Richland, Washington. To be published in October 2014.

[4] Nickell BM. 2008. *TEPPC Renewable Energy Cases*. [Online.] Accessed August 15, 2013, at http://www.wecc.biz/committees/StandingCommittees/PCC/LRS/Shared%20Documents/Wind%20Profiles/Renewable%20Energy%20Generation%20Paper.pdf

[5] 3TIER. 2010. *Development of Regional Wind Resource and Wind Plant Output Datasets*. NREL/SR-550-47676. Work performed by 3TIER for National Renewable Energy Laboratory, Golden, CO. [Online.] Accessed August 15, 2013, at http://www.nrel.gov/docs/fy10osti/47676.pdf.

[6] Orwig K, M Hummon, BM Hodge and D Lew. 2011. *Solar Data Inputs for Integration and Transmission Planning Studies*. NREL/PO-5500-52985. National Renewable Energy Laboratory, Golden, CO.

[7] Ma J, YV Makarov, C Loutan, and Z Xie. 2011. “Impact of wind and solar generation on the California ISO’s intra-hour balancing needs.” *IEEE 2011 Power and Energy Society General Meeting*, pp.1-6, 24–29, July 2011. DOI: 10.1109/PES.2011.6039410.

5.0 Reduction of Uncertainty in Forecast Errors for Load and Wind Generation Using Integrated ARIMA Forecasting, Signal Decomposition, and Principal Component Analysis

Our approach based on separation of slow and fast power system motions (represented as net load variations = load variations - wind variations – solar variations) and associated uncertainties using probabilistic methods helps to improve predictability and reduce uncertainty in the system. Slower motions of stochastic processes are more predictable due to more autocorrelation between the subsequent points. By employing this autocorrelation, it is possible to improve the forecast of parameters subject to uncertainty. By quantifying uncertainty around the improved forecast, it is possible to reduce the range of uncertainty.

5.1 Connection of Uncertainty with System Motions

A power system maintains balance between its generation, changing load, variable generation, and interchange. Increasing uncertainty means more balancing work to be done by conventional generation, energy storage, and controllable load. In other words, these resources have to move more with increasing uncertainty and variability. Because the balancing service is a paid service, more system motion results in additional costs. The statistical separation of fast and slow motions helps to better distribute the balancing job between slow and fast balancing resources; so that they are more efficiently used and the cost is reduced (the fast-responsive resources are usually more expensive). In this study, we apply various approaches to address the challenges in forecast error prediction and uncertainty reduction by: (1) adding seasonal effects to preserve seasonal patterns; (2) adding Cholesky decomposition to honor cross-correlation structure in the multivariate series; (3) decomposing signals into trend, seasonal, and noise components, each of which is more appropriate for time series forecasting (e.g., autoregressive integrated moving average ARIMA); (4) applying principal component analysis (PCA) to extract the dominant patterns across zones and zonal similarity for uncertainty reduction.

5.2 Generation of Forecast Errors Using Seasonal ARIMA Models

The load and wind error forecast generator generates multiple error forecasts while preserving statistical properties and characteristics from real data such as mean, variance, autocorrelation, seasonal behavior, and cross-correlation between different zones or buses. For the load and wind data described in the previous section, a seasonal autoregressive integrated moving average (ARIMA) model is fitted to each set of data. For the load data, the time-of-day biases are also extracted and preserved. R Libraries ‘RJDBC’ and ‘forecast’ are required for the analysis.

5.2.1 Data Preprocessing

Given the actual and forecast values, the forecast error is defined as

$$err = actual - forecast. \quad (1)$$

For the load forecast error data, the mean value for each hour of each day in a particular month (time-of-day bias) is computed. Therefore, each month has 24 means with each mean corresponding to an hour of the day. The bias is subtracted from error. Figure 23 shows the time-of-day bias for the load forecast

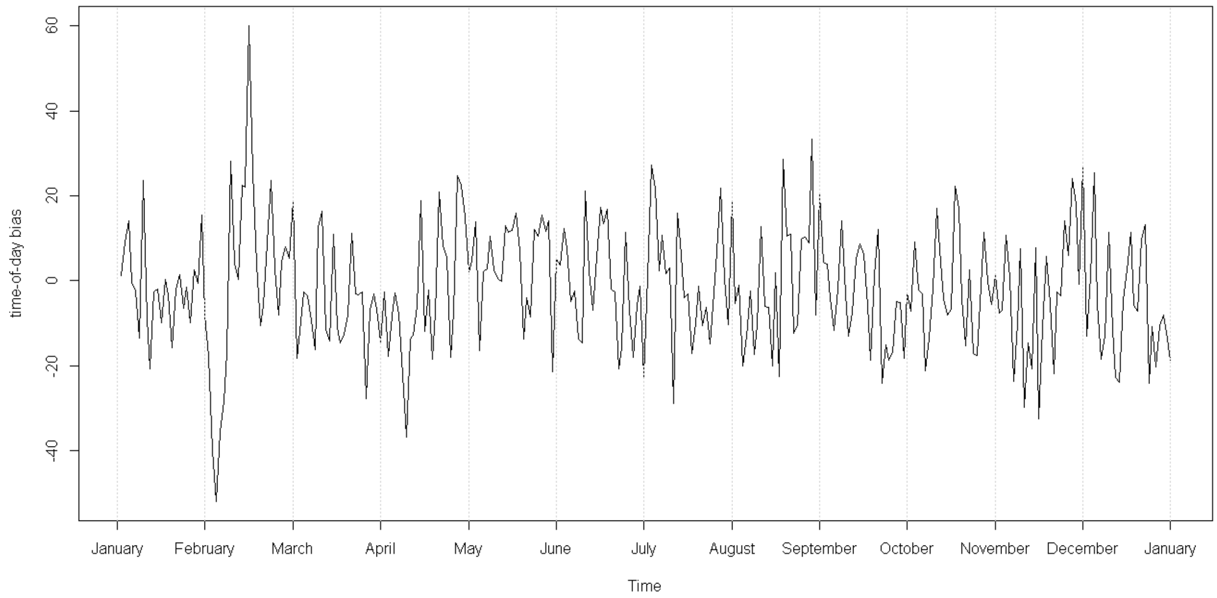


Figure 23.Time-of-day bias for the load forecast error data.

5.2.2 Fitting Seasonal ARIMA Models

The auto.arima function in the 'forecast' library is used to fit a seasonal ARIMA model to the unbiased error. A seasonal ARIMA model is defined by parameters, which is denoted by $ARIMA(p,d,q) \times (P,D,Q)_s$. The general detailed model can be given as

$$\left(1-\sum_{i=1}^{p}\phi_i B^i\right)\left(1-\sum_{j=1}^{P}\Phi_j B^{js}\right)(1-B)^d(1-B^s)^D y_t = c + \left(1-\sum_{i=1}^{q}\theta_i B^i\right)\left(1-\sum_{j=1}^{Q}\Theta_j B^{js}\right)\varepsilon_t, \qquad (2)$$

where B is the lag operator, i.e., $By_t = y_{t-1}$, y_t is the model output at time t, and the noise terms ε_t are independent, identically distributed (i.i.d.) random variables sampled from a normal distribution with zero mean.

This function finds the number of parameters and identifies their values using stepwise fitting. The model is selected based on the Akaike and Bayesian information criteria (AIC and BIC), which are standard measures for statistical model selection.

Using the fitted model, a forecast error series can be generated. To complete the generation process, the output of the seasonal ARIMA model is summed with a random sample from the model residuals and the previously computed biases.

In order to maintain the correlation structure across multiple zones/buses, the correlation matrix is computed from the matrix containing each zone/bus's forecast errors as columns. Using the Cholesky decomposition, the correlation weights (the upper triangular factor) can be extracted from the correlation matrix. The correlation weights are then multiplied against the generated forecast errors for that set of zones/buses.

In the case that some of a set of zones/buses are highly correlated, it is not possible to recover the correlation weights from the correlation matrix because of numerical instability. Instead, a subset of buses with lower correlation is extracted. The seasonal ARIMA models are only fitted to the members of the subset and generation of the forecast errors for the remaining buses is done by rescaling the corresponding subset member.

The Oracle database is accessed from R via SQL queries using the 'RJDBC' library. The generated results are saved in csv files that can be loaded into the Oracle database.

5.3 Improved Representation and Reconstruction of Load and Wind Forecast Errors Using Signal Decomposition and Automated ARIMA

The above ARIMA approach developed has the following features: (1) the AR parameters can capture autocorrelations (2) built-in Cholesky decomposition can handle cross-correlation (note that the night and day signals are mixed without decomposition); and (3) the algorithm honors the local mean and global mean information.

Here we integrated the automated-ARIMA with signal decomposition in order to alleviate the non-normality and non-stationarity issues such that ARIMA is more applicable. The forecast error data is decomposed in order to extract its trend (long-term upward or downward movement), seasonal (periodical patterns), and random (residual) components. Instead of using the Cholesky decomposition method derived above, we fixed the low frequency (e.g., trend and seasonal) components to preserve the cross correlation between buses/zones. This is based on the assumption that the low-frequency patterns will mostly repeat from year to year. The model is improved to better preserve the local continuity of the forecast errors. An AR model is trained on the random component for each bus/zone to preserve the continuity pattern. The maximum order for the AR model is restricted to be three to avoid possible over fitting.

5.3.1 Signal Decomposition of Time Series

For each bus/zone, the forecast error time series is decomposed to three components—trend, seasonal and random series. Figure 24 shows the 10 minute forecast error data for bus 10116 during February 2020. Figure 25 - Figure 27 show the 3 component series extracted from the forecast error data.

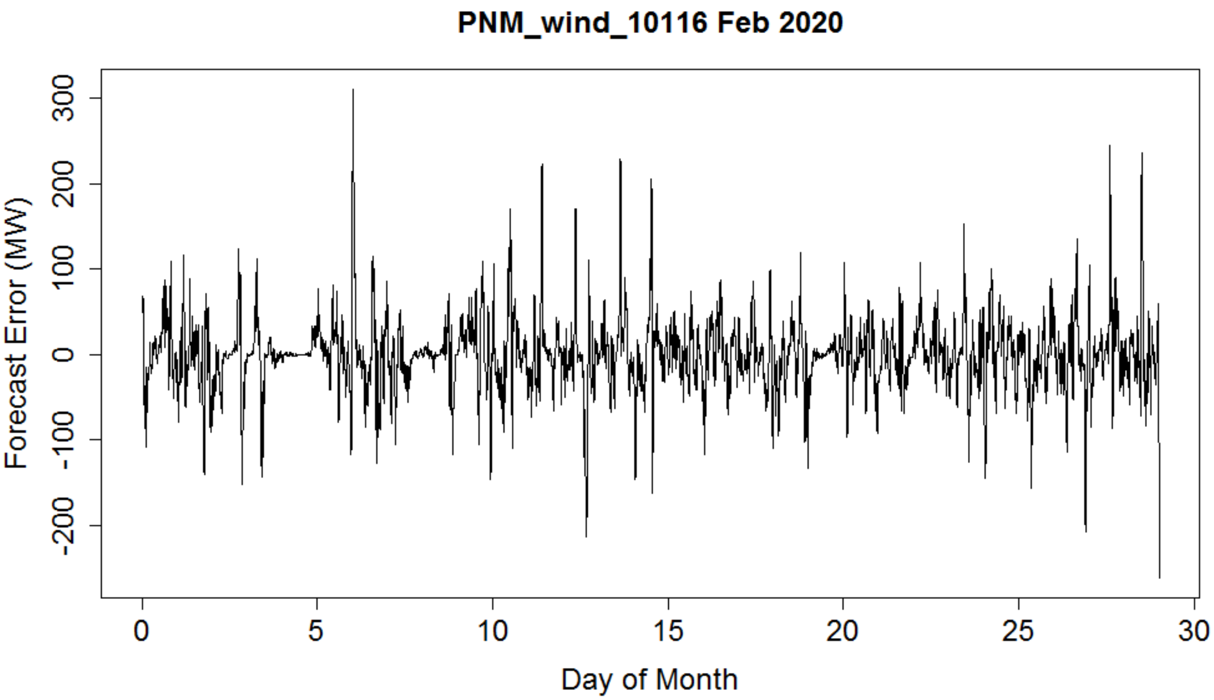


Figure 24. Wind forecast error for bus 10116 (10 minute resolution) during February 2020.

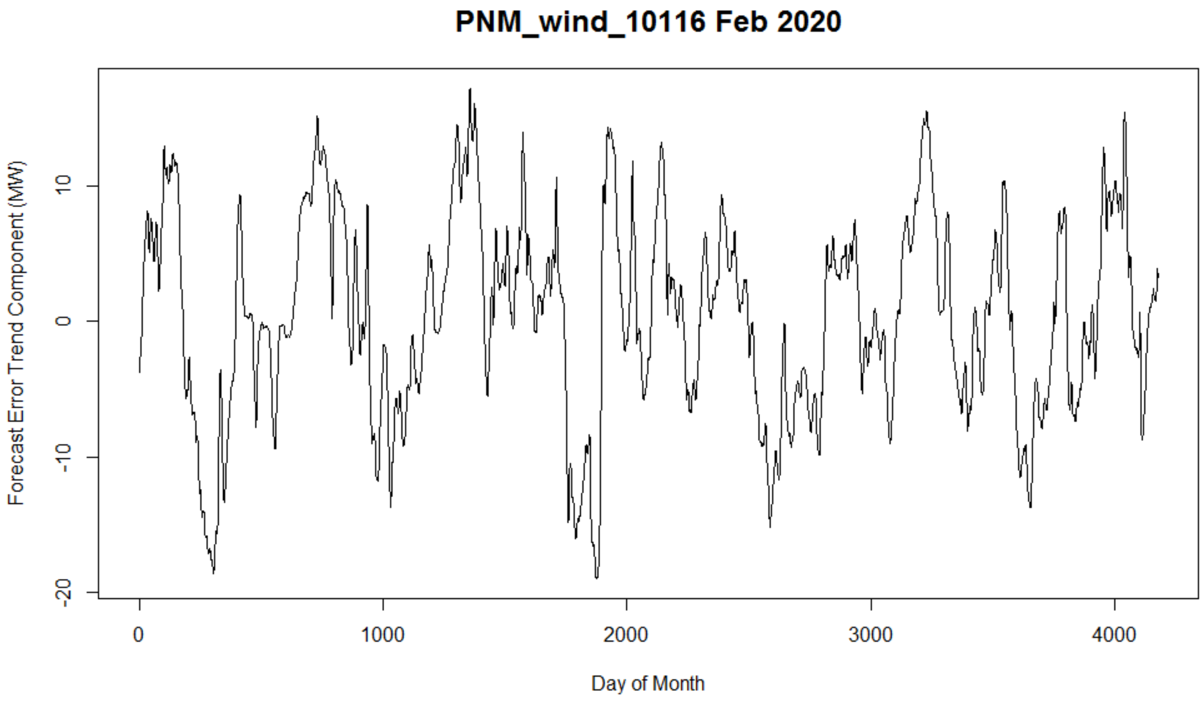


Figure 25. Trend component of wind forecast error for bus 10116 (10 minute resolution) during February 2020.

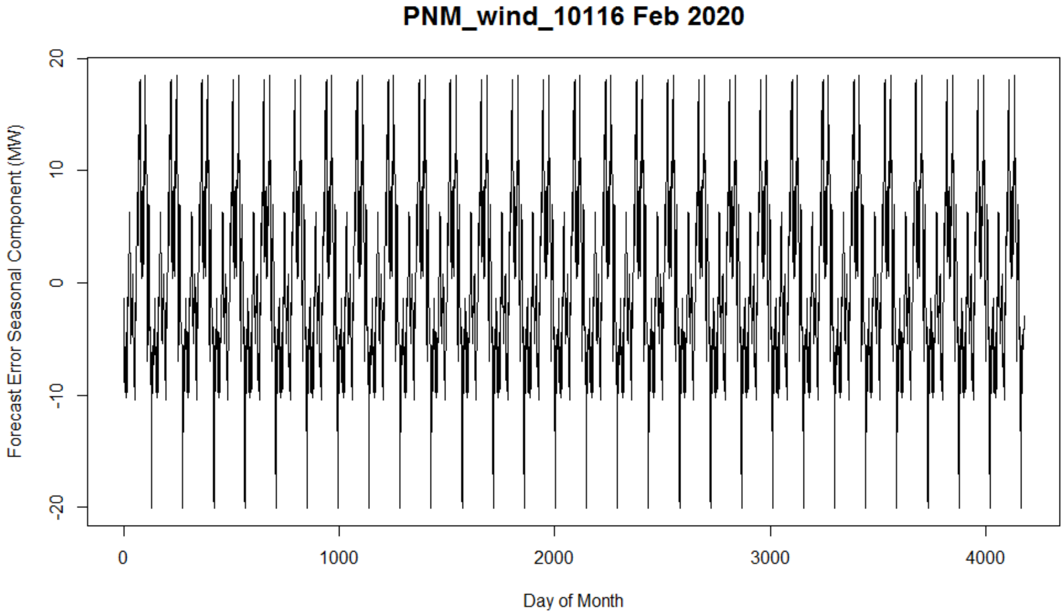


Figure 26. Seasonal component of wind forecast error for bus 10116 (10 minute resolution) during February 2020.

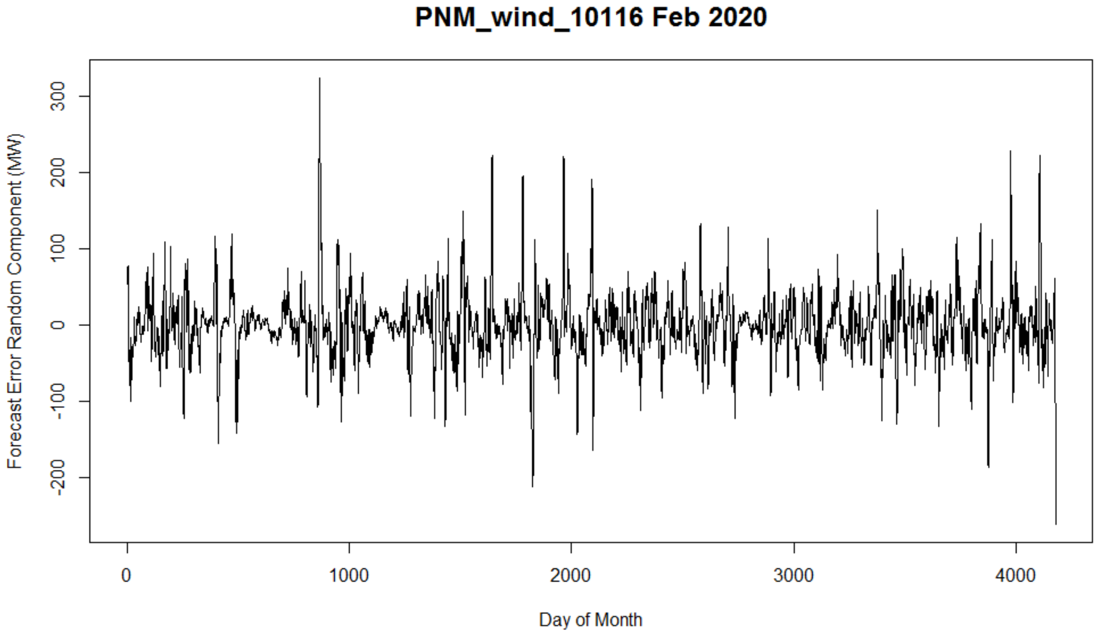


Figure 27. Random component of wind forecast error for bus 10116 (10 minute resolution) during February 2020.

5.3.2 Forecast Error Reconstruction

To generate a new set of forecast errors, an AR model of order 3 or less is selected for the random component of each error series based on AIC values. For example, an AR(3) model was fitted to the trend component series (1-year long) for bus 10116 as follows

$$y_t = 0.9579y_{t-1} - 0.028y_{t-2} - 0.097y_{t-3} + \varepsilon_t, \tag{3}$$

where y_t is the model output at time t, and the noise terms ε_t are independent, identically distributed (i.i.d.) random variables sampled from a normal distribution with zero mean. Figure 28 shows the generated random component from an AR(3) model that was fitted to the actual wind forecast error for bus 10116 during February 2020.

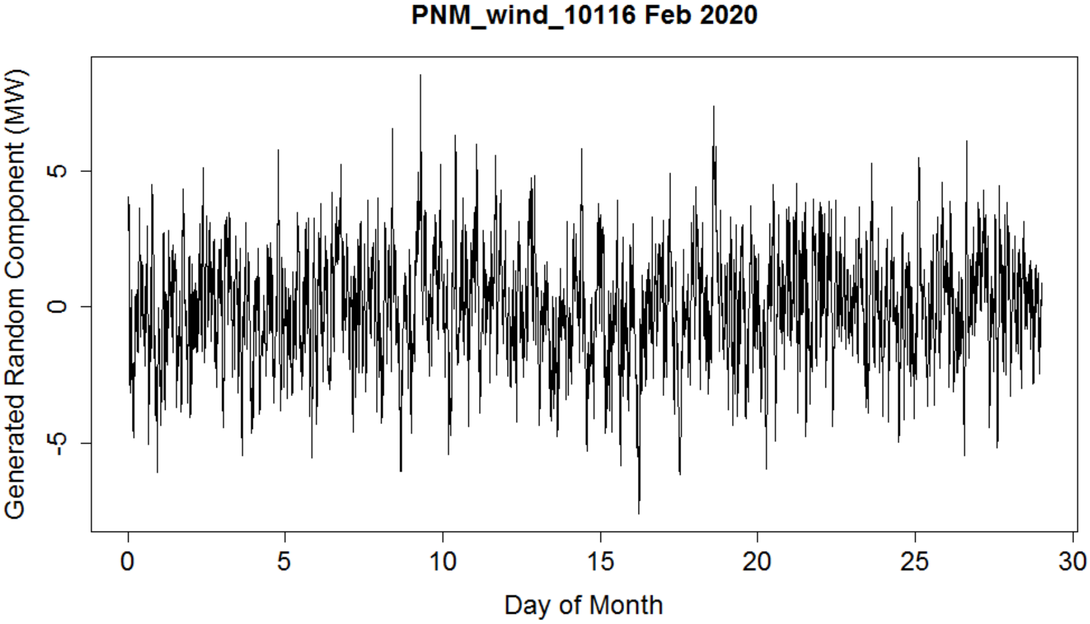


Figure 28. Generated random component from the fitted AR(3) model for wind forecast error at bus 10116 (10 minute resolution) during February 2020.

5.4 Evaluating the Residuals/Errors Using Principal Components to Represent/Reconstruct System Behaviors

It is observed that the load and wind generation and their forecasts are spatially and temporally correlated across the study zones (in our study, the load generation data was collected from 39 zones, and the wind generation data was collected from 190 buses). It is interesting and useful to express the system behavior by extracting the dominant patterns and/or system motion trajectories. In this section we describe tests and results of using a principal component analysis (PCA) approach on load data to evaluate the possibility to characterize the system motion similarity and express the overall behavior in a much lower dimensional subspace than the original system dimension (i.e., 39 in the case study).

PCA uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The first principal component has the largest possible variance and accounts for the most variability in the data, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components.

Here we tested PCA on a one-year-long hourly data (8784 data points) for the 39 zones of load data, and therefore the data matrix is of size 8784 x 39. The systems behave similarly across the zones –Figure 29 and Figure 30 show that the first two principal components dominate the overall variability in the data, and only the first six components are needed to explain more than 90% of the total variability. This provides guidance on how to reduce the system to a lower dimension subspace without introducing too much error (e.g., by setting a 10% error tolerance).

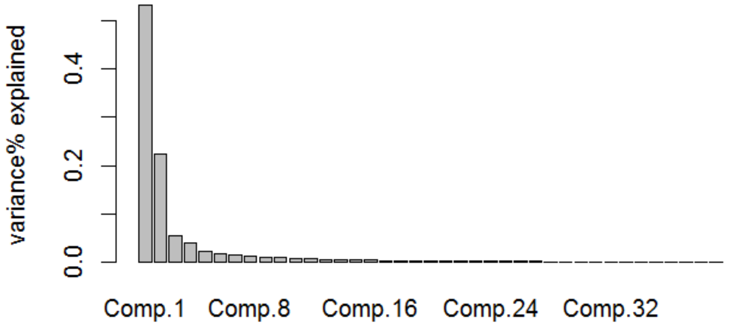


Figure 29. Percentage of total variance in original data that is explained by each principal component.

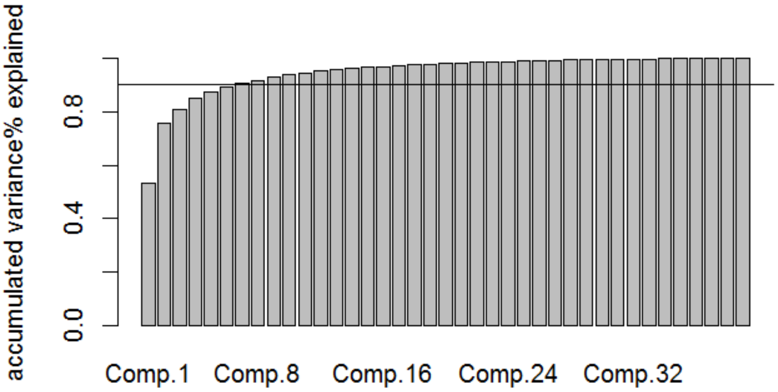


Figure 30. Accumulated percentage of total variance in the original data that is explained by the principal components.

Next, similar analyses were performed on incremental data to study the system movement behavior. Figure 31 and Figure 32 show that the data matrix for the incremental data also has strong correlation and can be reduced through PCA. Furthermore, Figure 31 shows that the first four components dominate the overall variability, but it needs the first 13 components in order to have 90% of total variances explained.

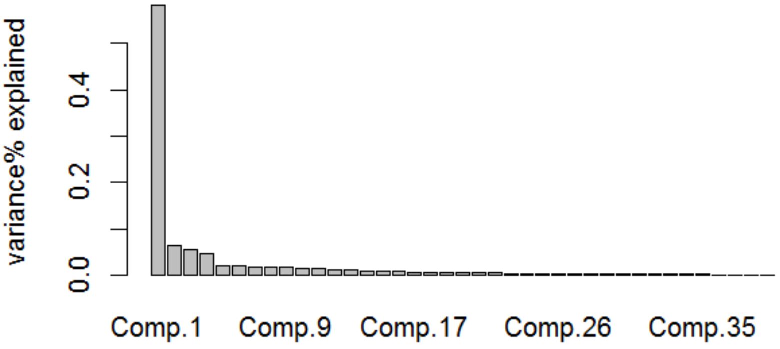


Figure 31. Percentage of total variance in incremental data that is explained by each principal component.

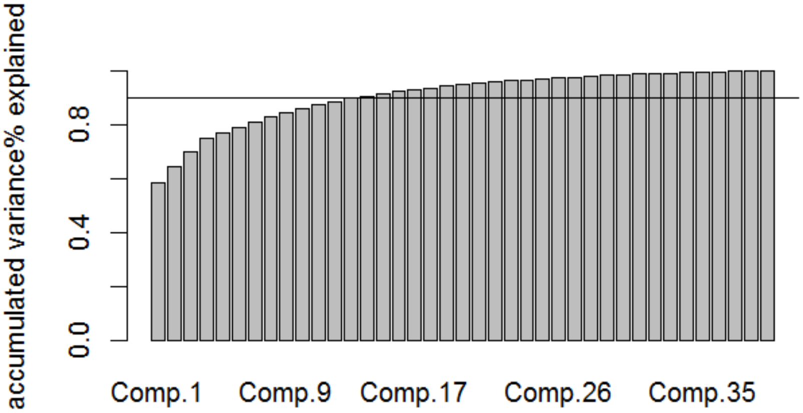


Figure 32. Accumulated percentage of total variance in the incremental data that is explained by the principal components.

The same effects can be seen when data for load and wind are combined (and represented as "net load" data). However, for the areas under study in this case, when wind and load are added together, there is no visible change to the plots, because the results are heavily dominated by the load forecast error, because the wind forecast error is very small by comparison.

5.5 Simulation for Short-Term Forecast Errors

In this section, an algorithm is proposed to simulate short-term load and wind forecast errors to serve as inputs to the TUT. The proposed algorithm is shown for load data, but also applies to net load data (represented as net load variations = load variations – wind variations – solar variations). Given a time instant and 4 hours' forecast error data prior to the time instant (i.e., 24 data points at 10 minute resolution), a number of different possible forecast errors for the next 4 hours (24 data points) will be generated using a sequential Gaussian simulation method. PCA will also be performed on the simulated data to reduce the simulation error. For the two sets of simulation data (without and with PCA), stationarity and normality of the residuals will be tested. The cross-correlation between different zones of the simulation date will be also compared with the historical data.

5.5.1 Sequential Gaussian Simulation

The historic data of load (and wind generation, if desired) forecast errors can be used to train ARIMA parameters as well as cross-correlation parameters among the 39 zones. With these parameters given, however, the hour- or day-ahead predictions of the forecast errors will be fixed values, which will only give one realization of predicted forecast errors. In order to represent the uncertainty in the predictions and also to enable evaluation of system output uncertainty, we adopt the sequential Gaussian simulation approach. At the first step, instead of a single predicted value, we generate multiple (e.g., 200) realizations of the predictions of forecast errors from multivariate Gaussian distributions with ARIMA prediction mean and cross-correlation matrix trained from historical data. The combination of the mean and variance predicted by the ARIMA model with multivariate Gaussian sampling is a new technique developed in this project. The generated predictions (Figure 33) are then treated as ‘actual observations’ and used as additional ‘historical’ data. This is to maintain local continuity as well as autoregressive and cross-correlation patterns in the original forecast errors data.

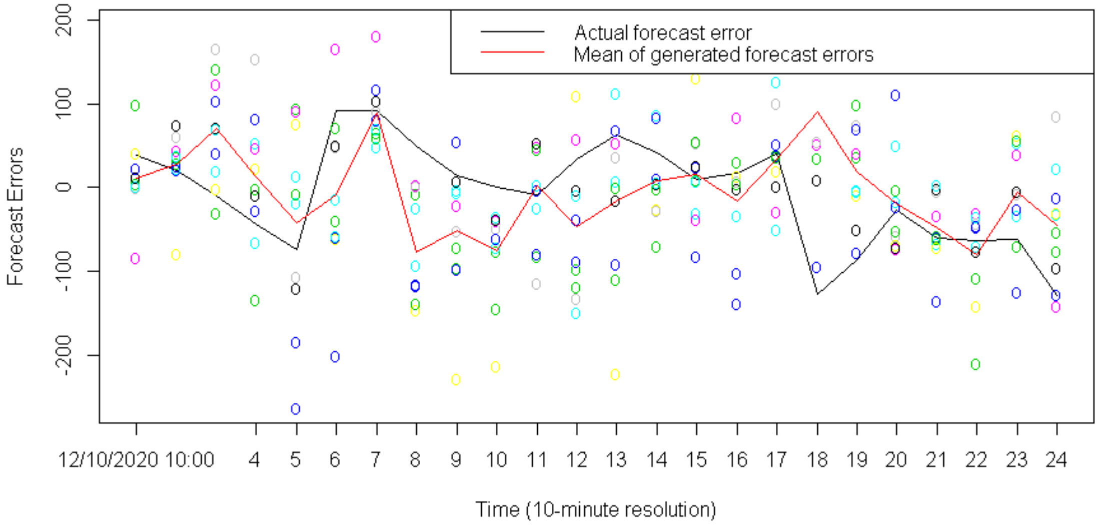


Figure 33. Randomly selected 10 sets of simulated load forecast errors for zone 1. The two curves are different, but their statistical characteristics are identical (mean, variance, autocorrelation pattern).

5.5.2 Error Reduction Using PCA

As demonstrated in section II, PCA can identify the similarities between the forecast errors at all 39 zones (or 190 buses for wind) and enable us to represent the overall variability using a subset number of components (e.g., 13 out of 39 for load forecast errors as in Figure 34, and 18 out of 190 for wind forecast errors as in Figure 35) to represent more than 90% of total variability. It would be interesting to see if we can reduce the uncertainty of short-term predicted forecast errors by combining PCA analysis and the sequential Gaussian simulator used in the previous section.

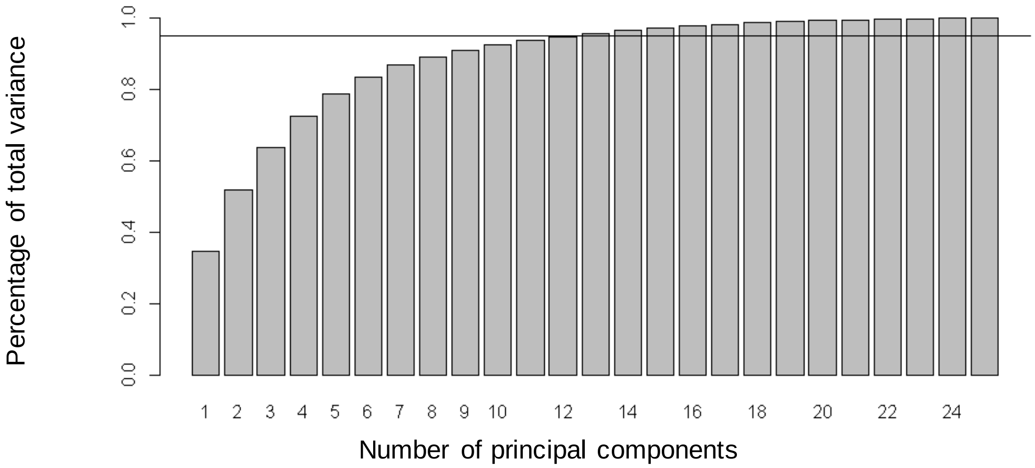


Figure 34. Percentage of total variance explained of actual load forecast errors within 4 hours prior to the given time instant (12/10/2020 10:00).

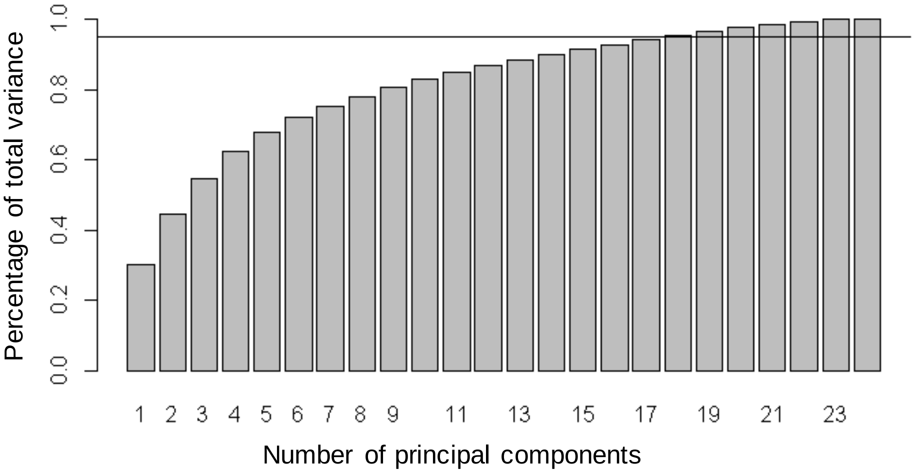


Figure 35. Percentage of total variance explained of actual wind forecast errors within 4 hours prior to the given time instant (12/10/2020 10:00).

Mathematically speaking, at any given time instant, the actual load forecast error data (10 minute resolution) for all 39 zones within 4 hours prior to the time instant can form a 24x39 dimensional matrix A. Each column of A is the time series of 4 hours' forecast errors for one zone. Each row of A corresponds to one time instant. Apply PCA to A, and then we have

$$A = U\Sigma V^T = SV^T, \tag{4}$$

where $U \in \mathbb{R}^{24\times 24}$, and $V \in \mathbb{R}^{39\times 39}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{24\times 39}$ is a rectangular diagonal matrix with nonnegative singular values on the diagonal. Therefore, the columns of V are the normalized principal components that give a new basis of the full space. The columns of the *score matrix* $S = U\Sigma$ are the coefficient vectors of the 24 records (39-dimensional points) in the new coordinate system.

The directions of the leading principal components represent the main cross-correlation between different zones. The forecast error can be reduced by projecting the full space to the subspace spanned by the leading principal components. In our study, we used the leading 14 principal components because they explained more than 90% of the total variance as shown in Figure 34.

For any set of the simulated load forecast error data for the following 4 hours after the given time instant, the data, denoted by $\tilde{A} \in \mathbb{R}^{24\times 39}$, has the same format as A. The score matrix of $\tilde{A}$ in the coordinate system given by V satisfies $\tilde{A} = \tilde{S}V^T$, which leads to $\tilde{S} = \tilde{A}V$. Therefore, the projection of $\tilde{A}$ onto the 14-dimensional subspace is

$$\tilde{A}_s = \tilde{S}(:,1:14) \cdot V(:,1:14)^T, \tag{5}$$

where $X(:,1:14)$ represents the first 14 columns of a matrix X.

In Figure 36, the 10 sets of simulated load forecast errors shown in Figure 33 were projected to the 14-dimensional subspace, and the projected time series for zone 1 were plotted. It can be seen that the samples are less spread out than the original simulated data. This section represents the different realizations from PCA ARIMA predictions rather than uncertainty reduction itself. In this section we develop a method to reconstruct processes with certain probabilistic characteristics. Section 7 will discuss the improvements depending on the number of principal components considered.

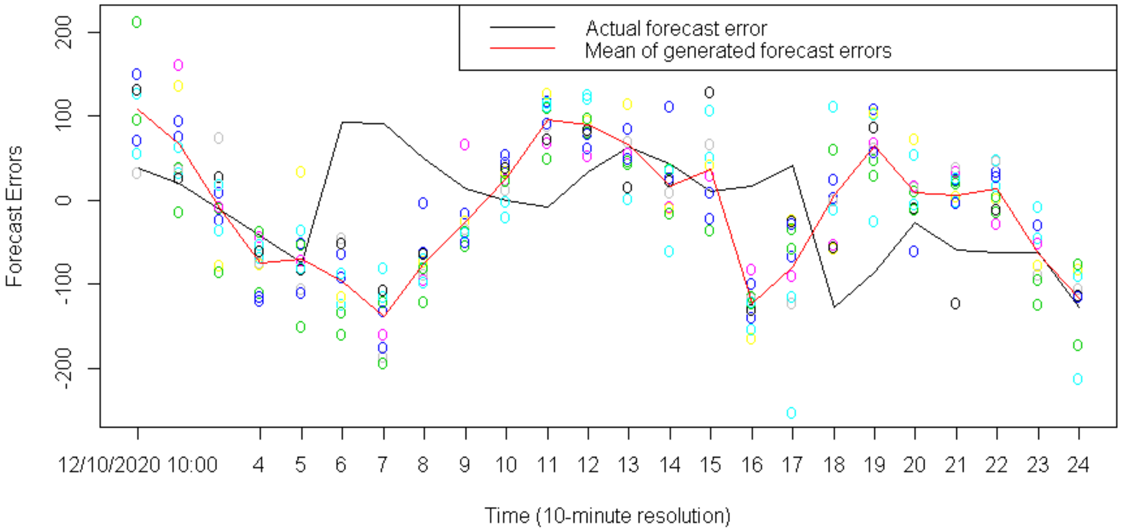


Figure 36. Randomly selected 10 sets of simulated load forecast errors for zone 1 using PCA.

5.5.3 Tests on Statistical Properties

To check whether the combined approach yields reasonable results, stationarity and normality tests are performed on the short-term predictions of forecast errors (Figure 37-Figure 42) [1]. In order to check the statistical properties of the analyzed signal (the load data) in terms of stationarity and normality, two well-accepted methods were used: The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test of stationarity [2] and the Shapiro-Wilk test of normality [3].

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was conducted to check the stationarity of the generated time series of forecast errors. The null hypothesis that a time series is stationary will be rejected if the test P-value is smaller than a threshold (e.g., 0.05).

The Shapiro-Wilk test of normality was performed to test the normality of the generated short-term predictions of forecast errors. Similar to the KPSS test, the null hypothesis that the test time series is normal will be rejected if the test P-value is smaller than a threshold (e.g., 0.05).

Figure 37 - Figure 42 show that most of the generated time series for load passed the stationarity and normality tests.

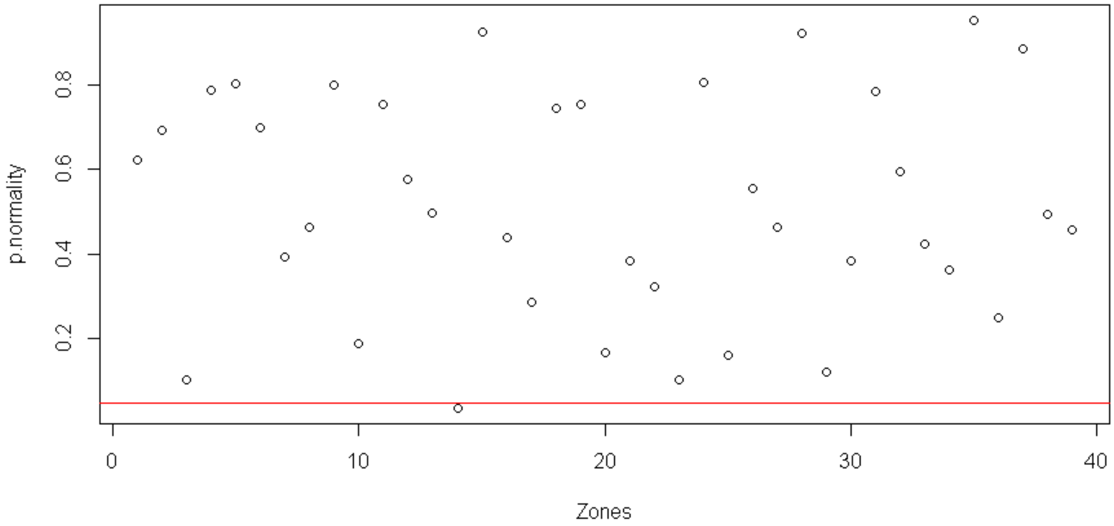


Figure 37. Normality test on the 1st set of simulation data for all zones.

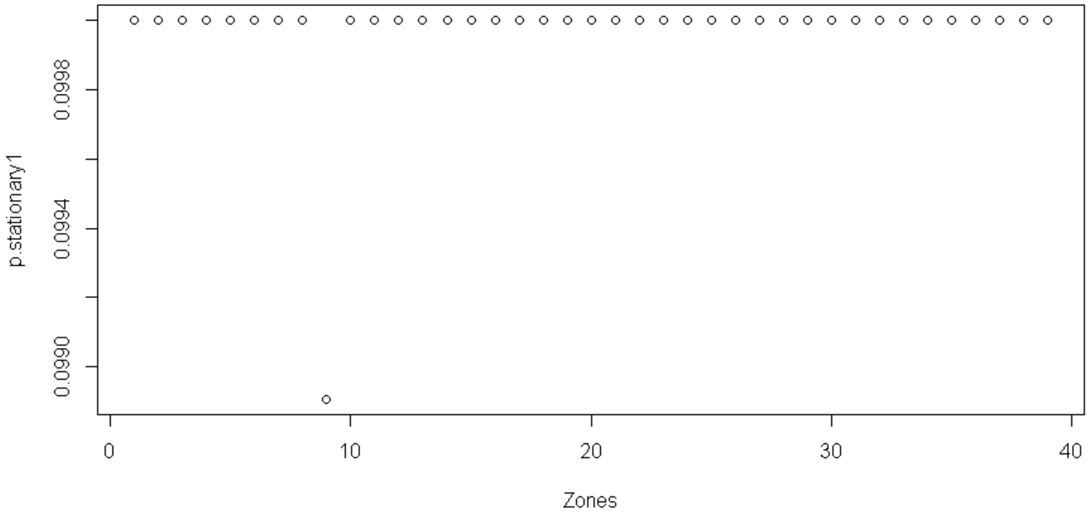


Figure 38. Stationarity test on the 1st set of simulation data for all zones.

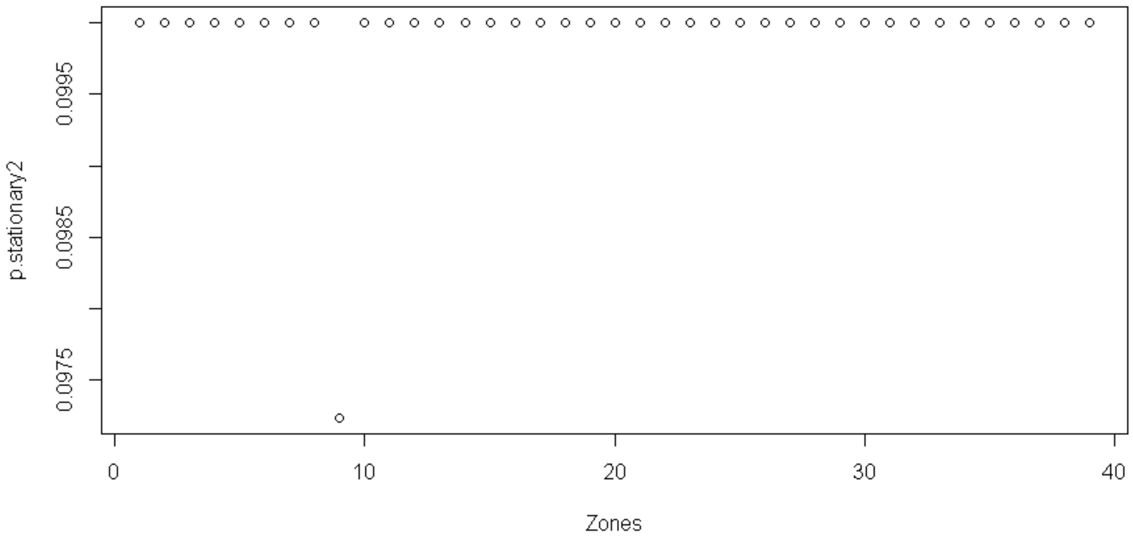


Figure 39. Stationarity test on the residuals of the 1st set of simulation data for all zones.

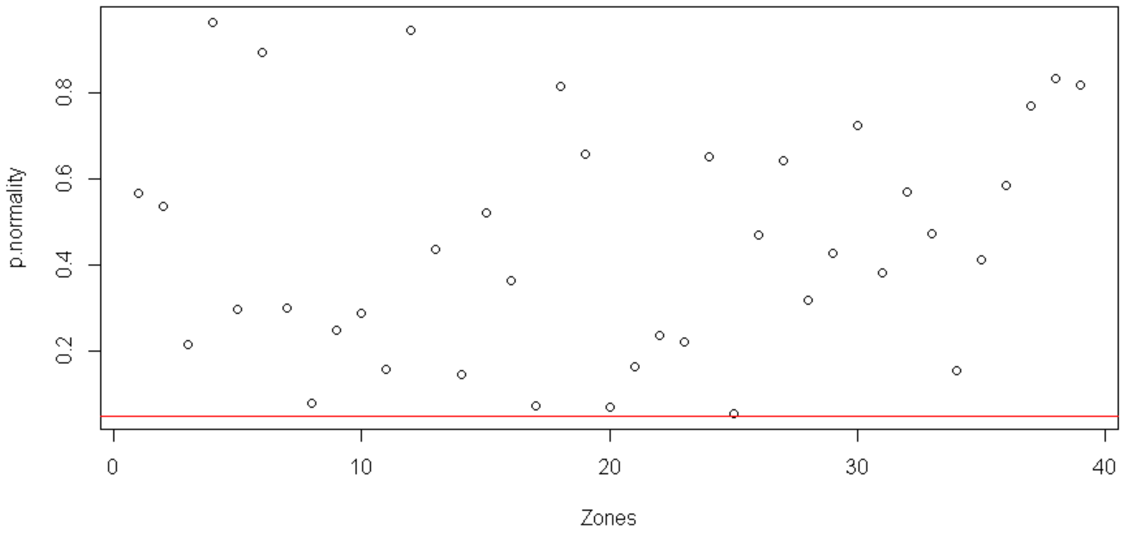


Figure 40. Normality test on the 1st set of simulation data for all zones using the PCA approach.

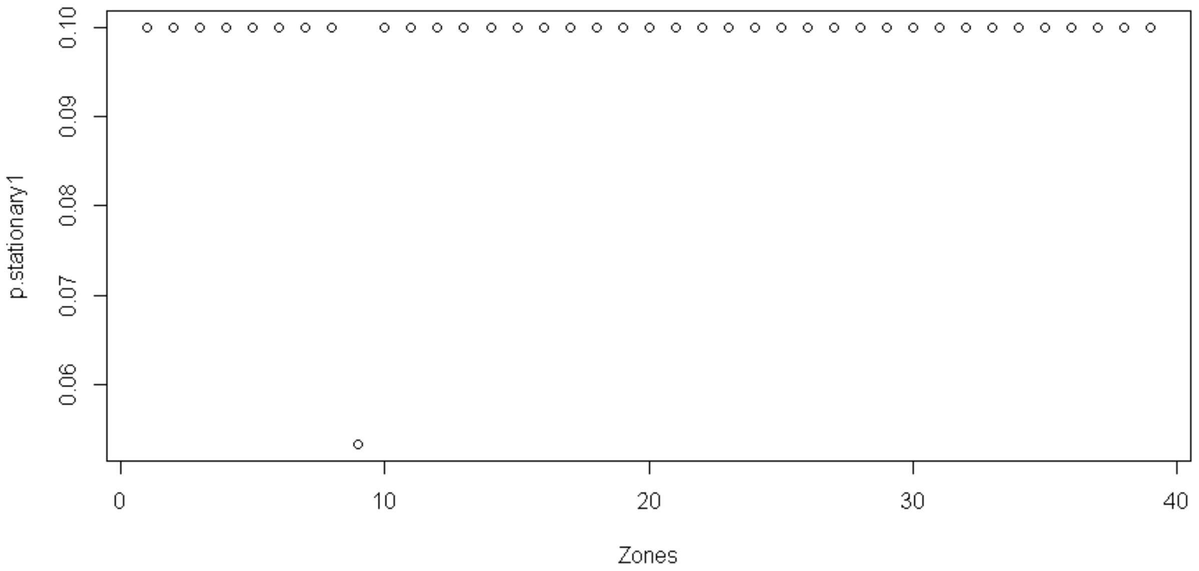


Figure 41. Normality test on the residuals of the 1st set of simulation data for all zones using the PCA approach.

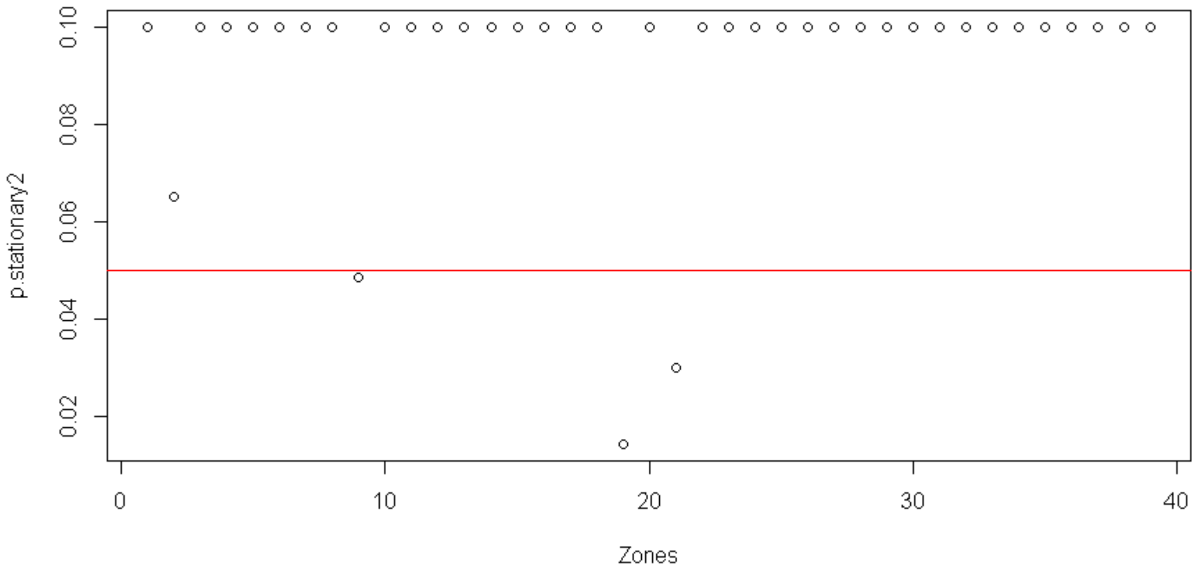


Figure 42. Stationarity test on the 1st set of simulation data for all zones using the PCA approach

Except for a few zones, the generated predictions of the forecast errors are more likely to be stationary and normally distributed, using sequential Gaussian simulation, with or without PCA implementation.

The cross-correlation between the zones is tested on both the original simulated forecast errors and the data obtained using PCA. Figure 43 shows the correlation matrices for the first 15 zones of the two sets of forecast errors. The names of the zones are listed in Table 3. A green color disk at the (i, j) position in the correlation matrix means that zone i and zone j are positively correlated. The stronger the correlation is, the skinnier the disk is, and the darker the color is. It can be seen that with the PCA approach, the correlation matrix is adjusted, resulting in enhancement/reduction of a certain zone’s correlation with all the other zones. This adjustment is useful when the number of zones is relatively large, and models for separate regional clusters are needed. The correlation matrix can give insight on how to cluster the zones. In addition, the correlation between the zones within a cluster is enhanced, while the correlation between different clusters is reduced.

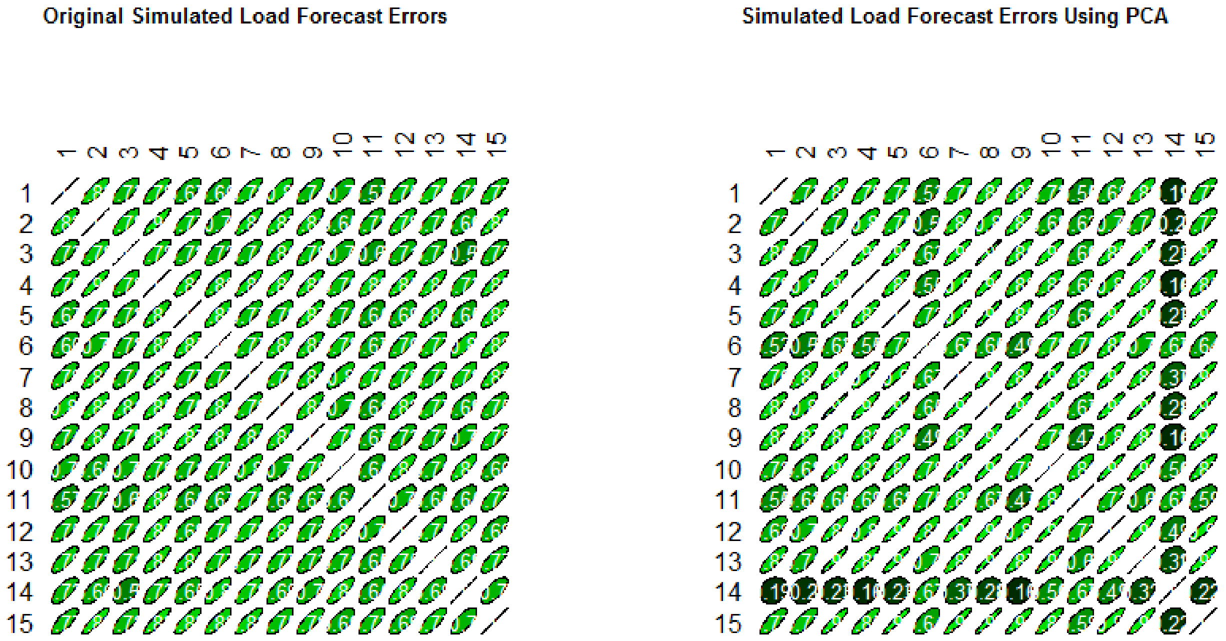


Figure 43. Correlation matrices for simulated load forecast errors, without and with using PCA

Table 3. Name of zones

1 AESO	11 PACE_ID	21 SPP	31 SCL
2 BCTC	12 PG&E_VLY	22 CFE	32 APS
3 PG&E_BAY	13 PGN	23 CHPD	33 TREAS
4 PNM	14 TPWR	24 DOPD	34 LDWP
5 WALC	15 WAUW	25 FAR EAST	35 PACW
6 MAGIC	16 AVA	26 NWMT	36 SRP
7 PSE	17 EPE	27 IID	37 TEP
8 SCE	18 GCPD	28 NEVP	38 TIDC
9 SMUD	19 PSC	29 PACE_UT	39 WACM
10 BPA	20 SDGE	30 PACE_WY	

The effects and results in this subsection based on load data alone can also be seen when data for load and wind are combined (and represented as “net load” data). However, for the areas under study in this case, when wind and load are added together, there is no visible change to the plots, because the results are heavily dominated by the load forecast error, because the wind forecast error is very small by comparison.

5.6 References

[1] Patrick Royston (1982) An extension of Shapiro and Wilk's W test for normality to large samples. Applied Statistics, 31, 115–124.

[2] Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". Biometrika 52 (3–4): 591–611. doi:10.1093/biomet/52.3-4.591. JSTOR 2333709. MR 205384. p. 593

[3] Patrick Royston (1982) An extension of Shapiro and Wilk's W test for normality to large samples. Applied Statistics, 31, 115–124.

6.0 Improving System Predictability Using Statistical Separation of Slow, Fast and Very Fast System Motions

Motions and dynamic behavior of a power system can be effectively characterized by system net load as well as dynamic events, where net load is the difference between actual system load and renewable generation. Based on recorded time-domain trajectories, one can identify the actual motions of the power system and provide actual stress directions that can be applied to the system models and calculate security margins. Load and renewable generation forecasts in near term and longer term are usually provided separately by service vendors, to predict future system motions for better balancing the system.

In this project, advanced statistical methods were applied to separate slow, fast and very fast system motions caused by variation of the net system load. The very fast motion is actually white noise, which is a completely unpredictable component of a signal. If the forecast error is exactly white noise, this is a perfect forecast.

$$\textit{Actual net load} = \textit{load forecast} + \textit{load forecast error} - (\textit{renewable forecast} + \textit{renewable forecast error}) = \textit{net load forecast} + \textit{net load forecast error} \qquad (6)$$

Accurate forecasts are critical for real-time operations and long term planning for power utilities and independent system operators. However, very few of them actually analyze the realistic forecast error signals in a balancing authority to better understand their motions and then find means of reducing forecast errors.

A new concept is proposed in this paper to decompose the actual system net load into three different components obtained from balancing authorities: (1) the first component refers to hourly, daily, and even longer term forecast, reflecting the system slow motion; (2) the second component is the estimate of forecast error that captures the fast motion of system movement in near term (they remove most of longer term period patterns and preserve relatively fast signals); and (3) the third component is the residual representing very fast motion that can be characterized as noises. The main idea is illustrated in Equation (7), where the second and third components constitute the net load forecast error shown in Equation (6). Separating these components and estimating the net load forecast error make it possible to reduce forecast errors and system uncertainties, resulting in many benefits for the power system planners and operators. To achieve this goal, badging decision trees [3], an ensemble method that builds multiple decision trees by repeatedly resampling training data with replacement and voting the trees for a consensus prediction, are used to effectively estimate the net load forecast errors, which leaves the main uncertainties of net load with the residuals. The wavelet transform method is used for analyzing our forecast errors and forecasted forecast errors to evaluate the performance of our approach for separating the system fast and very fast motions.

$$\textit{Actual net load} = \textit{net load forecast (slow motion)} + \textit{estimate of net load forecast error (fast motion)} + \textit{residuals (very fast motion)} \qquad (7)$$

In this section, two methodologies are introduced and described. Case studies and results are shown and discussed.

6.1 Methodology description

Two methodologies used in this work are introduced in this section. One is the predictive model for generating load, load forecast errors and net load forecast errors to separate the above mentioned system motions; another one is the wavelet analysis method to study these motions in the time-frequency domain and to evaluate the separation results for reducing uncertainties.

6.1.1 A Regression Tree Based Predictive Model

The decision tree technique is an effective supervised data mining tool to solve the classification problems in high data dimensions [4]. For a created database consisting of different cases that are represented by a vector of predictors (or variables) along with an objective, a decision tree is designed for successful classifications of this objective by using only a small number of these predictors.

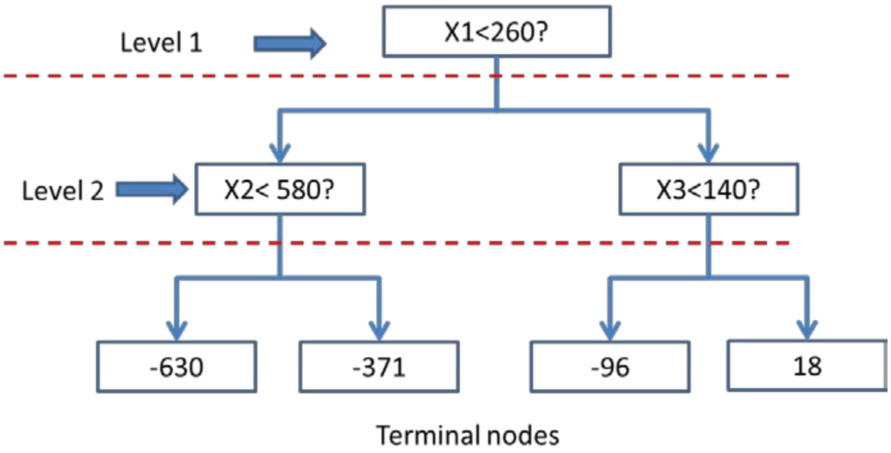


Figure 44. Example for a regression tree.

Regression tree (Figure 44) based methods are widely utilized for day-ahead load forecasts [5]. In this paper, we implemented the regression trees method for generating day-ahead load forecasts, and developed an effective knowledge base for training an auto-regression model for hour-ahead load forecast. Two knowledge bases were developed to forecast hour-ahead load forecast and hour-ahead net load forecast, respectively. The selected attributes for each knowledge base are listed as below:

1) For hour-ahead load forecast:
    - Load data
    - Temperature data
    - Day of week
    - Net load from the same hour the same day in the previous week
    - Net load from the same hour in the previous day
    - Net load from the previous two hours

2) For hour-ahead net load forecast:
    - Load data
    - Wind data

- Temperature data
- Wind speed
- Day of week
- Net load from the same hour the same day in the previous week
- Net load from the same hour in the previous day
- Net load from the previous two hours

To improve the performance of the predictive models using regression tree methods, all the available information prior to the operating hour being forecasted are included in the knowledge base for training the model. Thus, the size of the knowledge bases is increasing with time. This ensures the predictive models be trained using the most recent as well as historical information.

6.1.2 An Advanced Method for Analyzing Fast and Slow Motions

Wavelet transforms (WT) [1-2] are multi-resolution decompositions, which can be used to analyze signals and images. Wavelet transforms describe a signal by the power at each scale and position [6]. The wavelet power spectrum can be calculated at each level/frequency to demonstrate the contribution of the component to the overall variability of a signal. Therefore, the WT tells which components are important in the frequency domain. Wavelet transforms help identify these instances and reflect the changing magnitude of the components [7].

6.2 Case Studies

We performed two types of studies: Case A is to generate load forecast to separate the slow motion and fast motion; Case B is to predict forecast errors to reduce uncertainties and separate fast and very fast motions. In the Case B, there are two scenarios. One predicted load forecast errors while the other one predicted net load forecast errors. The net load forecast errors included both load and wind components.

In these studies, the realistic load and wind data in the year 2010 from two balancing authorities in the Western Interconnection of the U.S are used. The corresponding weather data, e.g., temperature profiles at major cities, is collected from National Oceanic and Atmospheric Administration website, which are used for training the predictive models.

6.2.1 Load Forecast

We collected 2010 hourly historical load and weather data of a WECC balancing authority to perform two load forecast tests based on the regression trees method: day-ahead load forecast and hour-ahead load forecast. We trained the model by using historical data of the month before the testing month (Fix model). The day-ahead forecast for July 2010 is shown in Figure 45.

The upper panel in Figure 45 shows a comparison between actual load and the generated forecast load. There are large discrepancies during several weekdays after the July 4th holiday. The residuals (actual - forecast) are the forecast errors as shown in the lower panel in Figure 45.

The predictive model is trained to use the historical data before every hour being forecasted and include more predictor variables such as data from the previous hour. The hour-ahead forecasts in July

2010 using regression trees are shown in Figure 45. The comparison between actual load and the forecasted load is given in the upper panel in Figure 46. There is no obvious difference between the two signals. The residuals, as shown in the lower panel in Figure 46, are very small.

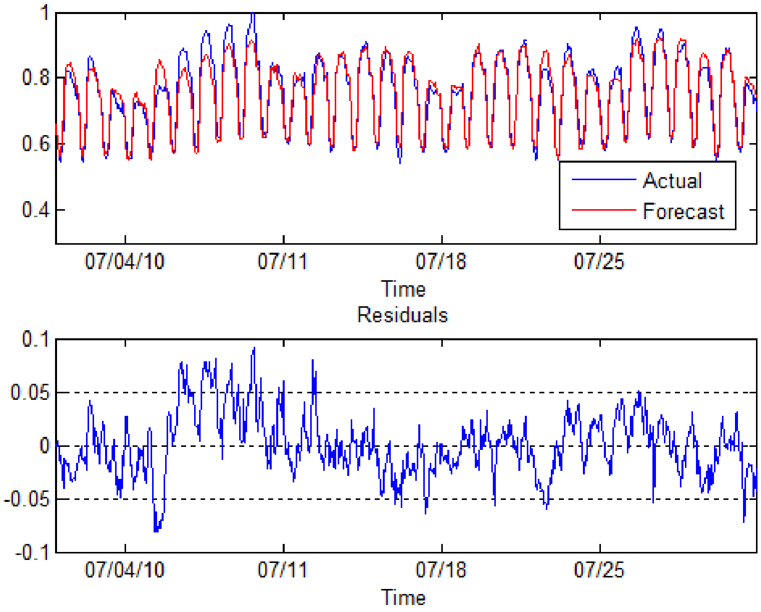


Figure 45. Day-ahead load forecast.
(note: data normalized by the peak load in 2010)

The results demonstrate that our approach can reasonably well separate the slow motion (i.e., load forecast) from the fast motion (i.e., forecast errors), as shown in Figure 45 and Figure 46.

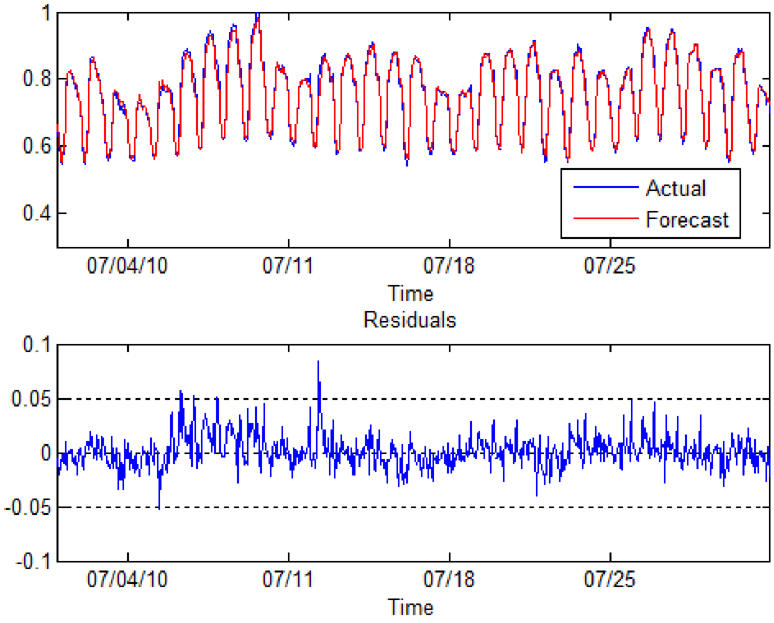


Figure 46. Hour-ahead load forecast.

(Note: Data normalized by the peak load in 2010.)

6.2.2 Forecast Errors Prediction

In this project we attempted to predict the forecast error based on state-of-the-art prediction methods. The ability to partially predict forecast errors would help to reduce the forecast error itself as well as ultimately reduce the prediction interval around the errors. Such a task is very challenging. Our results show that this, in fact, is feasible. We successful demonstrate that that is possible to reduce the forecast error if the signal has autocorrelation or cross-correlation with other signals.

In this study, historical hourly actual and forecast load data, and weather data of the year 2010 is used in the first scenario. In the second scenario, another set of 2010 hourly historical actual and forecast load and wind data, and weather data of another WECC balancing authority is collected. In this case, historical net load forecast errors are obtained. The regression trees method is utilized to perform hour-ahead forecast of the forecast errors.

6.2.2.1 Load Forecast Errors

The actual load forecast errors were based on historical data. We attempted to reduce the uncertainty of the load forecast errors based on the prediction methods described in section 5.

The predicted load forecast errors for July 2010 are shown in Figure 47. The upper part of Figure 47 presents the actual forecast error versus the predicted forecast error obtained by our method. The difference between the actual forecast error and the predicted forecast error is shown in the lower part of Figure 47. By comparing the actual forecast error and the new forecast error obtained by subtracting the predicted error from the actual error, one can see that the standard deviation of the error is reduced by~30%. The absolute mean error is reduced by 48%.

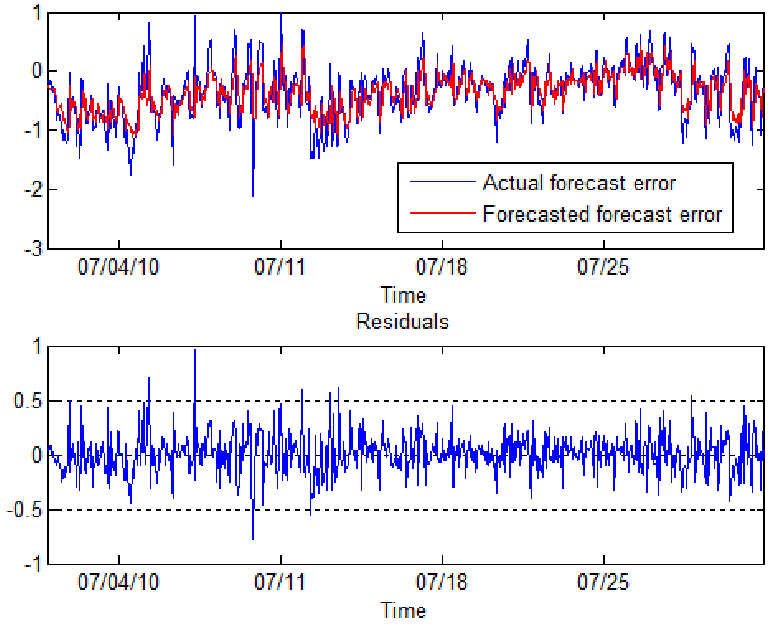


Figure 47. Load forecasted forecast error.
(Note: Data normalized by the maximum actual forecast error.)

6.2.2.2 Net load forecast errors

In the second scenario, the hour-ahead forecast of net load (load minus wind and solar generation) forecast errors is performed. The forecast results of net load forecast errors for July 2010 are shown in Figure 48. The upper part of Figure 48 presents the original forecast error and the predicted forecast error obtained by our method. The difference between the original forecast error and the predicted forecast error are shown in the lower part of Figure 48. In this case the reduced forecast errors (which are the residuals) can be obtained by subtracting the predicted forecast errors from the actual forecast errors.

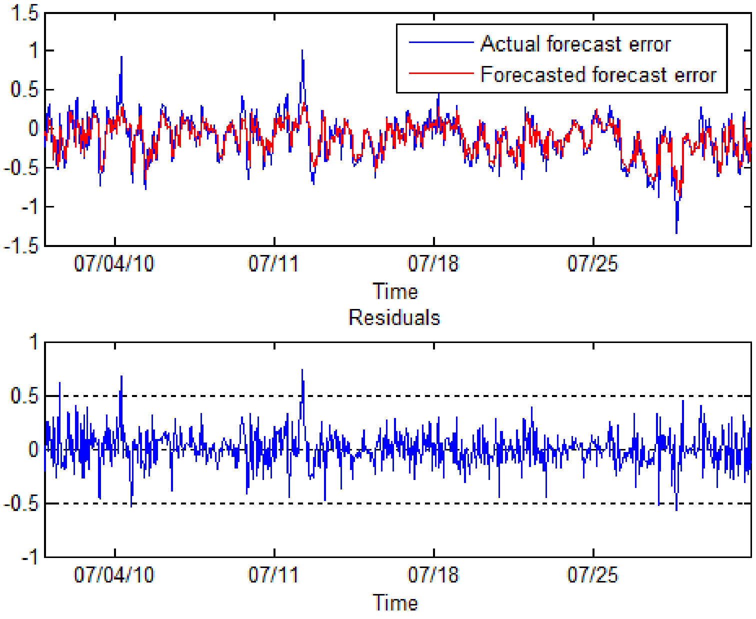


Figure 48. Net load original forecast error and preducted predicted forecast error.
(Note: Data normalized by the maximum actual forecast error.)

Comparing the actual forecast error and the reduced forecast error, the standard deviation is reduced to about 58.6%. The absolute mean error is reduced to 51.4%. The error distribution and absolute error distribution are shown in Figure 49.

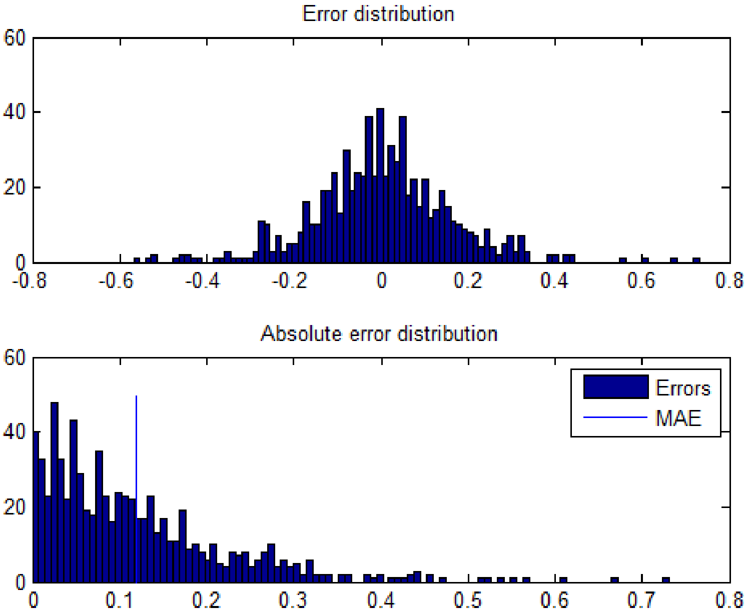


Figure 49. Error distribution and absolute error distribution (MAE: mean absolute error)

6.3 Wavelet Analysis

In this section, we analyzed the residuals using the wavelet analysis method.

The net load, net load forecast, and the forecast errors, reflect a mixture of multi-timescale variability, and the time series usually are of non-normal distribution with non-stationarity. Therefore, traditional statistical measures (mean, standard deviation) are inadequate to evaluate or validate the more prediction results. Here we adopt a wavelet decomposition approach to provide a time–frequency representation as a view of the corresponding time series variability and describe the signal in terms of wavelet functions of time and frequency.

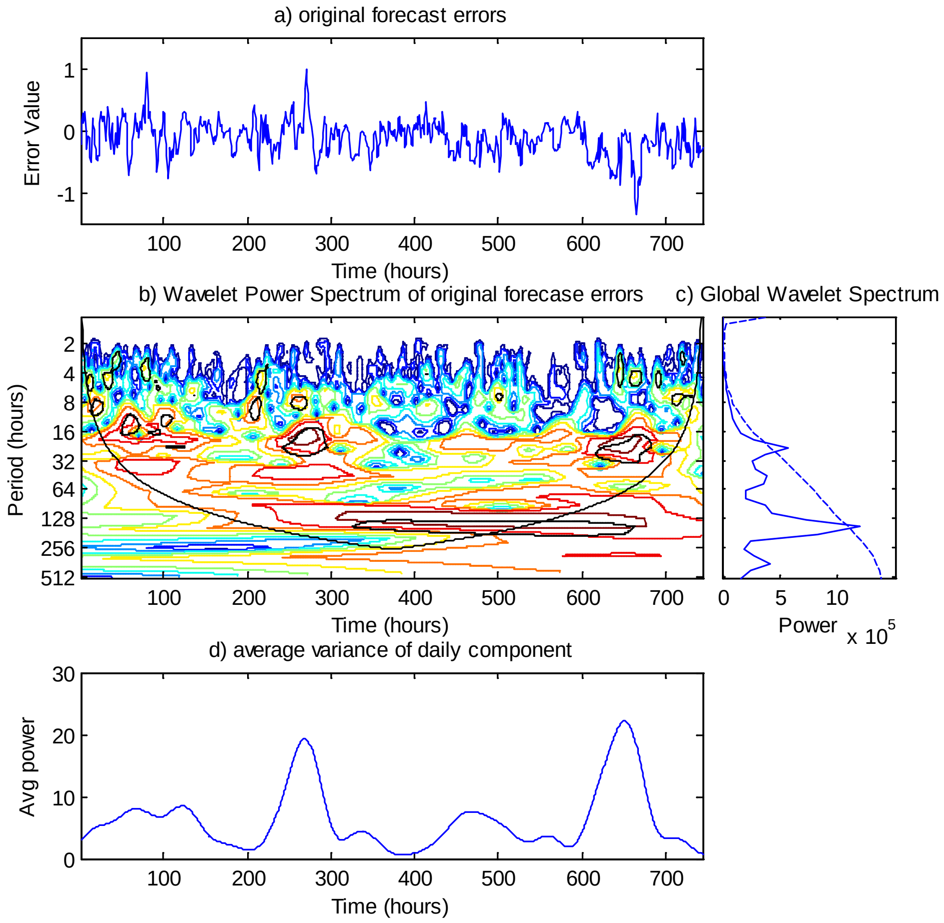
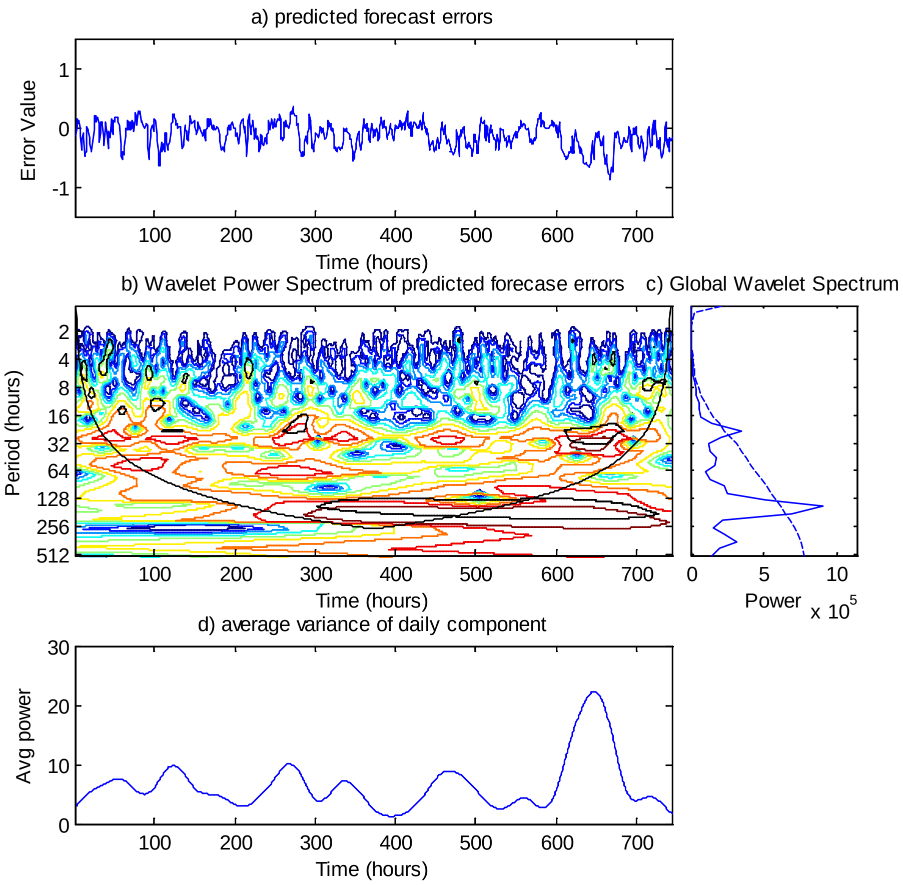


Figure 50. Wavelet analyses of net load forecast errors:

(a) Time series of net load forecast errors; b) Contour map of wavelet power of decomposed components at different time scales; c) The spectrum shows that data consists of a few major components (weekly + daily); d) The contribution of the short-term component is high during the middle and late times.



a) predicted forecast errors

b) Wavelet Power Spectrum of predicted forecase errors

c) Global Wavelet Spectrum

d) average variance of daily component

Figure 51. Wavelet analyses of predictions of new load forecast errors

(a) Predicted forecast errors; b) Contour map of wavelet power of decomposed components at different time scales; c) The spectrum shows that data consists of a few major components; d) The contribution of a short-term component is high during the late time period.

The predicted power spectrum of the predictions (Figure 51c) matches the spectrum of the original forecast errors (Figure 50c) very well, with a strong weekly component and a clear daily component.

Analysis of the residuals shows that the predictions have well captured the main signal in the forecast error series. The spectrum of the residuals is rather evenly distributed. There is still some but weak daily component remained in the residuals, while the weekly component is negligible (well captured). Regarding the mismatched daily component, the biggest mismatches occur around t = 270 hours (Figure 52d), which corresponds to the time of strong positive anomalies in the forecast errors.

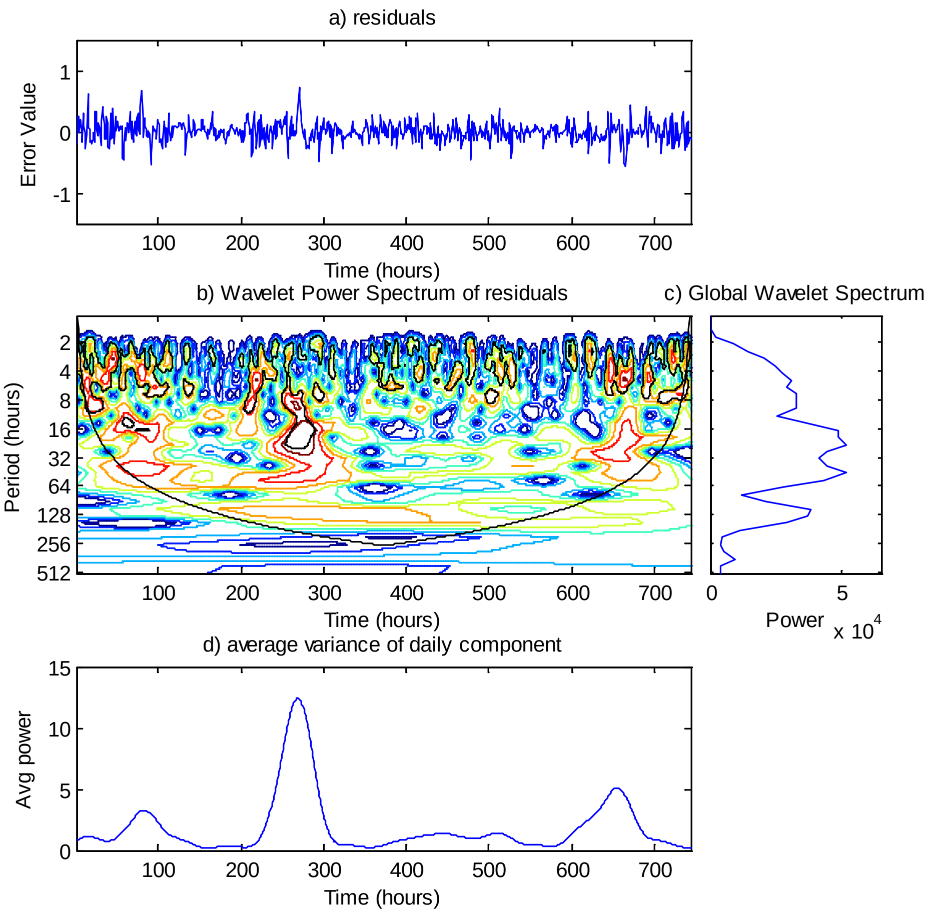


Figure 52. Wavelet analyses of residuals with net load forecast error prediction

(a) Residuals; b) Contour map of wavelet power of decomposed components at different time scales; c) The spectrum shows that data consists of weak daily component and negligible weekly component; d) The contribution of a short-term component is high during the middle time period.

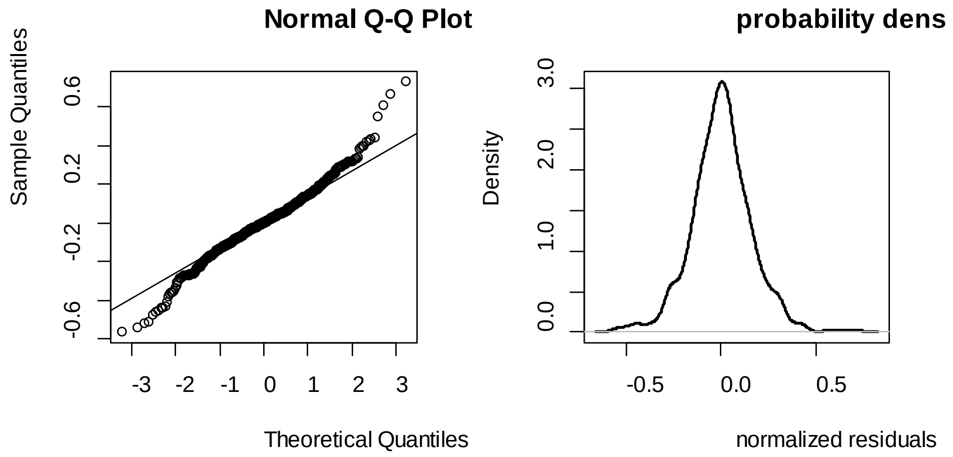


Figure 53. QQ plot and PDF of the residuals.

However, the residuals neither pass the Shapiro-Wilk normality test, nor the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for level stationarity.

Shapiro-Wilk normality test: p-value = 3.623e-08

KPSS Level = 0.6204, p-value = 0.02078

Augmented Dickey-Fuller Test: p-value = 0.01

As can be seen in the Figure 53b, the distribution of the residuals has long tails at both ends, and this is the main reason of non-stationarity and non-normality.

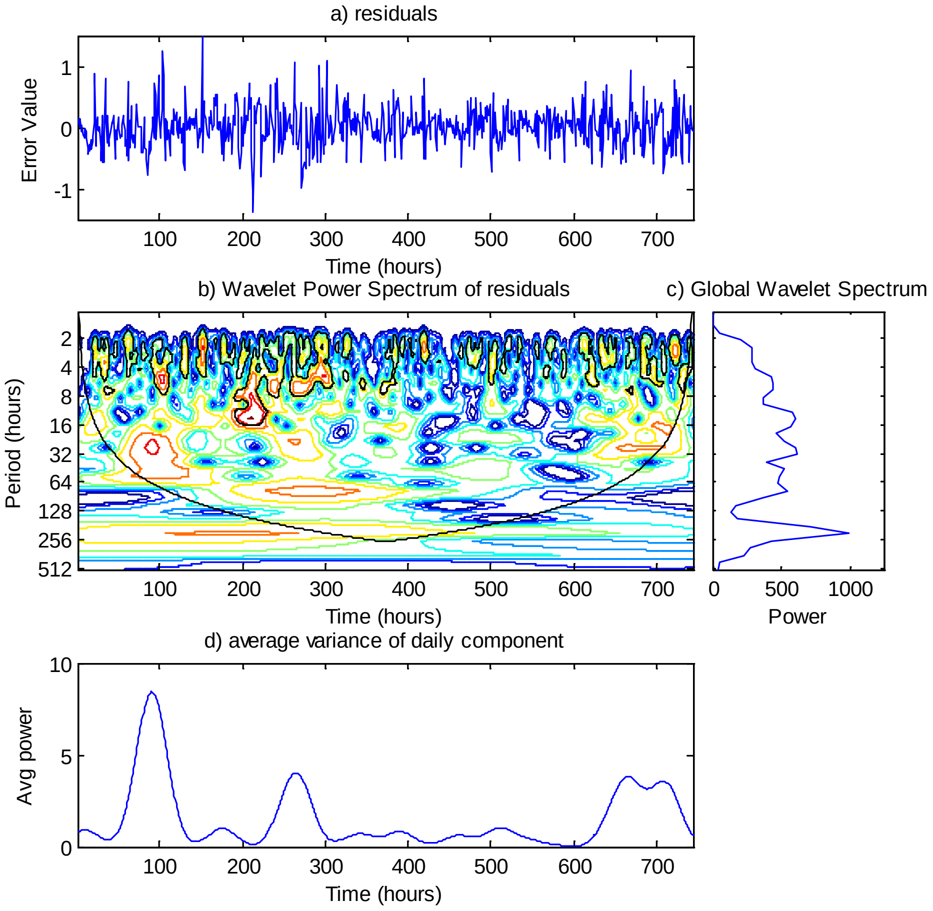


Figure 54. Wavelet analyses of residuals with load forecast error predictions

(a) Residuals; b) Contour map of wavelet power of decomposed components at different time scales; c) The spectrum shows that data consists of weak daily and weekly components; d) The contribution of a short-term component is high during the early and late times.

Figure 54 shows the wavelet analysis of the residuals on the data set using the same approach. Similar to Figure 52, one can see that there are not any dominant signals at any particular scales, which means that the major variability in the forecast errors has been captured.

6.4 Conclusion

In this study, we introduced a novel statistical separation method to improve system predictability and reduce uncertainties by separating slow, fast, and very fast motions in power system total load or net load. The method based on regression trees was developed to forecast load, and load/net load forecast errors using real power system data. From the wavelet analysis, both daily and weekly components are separated from the very fast motions of the signals, except for the mismatch in the daily component around the 270^{th} hours corresponding to strong positive anomalies in the forecast errors.

The fast motion can be extracted from slow motions by obtaining the real forecast errors in the system; while the very fast motion can be separated from fast motion by subtracting the predicted forecast errors from the actual forecast errors. Our results indicate that the system predictability is significantly improved, e.g., ~40% reduction.

6.5 References

[1] S. Mallat (1999) "A Wavelet," Tour of Signal Processing, 2nd ed. San Diego, CA: Academic, 1999.

[2] I. Daubechies 1990) The Wavelet Transform, Time-Frequency Localization and Signal Analysis, IEEE Trans. Inform. Theory, 36(6), 961-1005.

[3] L. Breiman, J. Friedman, C. Stone, and R. A. Olshen, Classification and Regression Trees (Wadsworth Statistics/Probability). Chapman and Hall/CRC, 1984.

[4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," Inf. Theory, IEEE Trans., vol. 36, no. 5, pp. 961–1005, Sep. 1990.

[5] "Electricity Load Forecasting Using Neural Networks and Bagged Decision Trees.".

[6] Z. Hou, Y. Makarov, N. Samaan, and P. Etingov, "Standardized Software for Wind Load Forecast Error Analyses and Predictions Based on Wavelet-ARIMA Models -- Applications at Multiple Geographically Distributed Wind Farms," in System Sciences (HICSS), 2013 46th Hawaii International Conference on, 2013, pp. 5005–5011.

[7] Z. Hou and X. Li, "Signal Detection of Global Climate Change and External Forcing Factors," Acta Scientiarum Naturalium Universitatis Pekinensis, 2000.

7.0 Principal Stress Direction Extraction and Motion Trajectory Approximation

The goal of this task is to identify the motions of the WECC system based on observations rather than engineering intuition. These motions give us the actual stress directions to be applied to the system models and to get the system security margin. To minimize the number of stress directions, we applied the principal component analysis (PCA). In this analysis, we considered 39 areas within the WECC system. PCA has been used to analyze total hour-to-hour load variations (including wind and solar generation) in different zones. Forecast data including loads, wind power and solar power have been analyzed through PCA. Approximation errors have been analyzed to quantify uncertainties.

PCA has been widely used to identify principal components in analysis of power system dynamic. For example, a PCA-based model reduction method was developed for power systems in [1]; based on PCA, reference [2] developed a wind speed prediction method; reference [3] used PCA to enhance wind power production costing simulation efficiency; and reference [4] used PCA to analyze the impact of wind power generation on European cross-border power flows.

In this study, we used PCA to extract principal power system stress directions, with wind power and solar power being considered. These principal stress directions are essential to characterizing system motion and quantification of uncertainties.

7.1 Principal Stress Direction Extraction

Suppose P_{ij} is the net real power[1] of zone i at time instant j; then $\Delta P_{ij} = P_{i(j+1)} - P_{ij}$ represent the load increment of zone i in the time period of $[j, j+1]$. Suppose we have n zones (in this case $n = 39$). Define vector

$$v_j = [\Delta P_{1j}, \Delta P_{2j}, \cdots \Delta P_{nj}] \tag{0}$$

We have different v_j for different days (over one month) as shown in blue solid curves in Figure 55. Here $j = 8$, n=39, namely v_j represent 39 loads increment between 8 and 9 a.m.. Let v_j^d represent v_j in the d-th day.

As can be seen from Figure 55, although v_j has different shapes in different days, the shapes are similar. It can be seen from, the first three principal components account for most of the shapes. The three principal components ($v_{p1,} v_{p2,} v_{p3}$) are shown in the following:

---

[1] Forecast of load minus forecast of wind and solar energy

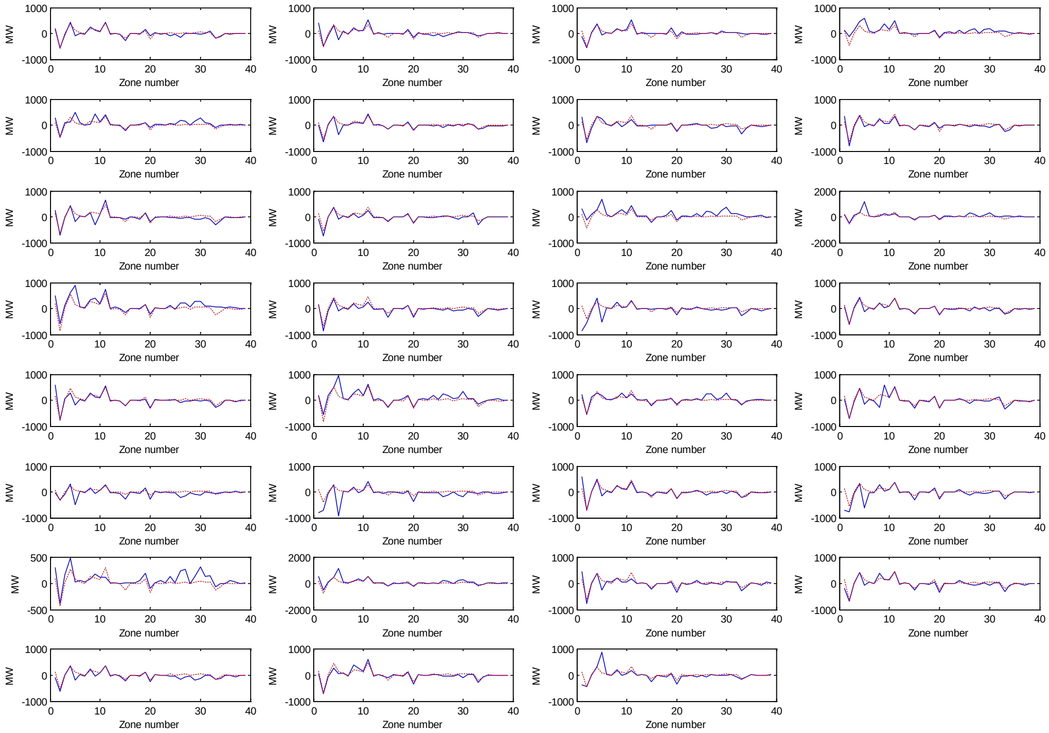


Figure 55. Plot of v_8 (representing increments in the interval between 8 and 9 a.m.) for different days in January in the parameter space of 39 zones.

7.2 Approximating System Motion Trajectories In Parameter Space

Using the singular value decomposition (SVD) to implement the idea of PCA, we have:

$$M = UDW^T \tag{0}$$

Here U is a 31×31 matrix; W is a 39×39 matrix; M is a 31×39 matrix and defined in the following:

$$M = \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^{30} \\ v^{31} \end{bmatrix} \tag{0}$$

D is a 31×39 matrix and is shown in the following:

$$D = \begin{bmatrix} D_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & D_{ii} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & D_{31} & 0 & 0 \end{bmatrix} \tag{0}$$

The first three rows of matrix W are the first three principal components. Singular values for different components are shown in Figure 56, and the first three principal components are shown in Figure 57.

Let $D_{ii} = 0$ (i>3), we have:

$$T = U * D$$

Here T is a 31×39 matrix equal to U multiplied by D; and $T(:,r) = 0$ for r = 1,..,3. Finally, we have

$$M \approx \tilde{T} \times \begin{bmatrix} v_{p1} \\ v_{p2} \\ v_{p3} \end{bmatrix} \tag{0}$$

Here we use the following notation: T(:,r) – is the r-th column of T.

According to equation (3), we have

$$v^{i} \approx T(i,1) \times v_{p1} + T(i,1) \times v_{p2} + T(i,1) \times v_{p3} \tag{0}$$

Using the first three principal components, the approximation is shown in Figure 58. The dotted lines represent approximated values.

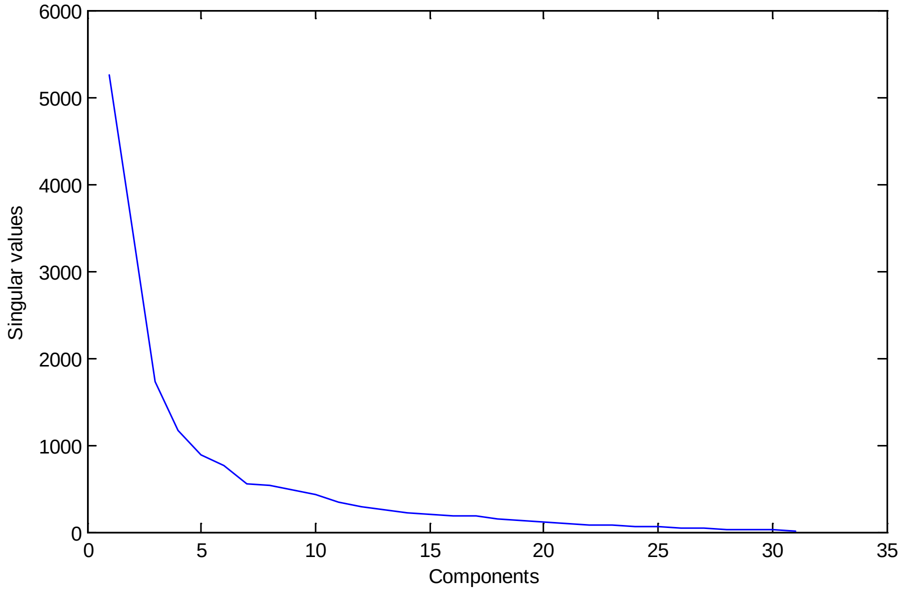


Figure 56. Singular values for different components.

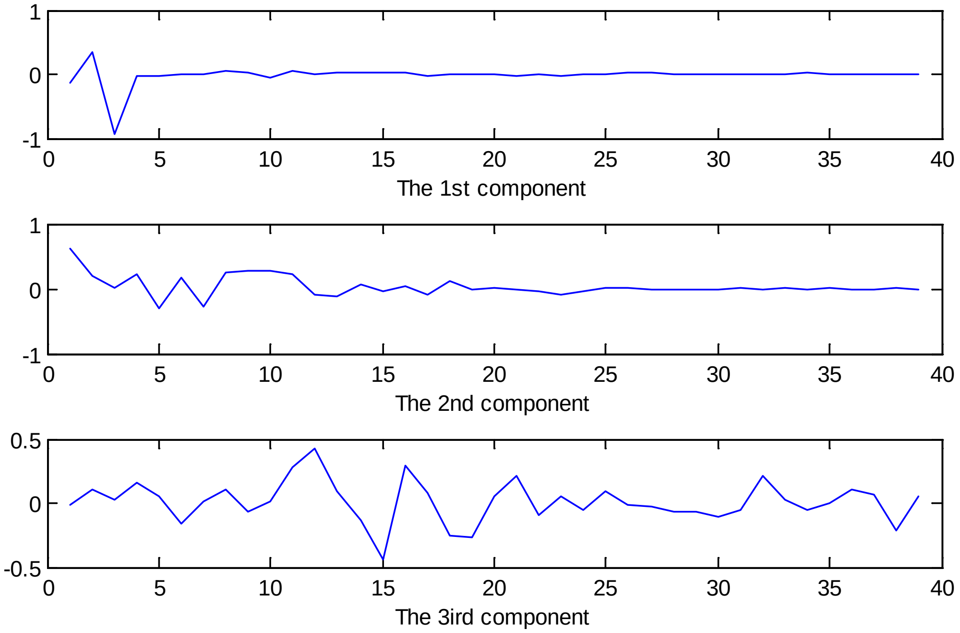


Figure 57. The first three principal components.

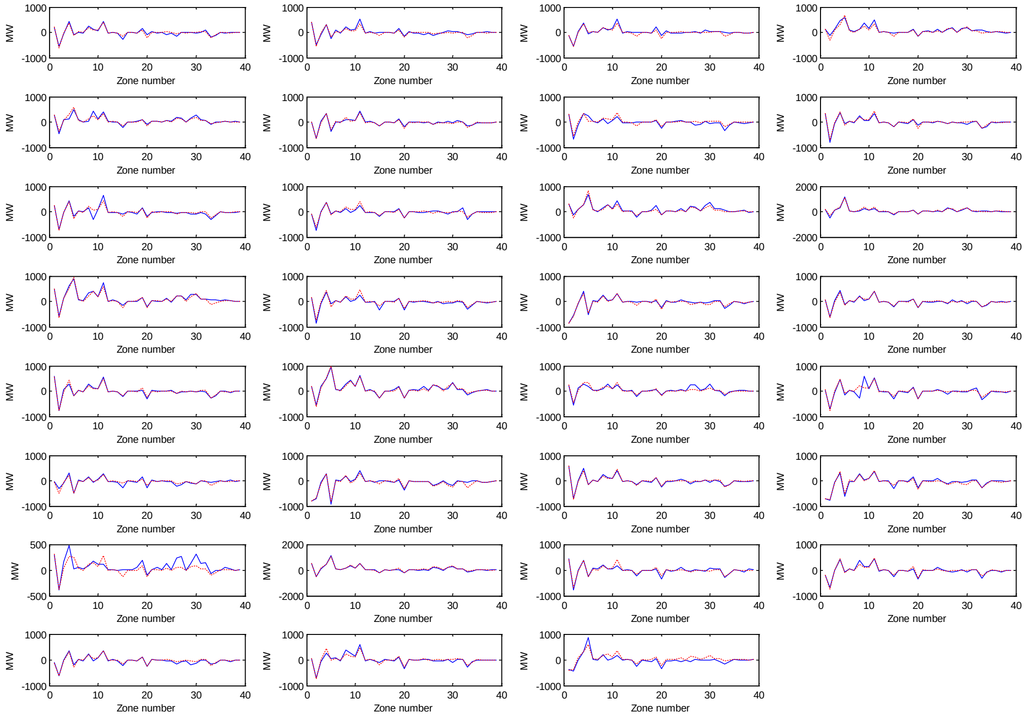


Figure 58. Approximation using first three principal components

Using the first two principal components, the approximation is shown in Figure 59. Figure 60 and Figure 61 show the approximation using only the first principal component. Remember v_j^d and v_j defined in Equation (1). So in the space of $P_{i\bullet}$, we have 31 vectors $(v_j^d, j = 8, d = 1, 2, \cdots, 31)$ for the time interval between 8 and 9 a.m.. These 31 vectors can be represented by the three vectors: v_{p1}, v_{p2}, v_{p3}.

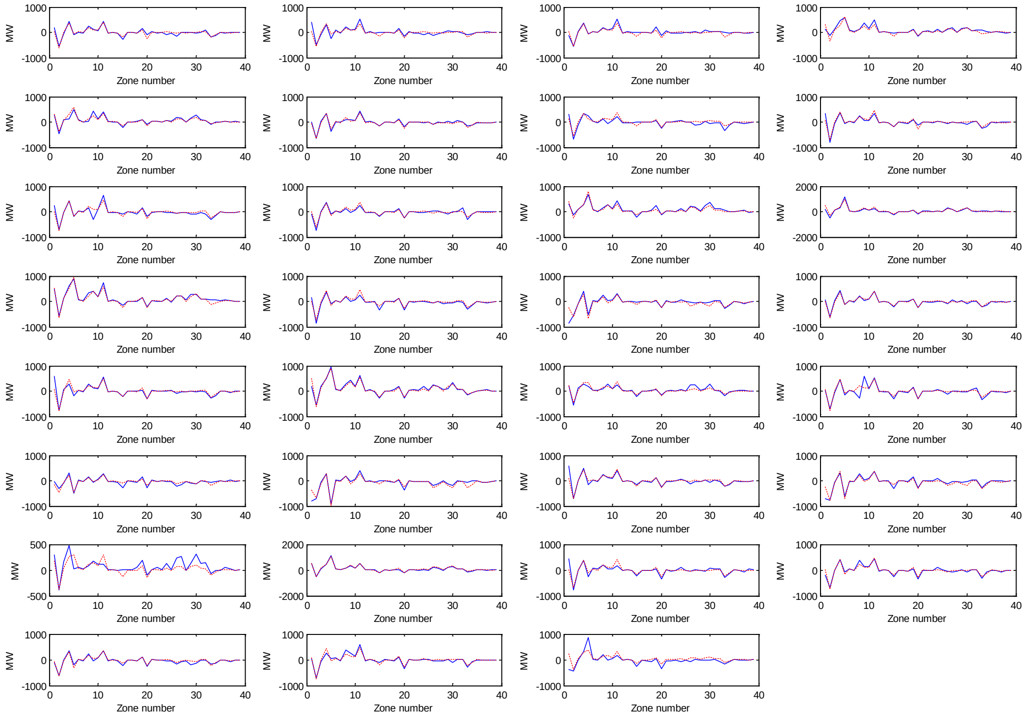


Figure 59. Approximation using the first two principal components.

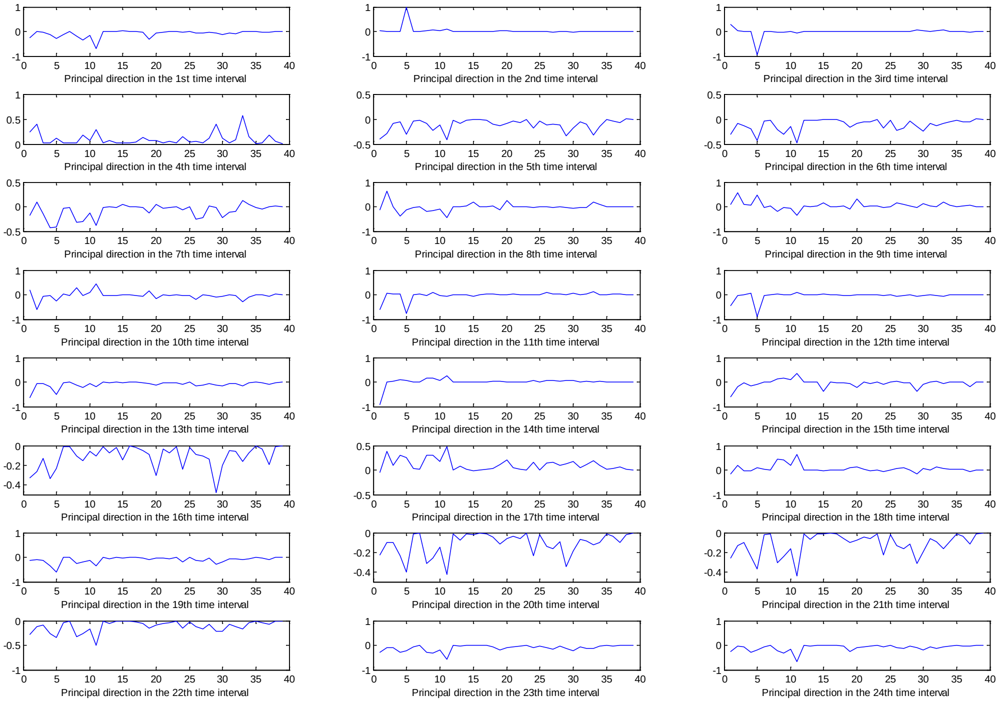


Figure 60. Principal directions for 24 intervals of January.

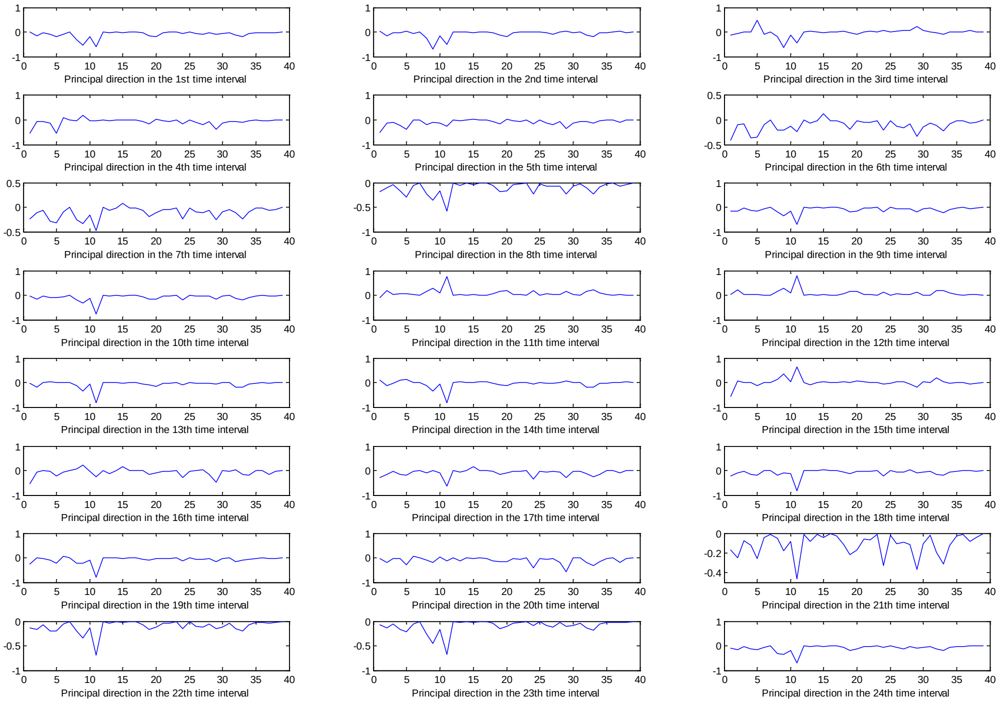


Figure 61. Principal directions for 24 intervals of July.

7.3 Conclusions

The following conclusions can be drawn from this study:

- In WECC, the system motion behaviors on different days have a high degree of similarity.
- Usually, the motion characteristics in the WECC system zones can be well captured by the first two or three principal components.
- PCA is a promising tool to analyze system motion characteristics.

7.4 References

[1] S. Wang, S. Lu, N. Zhou, G. Lin, M. Elizondo, and M. A. Pai, “Dynamic-Feature Extraction, Attribution, and Reconstruction (DEAR) Method for Power System Model Reduction,” Power Syst. IEEE Trans., vol. 29, no. 5, pp. 2049–2059, 2014.

[2] Q. Hu, P. Su, D. Yu, and J. Liu, “Pattern-Based Wind Speed Prediction Based on Generalized Principal Component Analysis,” Sustain. Energy, IEEE Trans., vol. 5, no. 3, pp. 866–874, Jul. 2014.

[3] D. J. Burke and M. J. O’Malley, “A Study of Principal Component Analysis Applied to Spatially Distributed Wind Power,” Power Syst. IEEE Trans., vol. 26, no. 4, pp. 2084–2092, Nov. 2011.

[4] M. Zugno, P. Pinson, and H. Madsen, “Impact of Wind Power Generation on European Cross-Border Power Flows,” Power Syst. IEEE Trans., vol. 28, no. 4, pp. 3566–3575, Nov. 2013.

8.0 Probabilistic Quantification and Observation of WECC System Transmission Limits

8.1 Introduction

The electrical power system is a complex network both from an operational as well as an analysis point of view. Large networks spanning several terawatts of generation and loads are interconnected, spanning several countries to operate a stable and sustainable power system. Many political and administrative entities that own parts of such integrated power systems often have bilateral power transfer limits across their boundaries (tie lines) for several reasons including stability, statutory export-import limitations of power, or tie-line capacity limits. Traditionally, Energy Management Systems (EMS) solutions were developed to remotely monitor and provide appropriate supervisory controls to operate the power system economically and stably. This includes economic dispatch, unit commitment, state estimation, etc. Most of the analyses (like security analysis and unit commitment) are performed offline or on a time-ahead basis (e.g., hour-ahead or day-ahead etc.) based on forecasted load so that responses to future events could be planned ahead of time. With increasing penetration of intermittent generation into bulk power system, generation uncertainty also plays a key role in time-ahead planning for efficient operation of the power system. Under these circumstances, stochastic tools are very helpful in both planning and operation of power systems.

Interacting wind, solar, and load forecast errors can create significant unpredictable impacts on the transmission system, congestion, voltage and reactive power stability margins. For instance, the worst case can be a combination of the significant positive errors in one area of the power system interconnection with the significant negative errors in another area. This could lead to a significant deviation of power flow from the scheduled values, and thus create congestion and voltage stability problems. These impacts will be increasing with increasing penetration levels of variable renewable generation in the Western Interconnection and California power systems. To maintain a secure system reliability level, the probability and the magnitude of the impacts should be evaluated and communicated to system operators. Based on this information, the system power flow limits, dispatches, voltage levels, and available reactive power margins could be adjusted to minimize the risk of system problems and failures to an acceptable level whenever it is required.

Generation re-dispatch can be an effective way to change the power flows along multiple transmission lines of the power system. Time-ahead analysis with generation and load forecasts takes into consideration a snapshot of the system and try to plan generation based on that. The probability distribution of the forecasting errors can provide worst-case scenarios that can cause congestion along important transmission lines. TUT provides a holistic view of the system superimposed by the density functions that arise from various uncertainties. This task develops a methodology based on generation dispatch to simultaneously reduce worst-case tie-line congestion caused by generation and load uncertainties.

8.2 Methodology Description

Forecasts are based on a forecasted snapshot of the system. But we know that there is a white noise probability density associated with each forecasted quantity. System-wide aggregation of these error distributions will produce a resultant distribution of power flows in the tie-lines between different areas. Although a single snapshot of the forecast shows no violation of the tie line power flow limits, there is a finite probability that the limits are violated as shown in Figure 62 for the California-Oregon Intertie (COI) interface of WECC.

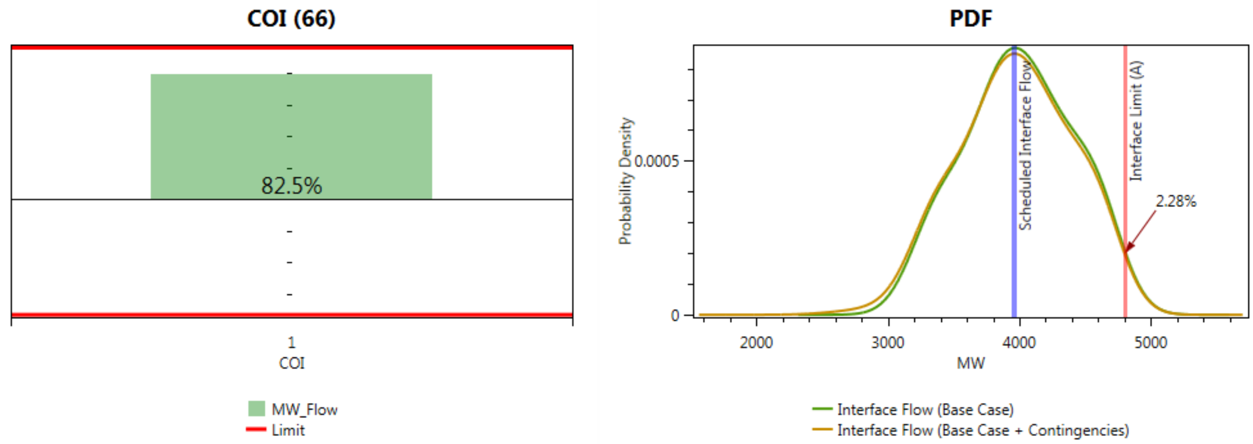


Figure 62. Power flow probability density function for the COI interface of WECC.

In Figure 62, there is a 2.28% probability that the power flow exceeds the interface limit (4800 MW). The distribution function is spread around the base power flow (3960 MW). If the base power flow is reduced by 400 MW, the entire probability distribution shifts to the left leaving the probability that the power flow exceeds interface limit to less than some tolerance (say 0.1%). Generation re-dispatch can be used to move the base power flow as shown in Figure 63.

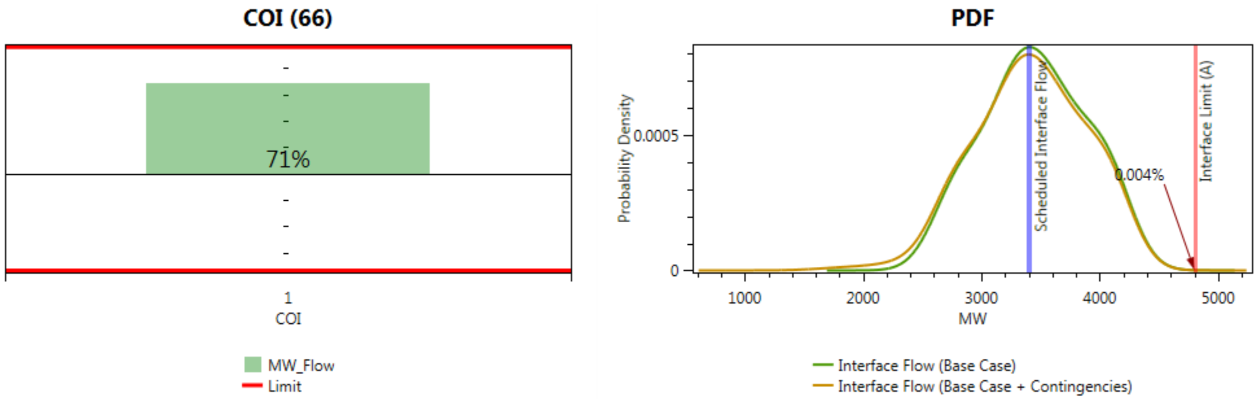


Figure 63. Probability density function for the COI interface with generation re-dispatch.

Similar distributions can be generated for every transfer in the system. A holistic mathematical formulation is necessary to calculate simultaneous generation re-dispatches in different areas that would reduce the probability of exceeding the interface limit to less than the tolerance limit.

8.2.1 Mathematical formulation

Power transfer distribution factor (PTDF) represents the sensitivity of power flow in a particular interface with respect to generation changes in respective areas. Mathematically, this can be described as,

$$\Delta F_i = PTDF_{ij} \times \Delta P_j \tag{1}$$

where,

$\Delta F_i \rightarrow$	represents the change in power flow in *i*th interface
$PTDF_{ij} \rightarrow$	represents the PTDF in *i*th interface for change in generation in jth area
$\Delta P_j \rightarrow$	represents the change in generation in jth area.

For a system with 'n' number of areas, each area generation can be re-dispatched. The resultant change in the flows of each interface is an aggregation due to changes in multiple area generations. This could be mathematically formulated as,

$$\Delta F_i = \sum PTDF_{ij} \times \Delta P_j \tag{2}$$

For each of the interface, acceptable limits for base power flow can be calculated when superimposed with the probability distribution of forecast errors will be within the interface transfer limits. This will provide $\Delta F_i{}^{min}$ and $\Delta F_i{}^{max}$ for each of the interface. We can use these limits as constraints of power flow changes.

$$\Delta F_i{}^{low} \leq \Delta F_i \leq \Delta F_i{}^{high} \tag{3}$$

These constraints are binding and equation (2) can be used to determine appropriate generation adjustment ΔPj in each area j. But this can produce an under-determinant set of equations that can result in infinite solutions. To resolve this issue, an objective function is developed based on the understanding that minimal generation adjustment is expected. Hence, the objective function is developed as

$$\text{Minimize, } \sum \Delta P_j{}^2 \tag{4}$$

This problem could be solved using quadratic optimization. Other forms of objective functions can be developed such as,

$$\text{Minimize, } \sum C_j \Delta P_j{}^2 \tag{5}$$

where $C_j \rightarrow$ represents the cost of generation adjustment in each area. An additional constraint, which can be added to the problem is $\sum \Delta P_j \;\; = 0$. Assuming that the analyzed case is a balanced case with acceptable frequency, this last equation is an equivalent of the power balance equation.
Results reported in this section have a preliminary nature. More work is needed in Phase 2 of this project.

8.3 Demonstration

The proposed formulation has been applied to WECC system particularly for the 2020 planning case. WECC has 21 areas. The limits of transfers for each of the interfaces are shown in Table 4.

Table 4. Interface base power flows and limits of flow changes

Path number	Interface_Name	Interface_MW_Flow	ΔF_low	ΔF_high
1	ALBERTA - BRITISH COLUMBIA	84.2	-1258	903
85	ALLSTON-KEELER	1374.3	-2580	-180
76	ALTURAS PROJECT	145.956	-370	100
21	ARIZONA - CALIFORNIA	-124.5	-5568	5814
53	BILLINGS - YELLOWTAIL	-68.125	-180	280
33	BONANZA WEST	387.739	-985	275
17	BORAH W	304.202	-2407	1827
19	BRIDGER W	762.067	-2908	1396
55	BROWNLEE EAST	25.349	-1864	1814
45	CA INDEPENDENT - MEXICO (CFE)	-413.79	56	764
50	CHOLLA - PINNACLE PEAK	374.177	-1540	786
66	COI	3493.69	-5975	200
100	COI+PDCI	2279.518	-4940	380
77	CRYSTAL - ALLEN	237.827	-1138	670
59	EAGLE MTN 230/161 KV - BLYTHE 16	40.591	-252	166
62	ELDORADO - MCCULLOUGH 500 KV	660.417	-3168	1843
58	ELDORADO - MEAD 230 KV LINES	-59.563	-1058	1182
18	IDAHO - MONTANA	-214.798	-37	477
16	IDAHO - SIERRA	-272.629	20	700
14	IDAHO-NW	-219.932	-580	2240
42	IID - SCE	150.955	-724	392
29	INTERMOUNTAIN - GONDER 230 KV	13.853	-178	150
28	INTERMOUNTAIN - MONA 345 KV	-263.343	-820	1550
60	INYO - CONTROL 115 KV TIE	-48.877	1	95
61	LUGO - VICTORVILLE 500 KV LINE	1343.852	-1980	780
8	MONTANA - NORTHWEST	665.069	-1650	1100
75	MP-SL	-374.247	280	1500
73	N JOHNDAY	3278.946	-11100	4400
43	NORTH OF SAN ONOFRE	1745.55	-4130	580

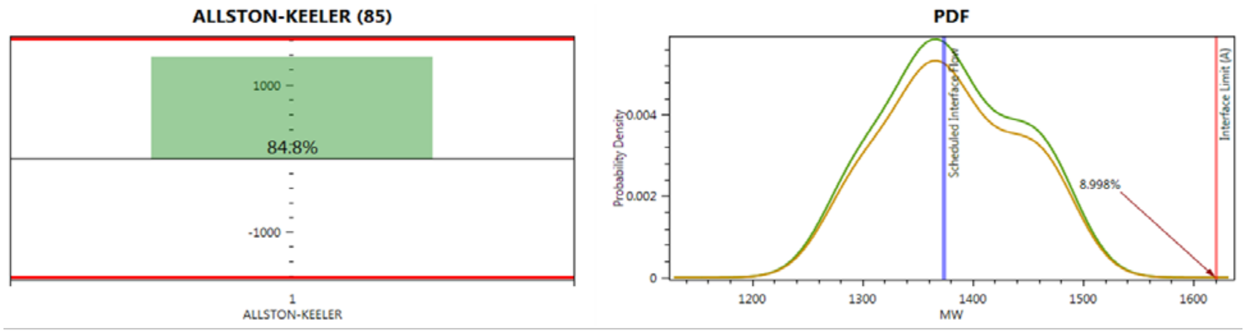
48	NORTHERN NEW MEXICO (NM2)	931.659	-2320	440
25	PACIFICORP/PG&E 115 KV INTERCON.	9.541	-45	63
15	Path 15	390.302	-4660	2980
26	PATH 26	2526.669	-3600	-1000
20	PATH C	-43.254	-880	980
32	PAVANT, INTRMTN - GONDER 230 KV	25.181	-205	360
24	PG&E - SPP	-11.812	-98	140
52	SILVER PEAK - CONTROL 55 KV	-1.256	-10	11
44	SOUTH OF SAN ONOFRE	404.449	-2540	1840
47	SOUTHERN NEW MEXICO (NM1)	59.403	-1048	956
41	SYLMAR - SCE	-296.806	-800	1280
30	TOT 1A	167.752	-610	310
31	TOT 2A	293.908	-790	290
34	TOT 2B	121.354	-810	500
78	TOT 2B1	83.76	-560	380
79	TOT 2B2	37.594	-288	177
36	TOT 3	564.598	-2035	855
37	TOT 4A	308.169	-950	390
38	TOT 4B	189.537	-760	400
39	TOT 5	633.353	-2155	755
40	TOT 7	61.568	-850	670
9	WEST OF BROADVIEW	1772.536	-4073	533
4	WEST OF CASCADES - NORTH	858.656	-10950	9270
5	WEST OF CASCADES - SOUTH	749.937	-7600	6100
10	WEST OF COLSTRIP	1977.045	-4518	498
11	WEST OF CROSSOVER	1949.331	-4428	488
6	WEST OF HATWAI	421.995	-3060	2080
46	WOR	609.764	-10648	9358

The optimization as defined in equations (2), (3) and (4) is applied. The resultant solution for area generation changes is shown in Table 5.

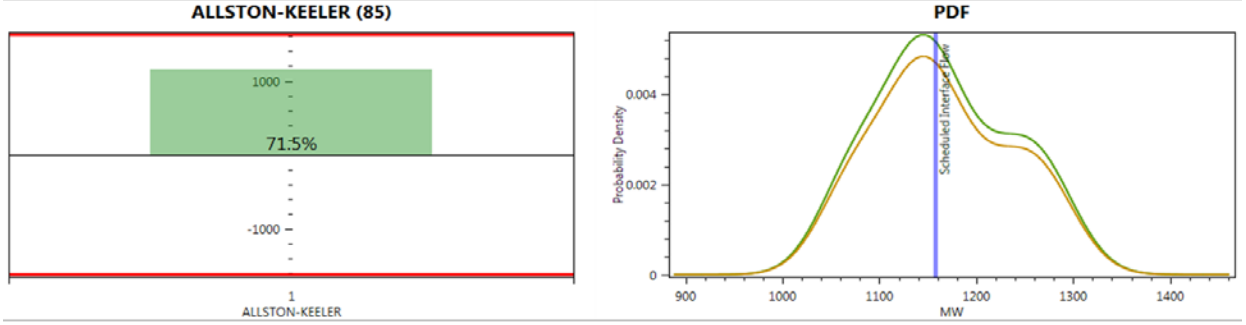
Table 5. Calculated generation changes to eliminate probabilistic interface congestions

Area	Base Gen (MW)	Change (MW)	Computed New Gen (MW)	Considered New Gen (MW)
NEW MEXICO	3300	589	3889	3889
EL PASO	1553	202	1755	1755
ARIZONA	29665	936.44	30601.44	30601.44
NEVADA	6900	796.63	7696.63	7696.63
MEXICO-CFE	3409	498	3907	3907
IMPERIALCA	1635	325	1960	1960
SANDIEGO	4381	198	4579	4579
SOCALIF	20842	-318.75	20523.25	20523.25
LADWP	5570	-821.19	4748.81	4748.81
PGANDE	30545	1168.99	31713.99	31533
NORTHWEST	34100	-111.33	33988.67	33988.67
B_C_HYDRO	10948	-708.62	10239.38	10239.38
FORTISBC	901	-404.9	496.1	496.1
ALBERTA	13995	-851.7	13143.3	13143.3
IDAHO	4611	-965.59	3645.41	3645.41
MONTANA	3054	409.45	3463.45	3463.45
WAPA U_M_	49	12	61	61
SIERRA	2676	-760.83	1915.17	1915.17
PACE	8517	-181.05	8335.95	8335.95
PSCOLORADO	7990	-15.35	7974.65	7974.65
WAPA R_M_	6686	3.8	6689.8	6689.8

By application of the calculated generation re-dispatches, the improvement in interface congestion is shown in Figure 65 – Figure 69.

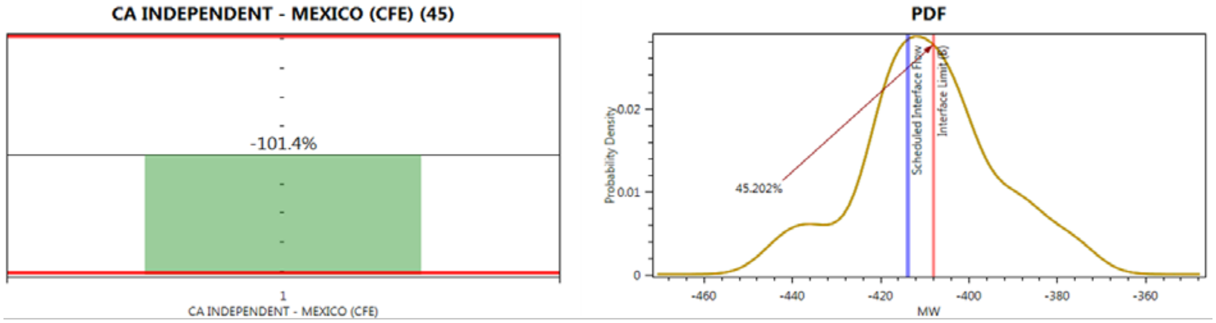


Base Case

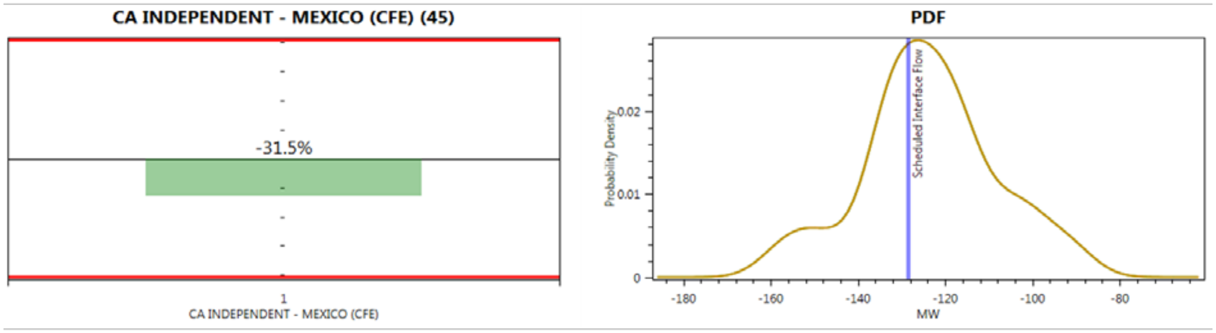


Updated Case with Computed Generation Redispatch

Figure 64: Improvement in congestion in the ALLSTON-KEELER interface.

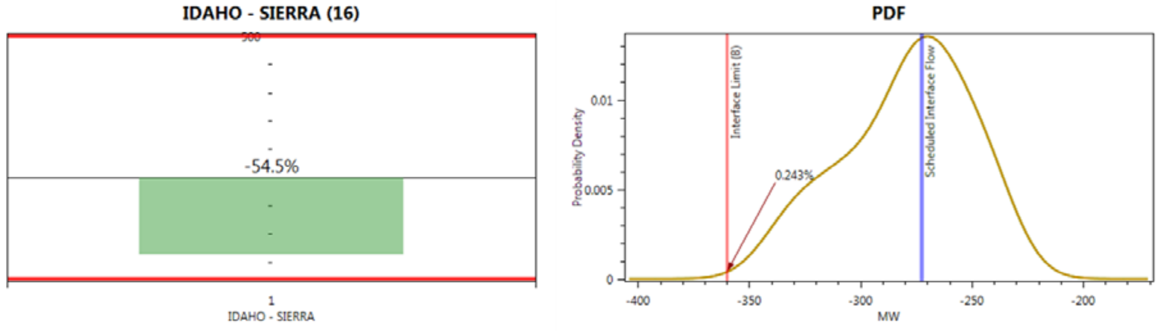


Base Case

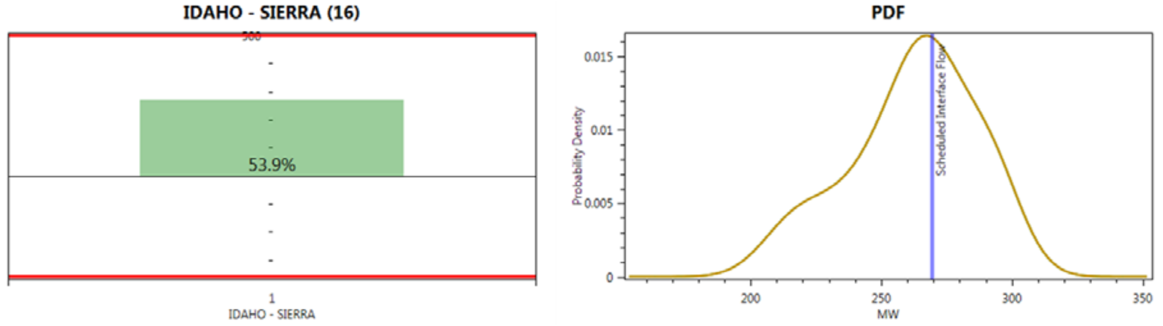


Updated Case with Computed Generation Redispatch

Figure 65: Improvement in congestion in the CA INDEPENDENT - MEXICO (CFE) interface.

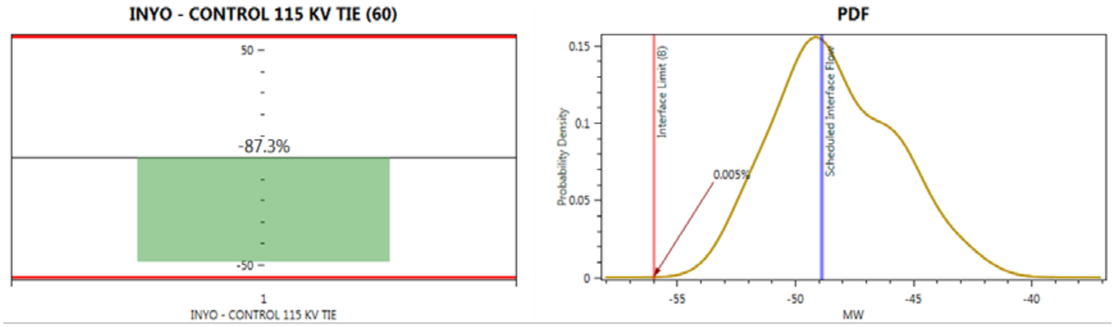


Base Case

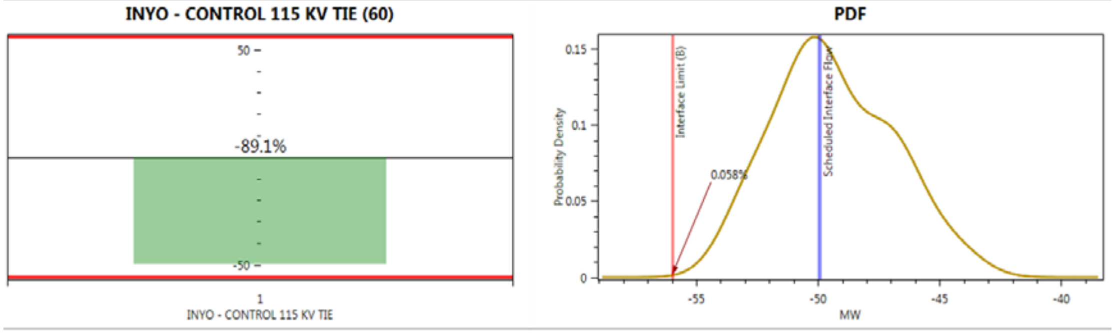


Updated Case with Computed Generation Redispatch

Figure 66: Improvement in congestion in the IDAHO - SIERRA interface.

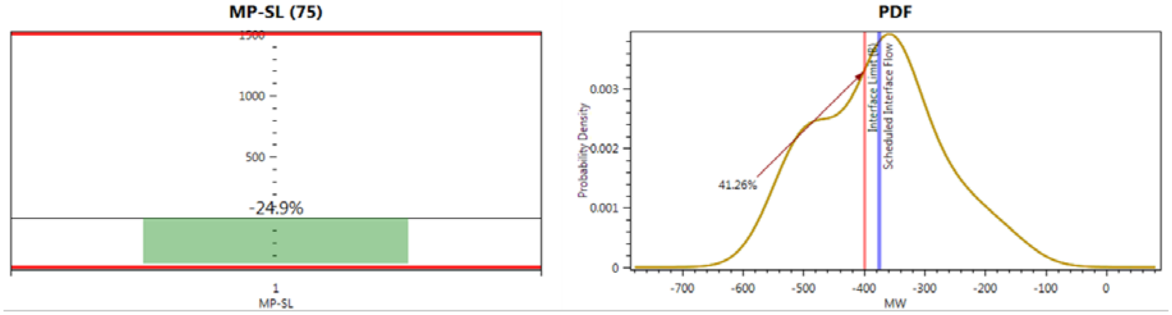


Base Case

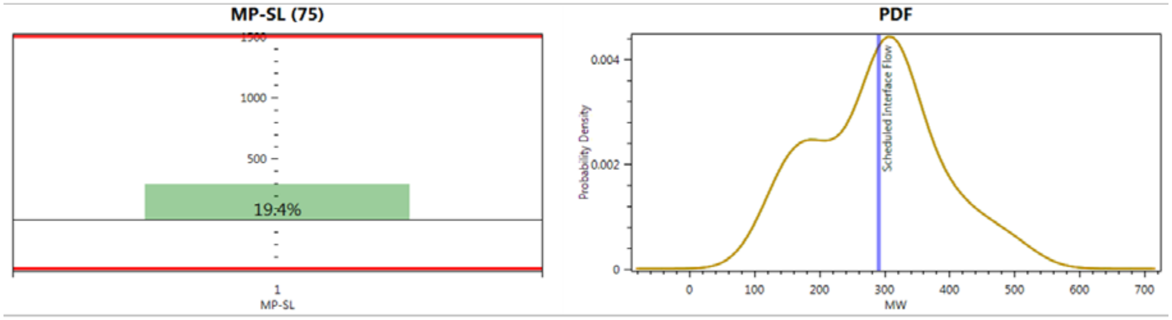


Updated Case with Computed Generation Redispatch

Figure 67: Improvement in congestion in the INYO - CONTROL 115 KV TIE interface.

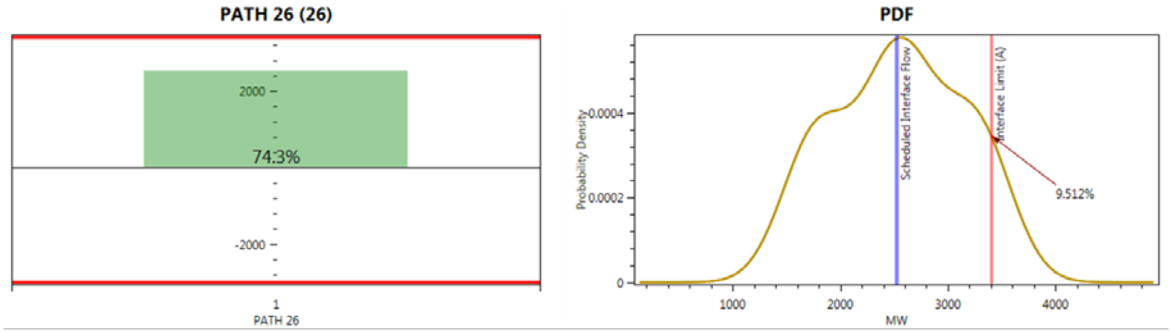


Base Case

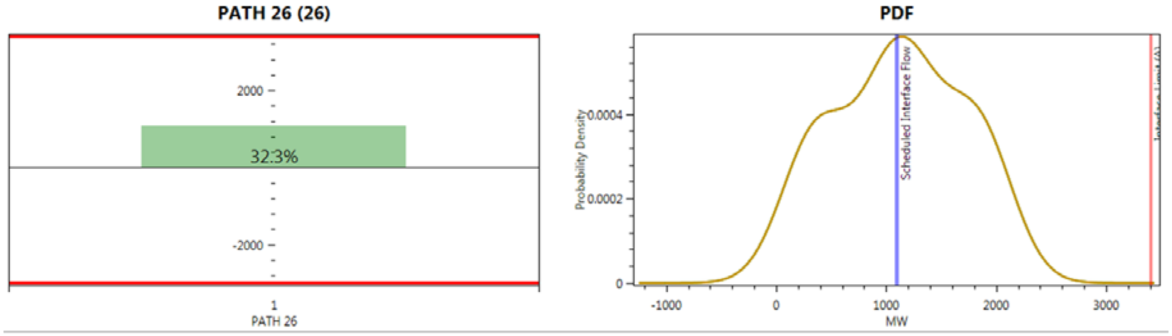


Updated Case with Computed Generation Redispatch

Figure 68: Improvement in congestion in the MP-SL interface.



Base Case



Updated Case with Computed Generation Redispatch

Figure 69: Improvement in congestion in the PATH 26 interface.

8.4 Conclusion

In this study, we proposed a novel methodology based on generation re-dispatch to eliminate the probability of transmission congestion due to uncertainties in generation and load forecasts. The transmission tool developed is very handy to simultaneously view the impact of uncertainties on all the transmission paths with aggregation of the probability distributions.

The results obtained are very encouraging in achieving the objectives of reducing transmission re-dispatches. The proposed methodology will be tested further with cost data for area generation.

9.0 Transmission Uncertainty and Prediction Tool (TUT)

Interactions among wind generation, solar generation, and load forecast errors can have a random impact on transmission system. This event can lead to additional congestion and voltage stability problems. To mitigate this problem, balancing authorities (BAs) should be able to re-dispatch their conventional generators ahead of time, if such impact can be predicted.

A worst-case example of interactions among wind and solar generation forecast errors and load forecast errors can having a random impact could be a combination of significant positive errors in one area (ΔP_1) of the system with significant negative errors in another areas (ΔP_2 and ΔP_3), as illustrated in Figure 70. This event can lead to deviations of power flows from their scheduled values and thus cause additional congestion and voltage stability problems. If the transmission impacts of uncertain parameters remain unpredicted (and the BAs therefore do not have the information to re-dispatch generation ahead of time), they can cause (1) unexpected reliability problems; (2) real-time energy price spikes because of the use of costly fast-responsive balancing resources; and (3) in some circumstances, curtailment of excessive wind and solar generation.

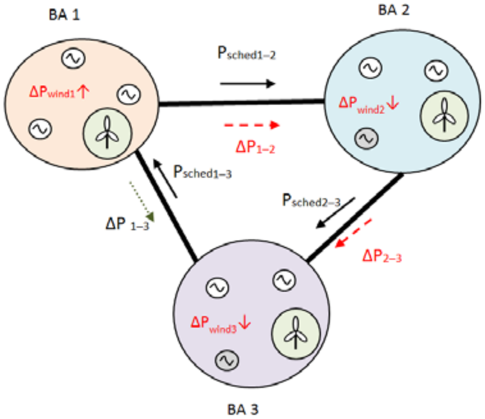


Figure 70. An illustration of transmission congestion caused by load, wind and solar forecast errors.

To maintain an adequate system reliability level, the probability and the magnitude of these potential transmission impacts should be predicted and communicated to the system operators or directly to the market/EMS systems. Based on this information, the system power flow limits, dispatches, voltage levels, and available reactive power margins could be adjusted to minimize the risk of system problems, failures, and real-time market spikes.

A new probabilistic methodology and new analytical software called the Transmission Uncertainty and Prediction Tool (TUT) have been developed at Pacific Northwest National Laboratory in order to help BAs facilitate higher penetration of renewable resources without compromising system reliability. This work was supported by the U.S. Department of Energy (DOE) Office of Electricity Delivery & Energy Reliability (OE) and Office of Energy Efficiency & Renewable Energy (EERE) and by the California Energy Commission (CEC). The developed methodology predicts the impact of variable resources on congested paths, worst-case voltage drops, and reactive power margins at low voltage points for several hours ahead of time and, if needed, proposes control actions to mitigate the problem if necessary.

The TUT analyzes the impact of wind generation and other sources of uncertainty on the base case and under user-specified contingencies, so that the most limiting contingency is determined and addressed. A conceptual design of the developed transmission uncertainty analysis methodology is shown in Figure 71. The developed methodology is based on the Monte Carlo simulations and advanced statistical analysis of different sources of uncertainty that can impact on the transmission network .The developed methodology also includes a linearized power flow model to calculate incremental active power flows in the transmission network caused by forecasting errors.

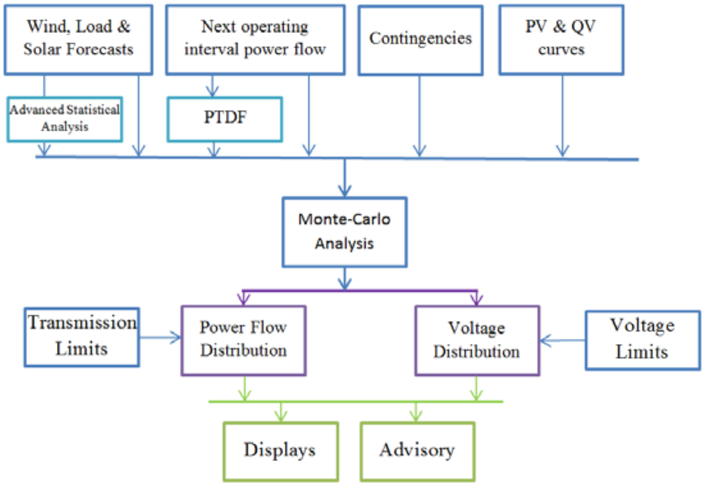
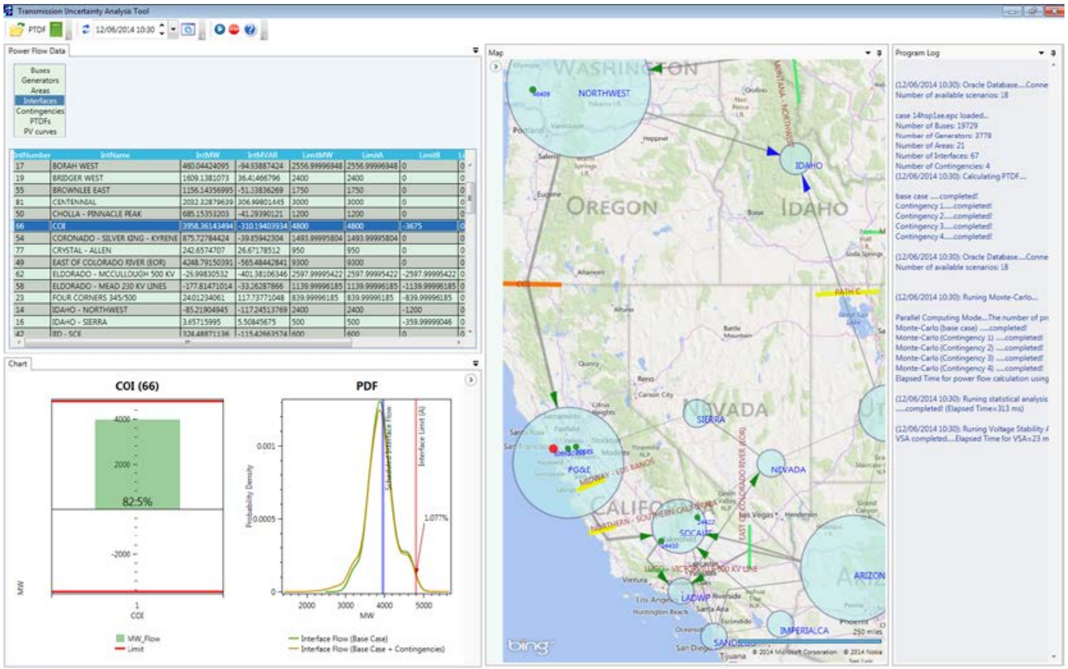


Figure 71. A concept of probabilistic transmission uncertainty analysis.

The impact of random variations of wind and solar generation and system load from their forecasted values on power transfers is revealed by simulating forecasting errors (of different kinds) distributed over a large geographical area. The statistical model based on the ARIMA method was developed for this purpose. Errors caused by variability (deviations from flat energy schedules) are also simulated. Based on this model and an incremental linearized model reflecting the impacts of system imbalances on power transfers, we determine the probability distributions of the flows by repeating simulations of random forecast errors and the corresponding increments of monitored flows. These distributions are used to find the probabilistic load margin. Also, these distributions and P-V and Q-V curves computed by a voltage stability application are used to determine probability distributions of the voltage magnitudes and reactive power margins.

The TUT is a standalone Windows application. It has been implemented using parallel computing techniques that allow us to significantly speed up the computational process. It has an advanced graphical user interface (GUI), which includes multiple customizable dashboards and transmission network visualizations based on geographical information systems (GIS) technology (Google Maps or Bing Maps). Real-time dispatchers can monitor any transmission interface to check the probability distribution of power flows at a specific look-ahead operating interval. The users will also be advised if the interface transfer limit potentially can be violated, and what the probability and size of this violation are. Congested transmission interfaces with potential voltage problems can be highlighted on the map. Depending on the severity level of potential transfer and voltage limit violations, different colors can be used. Figure 72 shows a real-case example, when the probabilistic analysis was applied to the WECC system model that includes 19729 buses and 3778 generators. One can see that based on the analysis the tool predicts potential violation of California Oregon Intertie (COI) transmission limit with 1.08% probability during next operating hour. Thus, real-time dispatchers will have time to re-dispatch units in California and Pacific Northwest systems to prevent this potential violation.


Transmission Uncertainty Analysis Tool
PTDF
12/06/2014 10:30
Power Flow Data
Buses
Generators
Areas
Interfaces
Contingencies
PTDFs
PV curves
Map
Program Log
Chart
COI (66)
4000
2000
82.5%
-2000
MW
1
COI
MW_Flow
Limit
PDF
0.001
0.0005
0
Probability Density
Scheduled Interface Flow
Interface Limit (A)
1.077%
2000
3000
4000
5000
MW
Interface Flow (Base Case)
Interface Flow (Base Case + Contingencies)
WASHINGTON
OREGON
IDAHO
NEVADA
CALIFORNIA
NORTHWEST
IDAHO
MONTANA - NORTHWEST
PATH C
SIERRA
PG&E
MIDWAY - LOS BANOS
NEVADA
EAST OF COLORADO RIVER (EOR)
NORTHERN - SOUTHERN CALIFORNIA
SOCALIF
LADWP
ARIZONA
IMPERIALCA
SANDIEGO
bing

IntNumber	IntName	IntMW	IntMVAR	LimitMW	LimitA	LimitB	L
17	BORAH WEST	460.04424095	-94.93887424	2556.99996948	2556.99996948	0	0
19	BRIDGER WEST	1609.1381073	36.41466796	2400	2400	0	0
55	BROWNLEE EAST	1156.14356995	-51.33836269	1750	1750	0	0
81	CENTENNIAL	2032.32879639	306.99801445	3000	3000	0	0
50	CHOLLA - PINNACLE PEAK	685.15353203	-41.29390121	1200	1200	0	0
66	COI	3958.36143494	-310.19403934	4800	4800	-3675	0
54	CORONADO - SILVER KING - KYRENE	875.72784424	-39.85942304	1493.99995804	1493.99995804	0	0
77	CRYSTAL - ALLEN	242.6574707	26.67178512	950	950	0	0
49	EAST OF COLORADO RIVER (EOR)	4248.79150391	-565.48442841	9300	9300	0	0
62	ELDORADO - MCCULLOUGH 500 KV	-25.99830532	-401.38106346	2597.99995422	2597.99995422	-2597.99995422	0
58	ELDORADO - MEAD 230 KV LINES	-177.81471014	-33.26287866	1139.99996185	1139.99996185	-1139.99996185	0
23	FOUR CORNERS 345/500	24.01234061	117.73771048	839.99996185	839.99996185	-839.99996185	0
14	IDAHO - NORTHWEST	-85.21904945	-117.24513769	2400	2400	-1200	0
16	IDAHO - SIERRA	3.65715995	5.50845675	500	500	-359.99999046	0
42	IID - SCE	324.48871136	-115.42663574	600	600	0	0

```
(12/06/2014 10:30): Oracle Database....Conne
Number of available scenarios: 18

case 14hsp1ae.epc loaded...
Number of Buses: 19729
Number of Generators: 3778
Number of Areas: 21
Number of Interfaces: 67
Number of Contingencies: 4
(12/06/2014 10:30): Calculating PTDF....

base case ......completed!
Contingency 1......completed!
Contingency 2......completed!
Contingency 3......completed!
Contingency 4......completed!

(12/06/2014 10:30): Oracle Database....Conne
Number of available scenarios: 18


(12/06/2014 10:30): Runing Monte-Carlo...

Parallel Computing Mode....The number of pr
Monte-Carlo (base case) ......completed!
Monte-Carlo (Contingency 1) ......completed!
Monte-Carlo (Contingency 2) ......completed!
Monte-Carlo (Contingency 3) ......completed!
Monte-Carlo (Contingency 4) ......completed!
Elapsed Time for power flow calculation using

(12/06/2014 10:30): Runing statistical analysis
......completed! (Elapsed Time=313 ms)

(12/06/2014 10:30): Runing Voltage Stability A
VSA completed....Elapsed Time for VSA=23 m
```

Figure 72. Probabilistic analysis of transmission flows in WECC system

The developed methodology and software tool have several very important advantages and opportunities for the BAs and other TSOs:

1. *Better quantification of available security margins*. Because the analyzed transmission impacts are caused by random variations of forecast errors in different parts of the system, they are not predictable in a deterministic sense. Based on a statistical analysis of multiple forecast errors, the tool provides a unique opportunity to adjust security margins depending on the risk (expected size and probability) of potential transmission violations.

2. *Better reliability level*. By adjusting the system security margins on critical paths in the system, the tool helps to prevent potential violations caused by random variations of system load and variable generation around their forecasted values.

3. *Better utilization of transmission assets*. In cases when the deterministic security margin is excessive, the tool will provide recommendations to reduce this margin based on the actual variability of the flows in the analyzed critical paths.

4. *Better situational awareness and predictive system monitoring*. The TUT algorithm is run for multiple look-ahead dispatch intervals and possible contingencies. Based on this information, system dispatchers will be informed about (1) potential violations and associated risks on all critical paths in the system; (2) the most critical contingencies; and (3) the expected time to violations.

5. *Preventive control*. As a result of its look-ahead feature, the TUT algorithm leaves some time for mitigation measures, helping to reduce the expected size and probability of violations to an acceptable level.

10.0 Parallel Implementation of Transmission Uncertainty and Prediction Tool

The Transmission Uncertainty and Prediction Tool (TUT) was developed to quantify the uncertainty of interface flows when power injection in each area appears stochastic properties. Prediction of transmission uncertainty in practical large-scale systems with large numbers of contingencies is prohibitively computationally intensive. The primary objectives of this task include: (1) analysis of computational cost of algorithms in TUT; and (2) design and implementation of parallel TUT to speed up the computation. The calculation in power transfer function distribution factor (PTDF) and Monte Carlo (MC) could be very time consuming, depending on the system size and number of MC. For example, for WECC system with 19279 buses and 3778 generators, when there are 100 MCs, the total simulation time is around 45 seconds, where PTDF calculation accounts for more than 80% of total time. When the MC number is increased to 2400, the total simulation time is 175 seconds, where PTDF calculation only accounts for about 23% of total time. We have explored parallel implementation of PTDF and MC simulation, and found that TUT with parallel PTDF calculation exhibits much improved performance compared to its sequential version, about 2 to 3 times speedup with 4 threads on a desktop.

10.1 Computational Cost Analysis of Algorithms in TUT

Computation in TUT consists of two modules: i) PTDF, and ii) probability density function (PDF) of interface flows.

In the first module, distribution factor on each interface flow as a function of each area generation is calculated for the base case and each contingency. PTDF is calculated for each area and interface flow. Let NC denote the number of contingencies (including the base case), NI denote the number of interface, and NA denote the number of areas. After PTDF calculation, we obtain a PTDF matrix (NI by NA) for each contingency, resulting NC such kind of matrix in total.

In the second module, PDF of interface flow is calculated through Monte-Carlo runs, which are used to capture the uncertainties of area power injection. The calculation can be divided into two parts. The first part takes the area power injection vector in each Monte-Carlo run together with PTDFs as inputs. PDF is calculated for each interface in each contingency. The second part combines PDFs under all contingencies to calculate the final PDF for each interface flow.

10.1.1 PTDF Calculation

PTDF is the incremental distribution factors associated with change in area power injection. These values provide a linearized approximation of flow changes of transmission lines and interfaces in response to change in area power injection. PTDFs may be calculated using either the full power flow Jacobian or only a portion of it. In current version of TUT, PTDF is obtained by calling PowerWorld®, which automatically generates PTDFs based on linearized AC approximation, lossless DC approximation, or lossless DC with phase shifters approximation. This process is illustrated in Figure 73.

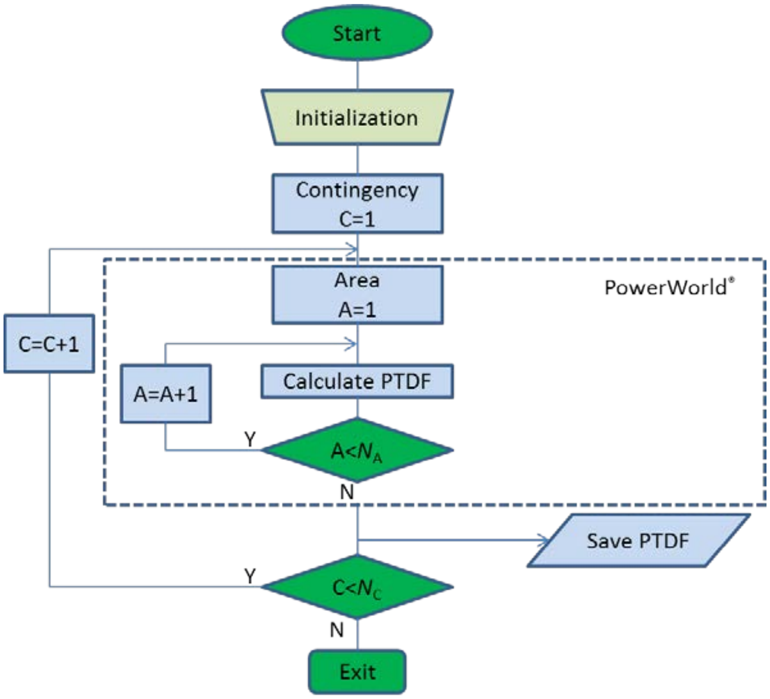


Figure 73. Flow chart of PTDF calculation calling PowerWorld®

In PowerWorld®, PTDFs are calculated using factored power flow Jacobian,

$$\Delta \mathbf{x} = J(\mathbf{x}^*)^{-1} \Delta \mathbf{P} \tag{1}$$

where $\mathbf{x}^*$ is the based operating point from power flow solution, $\Delta\mathbf{P}$ is the change in power injections associated with power transfer, and $\Delta\mathbf{x}$ is the change of system voltages and angles with are used to calculate flows. The PTDF calculation for each area in PowerWorld assumes the generators in this region/area participate according to their participation factors. The power flow solution to find $\mathbf{x}^*$ and computation in (1) dominate the overall PTDF calculation. The computational complexity is O(n^2), which n is size of Jacobian matrix. This computation will be performed N_C times for all the contingencies.

10.1.2 Interface Flow PDF Calculation

The PDF calculation can be divided into two parts. In the first part, PDF of each interface flow is calculated for each contingency, as illustrated in Figure 74. The overall PDF of each interface flow is then calculated by combining the corresponding PDFs over all contingencies, as illustrated in Figure 75. Both parts essentially perform matrix vector multiplication and addition. The computation complexities are N_A*N_I *N_{MC}*N_C and N_I *NMC*N_C, respectively. It is obvious that the first part of PDF calculation is more time consuming than the second part when is N_A big. As for comparison between computational time between PTDF and PDF, it depends on the how big n is compared with N_A, N_I , and N_{MC}. Both PTDF and PDF calculation can dominate the total simulation time.

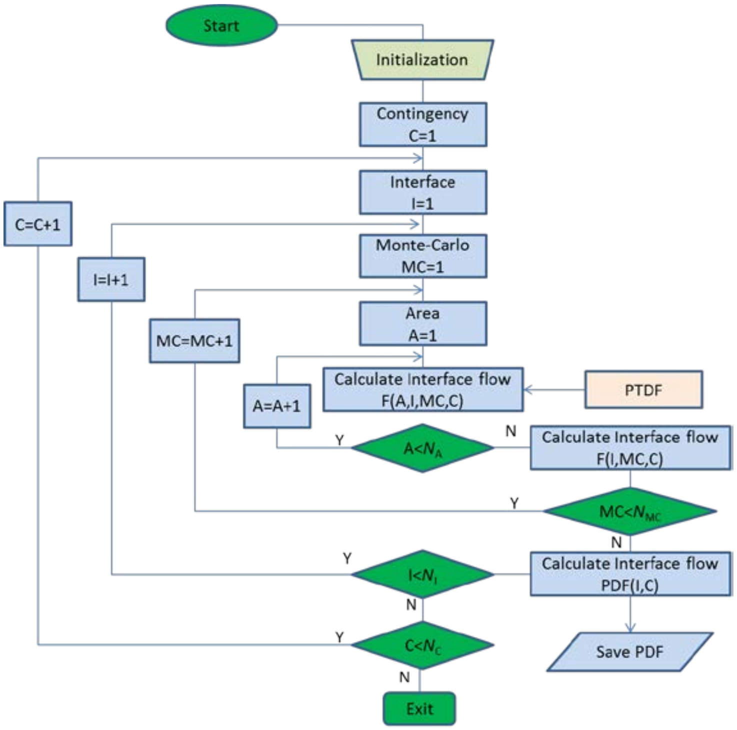

Start
Initialization
Contingency
C=1
Interface
I=1
C=C+1
Monte-Carlo
MC=1
I=I+1
Area
A=1
MC=MC+1
Calculate Interface flow
F(A,I,MC,C)
PTDF
A=A+1
Y
A<N_A
N
Calculate Interface flow
F(I,MC,C)
Y
MC<N_MC
N
Y
I<N_I
Calculate Interface flow
PDF(I,C)
N
Y
C<N_C
Save PDF
N
Exit

Figure 74. Flow chart of PDF calculation: PFD of interface flow in each contingency

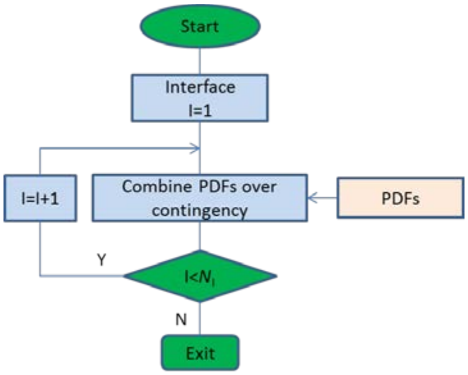

Start
Interface
I=1
I=I+1
Combine PDFs over
contingency
PDFs
Y
I<N_I
N
Exit

Figure 75. Flow chart of PDF calculation: Combine PDF over all contingencies

10.2 Parallel Implementation

The parallel implementation of matrix vector operation in the first part of PDF calculation is straight forward. It can be realized using parallel mechanism provided by visual basic.net. PTDF computation is performed by calling PowerWorld®. Its parallelization requires SimAuto® function provided by PowerWorld®.

PowerWorld® provides distributed computing add-on which takes advantage of modern multi-processor computers to reduce computation time. This is exactly the functionality required to compute PTDF on a multi-processor shared memory machine. The distributed computing add-on relies on SimAuto® on different processors and a master process to communicate with these processes using Microsoft® Distributed COM, as shown in Figure 76. However, an extra license is needed for this add-on.

In order to avoid additional cost on distributed computing add-on, a similar framework is proposed to perform parallel computing, as shown in Figure 77. One limitation of this method is that multiple instances of PowerWorld® are created, and each instance needs to maintain its own input and output, which increase the memory usage.

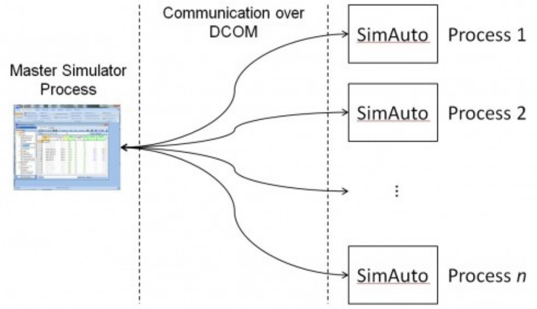

Communication over DCOM
Master Simulator Process
SimAuto
Process 1
SimAuto
Process 2
⋮
SimAuto
Process n

Figure 76. Parallel computation of PTDF through SimAuto®, with distributed computing add-on

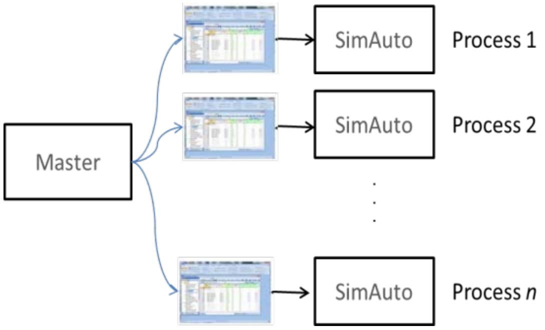
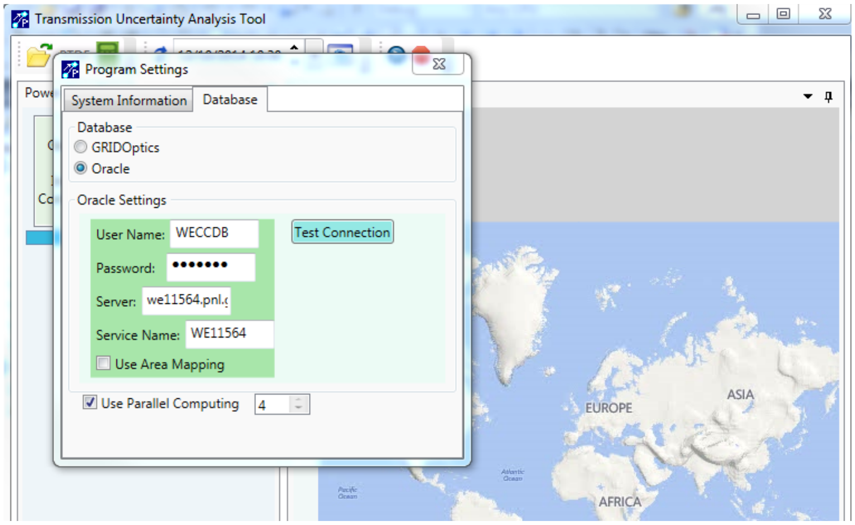

Master
SimAuto
Process 1
SimAuto
Process 2
SimAuto
Process n

Figure 77. Parallel computation of PTDF through SimAuto®, without distributed computing add-on

The tool also enables users to select number of thread for parallel computation. This parameter can be set in "Program Settings", as shown in Figure 78. The maximum number of threads which can be used is limited by the available threads in a computer, which is reflected in the list box.



Transmission Uncertainty Analysis Tool

Program Settings

Database
○ GRIDOptics
● Oracle

Oracle Settings
User Name: WECCDB
Password: •••••••
Server: we11564.pnl.
Service Name: WE11564
☐ Use Area Mapping

☑ Use Parallel Computing 4

Figure 78. Control items to parallel PTDF computation

10.3 Test Results

Western interconnection system with 19279 buses, 3778 generators, 21 areas, 67 interfaces and 40 contingencies is used for case study. The number of MC is set to 2400. A series of tests is performed using the parallel TUT on a laptop with i7-3740QM CPU and 16GB memory. Using 2 processes can provide approximately 1.4 fold speed up and using 4 processes can provide approximately 2 fold speed up for PTDF and more than 3 times for MC. The computation time and speedup for serial and parallel (with 4 threads) implementation are shown in Table 6. The PTDF computational time as a function of thread number is shown in Figure 79.

Table 6. Computation time of PTDF and Monte Carlo

	PTDF	**Monte Carlo**
Serial implementation (seconds)	38.39	130.15
Parallel implementation using 4 threads (seconds)	18.42	42.56
Speedup	2.08	3.06

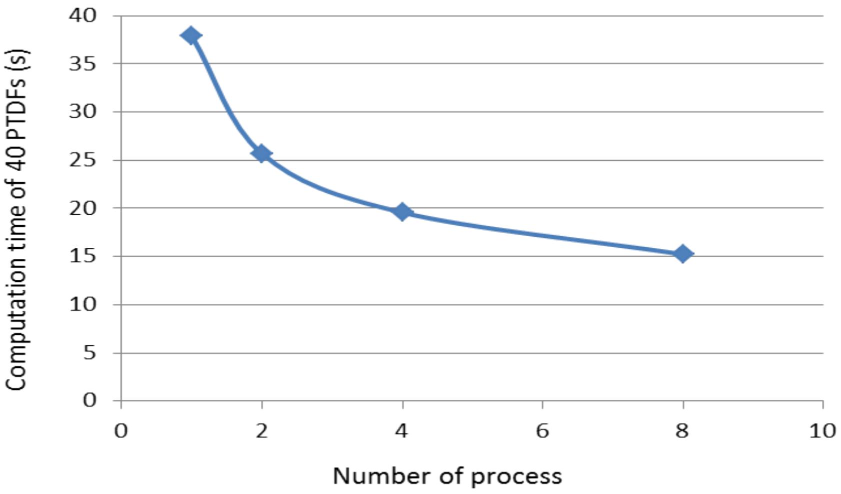


Figure 79. Performance of parallel PTDF computation

10.4 Future Work

In current parallel version of TUT, the calculation of PTDF is realized through PowerWorld®. One disadvantage is multiple instances of PowerWorld® are used. In addition, the speed is also limited by the computing power of desktop where PowerWorld® is running. In order to further improve the speed by using high performance computing, a script in FORTRAN or C++ is desired to perform PTDF calculation.

11.0 Conclusions and Future Work

11.1 Conclusions

1. A zonal geographically distributed model was developed for the Western Electricity Coordinating Council (WECC) system reflecting various sources of uncertainty in their interaction (load, wind and solar generation). The model forms a basis for developing and testing probabilistic methods.

2. Our approach based on separation of slow and fast power system motions (represented as net load variations = load variations – wind variations – solar variations) and associated uncertainties using probabilistic methods helps to improve predictability and reduce uncertainty in the system. Slower motions of stochastic processes are more predictable due to more autocorrelation between the subsequent points. By employing this autocorrelation, it is possible to improve the forecast of parameters subject to uncertainty. By quantifying uncertainty around the improved forecast, it is possible to reduce the range of uncertainty.

3. Connection of uncertainty with system motions. A power system maintains balance between its generation, changing load, variable generation, and interchange. Increasing uncertainty means more balancing work to be done by conventional generation, energy storage, and controllable load. In other words, these resources have to move more with increasing uncertainty and variability. Because the balancing service is a paid service, more system motion results in additional costs. The statistical separation of fast and slow motions helps to better distribute the balancing job between slow and fast balancing resources; so that they are more efficiently used and the cost is reduced (the fast-responsive resources are usually more expensive).

4. In this project, advanced statistical methods were applied to separate slow, fast and very fast system motions caused by variation of the net system load. The very fast motion is actually white noise, which is a completely unpredictable component of a signal. If the forecast error is exactly white noise, this is a perfect forecast. The advanced statistical methods applied were:

    d. Wavelet transforms – ARIMA method. Wavelet transforms were used to extract the slower component in the total load forecast error. The ARIMA model was used to predict this slower component. We demonstrated an uncertainty reduction of 10-12% by applying this combination of methods.

    e. We demonstrated, that the proposed decision tree regression analysis method helps to reduce the uncertainty (and associated balancing effort) up to 40%, and that the residual (very fast motion component) becomes close to white noise.

    f. Principal component analysis (PCA). By applying this method, we employ cross-correlation between the total load forecast errors in different zones of the WECC system. Thai is a new idea. The PCA helps to reduce the number of dimensions, where the predictable forecast components of multiple zonal errors can be located. This helps to improve the accuracy of the forecasts, and make the residual error close to white noise.

5. We developed and demonstrated new methods for quantification and adjustment of transmission limits using probabilistic methods (in the WECC system). Using the Transmission Uncertainty and Prediction Tool (TUT), we demonstrated how the transmission limits in the WECC system can be quantified and adjusted to minimize the risk of violations caused by uncertainty. This adjustment can also help to increase the utilization of transmission facilities within capacity and reliability limits.

6. PNNL Transmission Uncertainty and Prediction Tool (TUT). Interactions among wind, solar gen load forecast errors can have a random impact on transmission system. This event can lead to additional congestion and voltage stability problems. To mitigate this problem, Balancing Authorities (BAs) should be able to re-dispatch their conventional generators ahead of time, if such impact can be predicted. A new probabilistic methodology and new analytical software called the Transmission Uncertainty and Prediction Tool (TUT) was developed at Pacific Northwest National Laboratory using DOE OE, DOE EERE, and CEC funding. The developed methodology predicts the impact of the uncertainties on congested paths, worst-case voltage drops, and reactive power margins for several hours ahead of time and, if needed, proposes control actions to mitigate the problem if necessary.

7. We demonstrated how TUT can benefit from parallelization.

11.2 Recommendations for Future Work

Based on the first successful phase of the project, the following recommendations can be made for the near-term and long-term future work.

11.2.1 Proposed Fiscal Year 2015 Work

- Move the TUT to a near production level of development. The Tool will be transformed into a "fully" standalone application. Currently we need to use Powerworld to calculate several characteristics needed for simulations. Several capabilities will be added to eliminate this dependency and generate a standalone tool:
    a. Power Transfer Distribution Factors (PTDF) calculation capabilities.
    b. PV curves computation. There is a potential synergy with the non-iterative voltage stability project.
    c. Usage of parallel computational technique to speed up the process.
    d. Contingency analysis.
    e. Methodology to calculate probabilities of contingencies.
- Stochastic optimization and control. With the increasing uncertainty, the role and implementation of optimization and control algorithms is changing. Instead of finding a single global minimum, the optimization algorithm should search for a solution that is covering the uncertainty range. Instead of moving a single operating point to the desired region, one should move all points contained in the uncertainty region. The project will reformulate key optimization and control problems to reflect the random nature of the new power systems.

- Tail events quantification and assessment. The tail vents are observed whenever the system state exceeds the limits suitable for the “mainstream” power system operations. The tail events are not very well studied or addressed in the modern system planning and operation. The proposed task will be one of the first efforts to quantify these events, predict them, and develop uncertainty-based controls to avoid them.

11.2.2 Long-term Recommendations – Big Picture

The purpose of this long-term effort will be to attract interest from industry organizations and vendors to make a subsequent implementation project feasible. Ultimately, this project could lead to an industry-wide effort (a major nation-wide initiative) to transform grid planning and operation from the existing deterministic platform to the new probabilistic platform. As a big picture, we envisage the following follow up steps:

- Develop a comprehensive framework and roadmap for probabilistic planning and operations based on state-of-the art criteria, methodologies, software tools, and technologies
- Cooperate closely with the industry, regulators, Universities, government, national labs, software vendors, and other interested organizations to make sure that the best ideas, know-how and skills are reflected in the framework and the roadmap
- Initiate and lead a nation-wide effort for implementing this framework within the next 5 years as a standard practice in planning and operations
- Facilitate continuing education, dissemination, and technology transfer in the area of probabilistic methods and applications
- Create and lead a Probabilistic Technology Interest Group (PTIG) as a tool to implement the Roadmap and forum for organization interested in the area. The Group could be based on UWIG (membership fees) or NASPI models (funded by DOE).

11.3 Publications

This project produced the following publications:

Etingov P.V, Makarov Y.V., Vyakaranam B., Wu D., Meng D., Huang Z., and Kannberg L. “Predicting and Mitigating Wind Generation Impact on the Transmission Network,” AWEA WindPower 2015 Conference, Orlando, FL, May 18-21, 2015 (submitted).

Zhang Y, R Diao, YV Makarov, and Z Hou. 2014. " Improving System Predictability Using Statistical Separation of Slow, Fast and Very Fast System Motions." In 48th Annual HICSS conference January 5-8, 2015 on the island of Kauai, HI. PNNL-SA-103477, Pacific Northwest National Laboratory, Richland, WA, (submitted).

Y. Makarov, P. Etingov, J Ma. “Incorporating Forecast Uncertainty in Utility Control Center”. Book chapter in L.E. Jones (Ed.), Renewable Energy Integration: practical management of variability, uncertainty and flexibility in power grids. Academic Press Elsevier, 2014.

Hou Z, YV Makarov, PV Etingov and NA Samaan, " Uncertainty Reduction in Power Generation Forecast Using Coupled Wavelet-ARIMA," in the Proc. of in the Proc. of IEEE/PES General Meeting, Washington DC, July 2014.

Ryan Hafen, Nader Samaan, Ruisheng Duo, Yuri Makarov and Ning Lu, "Joint Seasonal ARMA Approach for Modeling of Load Forecast Errors in Planning Studies," Accepted in the Proc. of IEEE/PES Transmission and Distribution Conference and Exposition, Chicago, April 2014

Ning Lu, Ruisheng Diao, Ryan Hafen, Nader Samaan, and Yuri Makarov, “A Comparison of Forecast Error Generators for Modeling Wind and Load Uncertainty,” in the Proc. of IEEE/PES General Meeting, Vancouver, Canada, July 2013.

Hou Z, YV Makarov, NA Samaan, and PV Etingov, "Standardized Software for Wind Load Forecast Error Analyses and Predictions Based on Wavelet-ARIMA Models - Applications at Multiple Geographically Distributed Wind Farms." in the Proc. of the 46th Hawaii International Conference on System Sciences (HICSS), January 7-10, 2013, pp. 5005-5011

On the Configuration of the US Western Interconnection Voltage Stability Boundary
Makarov YV, B Vyakaranam, D Wu, Z Hou, ST Elbert, and Z Huang. 2014. "On the Configuration of the US Western Interconnection Voltage Stability Boundary." In IEEE PES Transmission and Distribution Conference and Exposition, April 14-17, 2014, Chicago, Illinois, pp. 1-5. IEEE, Piscataway, NJ. doi:10.1109/TDC.2014.6863289

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by Battelle *Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)
www.pnnl.gov