SANDIA REPORT

SAND2013-7886
Unlimited Release
Printed September 2013

Quantifiably Secure Power Grid Operation, Management, and Evolution: A Study of Uncertainties Affecting the Grid Integration of Renewables

G. A. Gray, J-P. Watson, C. Silva, R. Gramacy

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: 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 any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent 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, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2013-7886
Unlimited Release
Printed September 2013

Quantifiably Secure Power Grid Operation, Management, and Evolution: A Study of Uncertainties Affecting the Grid Integration of Renewables

Genetha Anne Gray
Sandia National Laboratories
P.O. Box 969, Livermore, CA 94551-0969
gagray@sandia.gov

Jean-Paul Watson
Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM 87185-9999
jpwatson@sandia.gov

Cesar Silva Monroy
Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM 87185-9999
casilv@sandia.gov

Robert Gramacy
University of Chicago, Booth School of Business
5807 South Woodlawn Ave, Chicago, IL 60637
rbgramacy@ChicagoBooth.edu

Abstract

This report summarizes findings and results of the Quantifiably Secure Power Grid Operation, Management, and Evolution LDRD. The focus of the LDRD was to develop decision-support technologies to enable rational and quantifiable risk management for two key grid

operational timescales: scheduling (day-ahead) and planning (month-to-year-ahead). Risk or resiliency metrics are foundational in this effort. The 2003 Northeast Blackout investigative report stressed the criticality of enforceable metrics for system resiliency – the grid's ability to satisfy demands subject to perturbation. However, we neither have well-defined risk metrics for addressing the pervasive uncertainties in a renewable energy era, nor decision-support tools for their enforcement, which severely impacts efforts to rationally improve grid security. For day-ahead unit commitment, decision-support tools must account for topological security constraints, loss-of-load (economic) costs, and supply and demand variability – especially given high renewables penetration. For long-term planning, transmission and generation expansion must ensure realized demand is satisfied for various projected technological, climate, and growth scenarios.

The decision-support tools investigated in this project paid particular attention to tail-oriented risk metrics for explicitly addressing high-consequence events. Historically, decision-support tools for the grid consider expected cost minimization, largely ignoring risk and instead penalizing loss-of-load through artificial parameters. The technical focus of this work was the development of scalable solvers for enforcing risk metrics. Advanced stochastic programming solvers were developed to address generation and transmission expansion and unit commitment, minimizing cost subject to pre-specified risk thresholds. Particular attention was paid to renewables where security critically depends on production and demand prediction accuracy. To address this concern, powerful filtering techniques for spatio-temporal measurement assimilation were used to develop short-term predictive stochastic models. To achieve uncertainty-tolerant solutions, very large numbers of scenarios must be simultaneously considered. One focus of this work was investigating ways of reasonably reducing this number.

Acknowledgment

The authors would like to thank Patty Hough, Ali Pinar, Rich Chen and Cliff Hanson for their contributions to this work. We further thank the LDRD office for their funding support.

Contents

1 Motivation and Background 9
2 Some of the Basics of Uncertainty and Variability 11
2.1 Uncertainty 11
2.2 Variability 12
3 Unit Commitment 13
3.1 Data 13
3.2 Forecasting Models 14
3.3 Comparing Scenario Generation Methods 17
4 Expansion Planning 21
4.1 Current Practices 21
4.2 Approaches in Research & Development 22
5 Conclusions and Future Work 25
References 27

1 Motivation and Background

In the United States, the electrical grid has changed very little in the last 100 years, and in fact still works mainly on the principle of one-way flow of consumption [15]. The drivers for the new smart grid include new energy generation methods, the incorporation of renewables, balanced loads and reduced peaking, improved reliability and security, and the desire for a two-way flow system. To address these goals, solutions to some fundamental short-term operations and long-term planning problems in power systems are needed. However, these problems must be addressed given uncertain information. For example, the natural fluctuations in wind speeds have caused concerns about the integration of wind power. Questions have arisen about the stability of the grid system in the face of its intermittent capacity. Despite such concerns, policies have been written to integrate renewables into existing electrical grids. Therefore, planning activities must focus on techniques that minimize disruptions and account for uncertainties. In this project, we investigate some of the issues associated with uncertainties and variabilities associated with wind and solar energy and suggest some techniques to overcome them.

One of the most basic problems in electrical grid planning and operations is unit commitment (UC). The UC problem is to determine an optimal on/off schedule for a set of power generating units that both meets load demands and satisfies operational constraints. A UC plan is considered for both short (e.g., hours or days) and long (e.g., weeks or months) time horizons. Uncertainties resulting from load forecasts, network outages, and discrete events must be considered in order to make a robust UC schedule. For example, UC can be examined with uncertainty due to wind speeds or solar availabilities. This is a fundamental power systems operational problem that can be formulated as a mixed integer stochastic optimization problem (See for example [19] and references therein). Another such problem in electrical grid planning and operations is network expansion. Here, the focus is determining how best to upgrade the system in order to meet future demands. Like UC, network expansion can be posed as a mixed integer stochastic optimization problem. In planning contexts, like all generation and transmission expansion problems, inherent uncertainties in future demand, budgets, and technologies add to the challenge of obtaining a robust solution.

The computational challenges associated with solving stochastic mixed-integer problems like the UC and the network expansion problems are significant. There are two primary and related factors [1]. First, the (finite) number of scenarios required to approximate the joint distributions of uncertain parameters leads to notoriously difficult deterministic mixed-integer optimization problems. Second, unrealistic modeling simplifications are often required to achieve computationally tractability, leading to more costly and potentially infeasible solutions. High-performance computing technologies have been proposed to mitigate both concerns. As part of this project, we reviewed these technologies and presented some results for their application to the electrical grid planning and operation problems.

Another problem of great interest is that of grid expansion. In the US, grid expansion is needed to meet the needs of customers, to incorporate renewables, to increase surety, to provide security of resources, and to upgrade the current infrastructure. In this project, particular interest was paid to the addition of solar farms to the grids. Photovoltaics, the conversion of solar radiation into direct current electricity, has advanced significantly in recent years and shows great promise to aid

grid expansion. To make decisions about where to invest in new solar projects, stakeholders must calculate projected annual energy projections (AEPs). AEPs are dependent on solar irradiation data which can vary greatly both temporally and spatially. In this project, we investigated the root of these issues and how the uncertainties could be incorporated into the larger problems of grid expansion planning and long-term UC.

It should be noted that there is a wide range of problems related to grid operations and planning problems. This report addresses only two such problems and focuses on dealing with the variabilities and uncertainties related to these problems. There exists a vast array of literature and ongoing research about the incorporation of renewables and such an effort is needed to solve the problems (see for example [22, 14] and references therein). This LDRD project and subsequent report focuses on a very small subset of these problems and is arranged as follows: In Section 2, we discuss the mathematical and statistical basics needed to examine and address uncertainty and variability. Then, in Section 3, we consider the problem of unit commitment and the inclusion of load demand and wind supply uncertainties. Next, in Section 4, we examine the problem of expansion planning in terms of adding solar plants given resource uncertainty. Finally, in Section 5, we draw some conclusions and describe possible future directions and extensions of this work.

2 Some of the Basics of Uncertainty and Variability

Operating, managing and evolving the grid is done is a decision making environment. In general, all decision environments possess a common set of constituent elements as noted in [9]. They are as follows:

Step 1: A decision maker is committed to making a decision.

Step 2: An objective is identified (although in the context of complex decision environments, there might be multiple objectives).

Step 3: A finite number of mutually exclusive events or scenarios are identified. (Note that in complex environments, the scenarios may not be mutually exclusive).

Step 4: A finite number of decision alternatives or actions are identified.

Step 5: Outcomes for each scenario and action are identified.

Step 6: Quantified uncertainty is applied and propagated (or communicated) for each alternative.

Step 7: A return on the decision alternatives (also called the payoff) is determined.

Step 8: A best alternative is chosen and enacted.

These elements incorporate many different decision criteria such as available resources and specification requirements. Uncertainty is specifically called out in step 6 of the decision making framework. However, uncertainty permeates the entire decision making process. How uncertainty is represented, incorporated, and propagated within scenarios, their underlying data, and along the decision-making path itself is challenging; and novel mathematics is required to address it appropriately. This project investigated some of the issues surrounding uncertainty as they relate to operating, managing, and evolving the electrical grid.

2.1 Uncertainty

There are two basic types of uncertainties- aleatory and epistemic [7]. Aleatory uncertainty is the inherent variability in an object or system. It cannot be removed or reduced, but can be represented probabilistically if enough data are present. Epistemic uncertainty is a subjective uncertainty that results from lack of knowledge. It is also referred to as reducible or model-form uncertainty because the addition of data can increase knowledge and improve underlying models.

Uncertainty quantification (UQ) is a fundamental tool for determining how likely certain outcomes would be given that certain aspects of the system model are unknown. UQ is used to identify, characterize, reduce, and, if possible, eliminate uncertainties. For example, weather forecasts use UQ to define a percentage chance that a certain outcome might occur. (*i.e.* ”There is a 75% chance

of rain.") UQ techniques are used to analyze both computational and experimental data and may be used to compare the two sets. In fact, the effective identification, definition, quantification and communication of uncertainties is often the most important aspect of model development, and it can have a significant impact on the overall success of the modeling activity [7, 13, 21].

2.2 Variability

Variability is used to describe how clustered or spread out a data set is. It is related to uncertainty in that it can be used to describe the range of outcomes. For example, for a given variety of consumer behaviors, a range of operational scenarios can be generated. There is uncertainty in the consumer behavior and it is represented by its variability.

There are a number of basic measures of variability that can be useful in analyzing data sets. They include:

- *mean*: In this context, the mean is merely the simple average value of the data set. It is calculated as the sum of all the values divided by the number of values.
- *range*: The range measures the difference between the lowest and highest values in the data set. For example, the temperature on the warmest day of the year may have been 103 degrees and the temperature on the coldest day may have been 31. Then, the range of temperatures for the year would be 72.
- *interquartile range*: The interquartile range elucidates the spread of the central 50% of the data. It compares the data at the 25th percentile with that of the 75 percentile. This eliminates outliers and gives the analyst a better idea of the overall dataset. Note that in the case of grid related decision making activities, this value may not be useful as it is often the outlier values that concern operators.
- *variance*: The variance also measures how much a set of data is spread out. However, unlike the range and interquartile range, the variance considers every value in the set instead of just two of the values. It is calculated as the average of the squared differences from the mean.
- *standard deviation*: The standard deviation is the square root of the variance and is also a measure of the amount by which every value within a dataset varies from the mean. It is an indicator of how tightly the values in the dataset are clustered around the mean. Thus, if the values are densely centered near the mean, the standard deviation will be relatively small.

We encourage interested readers to consult a basic statistics text for more details.

3 Unit Commitment

The uncertainty and variability in the output of wind farms pose a significant challenge to grid integration. Differences between forecasted and actual volumes can arise on both the supply and the demand sides of the system. One the supply side, these include outages of generating plant (planned or unplanned) or distribution networks as well as unpredicted changes in wind speeds. On the demand side, these include changes in consumer usage due to unusual or rare events or weather patterns. In this project, we focused on these complications as they relate to the problem of unit commitment (UC), the least-cost dispatch of available resources to meet the electrical load.

To reduce the impacts of the uncertainties associated with renewables such as wind, the UC problem must be approached with sophisticated techniques. Here, we model and solve the corresponding operational problems stochastically. The optimization objective is to meet projected demand at a minimum expected cost, and the uncertainty in the renewables is captured in a set of possible, projected scenarios. We focus on analyzing electrical load demands and weather information to create an appropriate set of scenarios. The set balances the need to cover the wide range of variability with the need to minimize the computational workload and problem complexity.

3.1 Data

The data utilized in this work represents the Western Electricity Coordinating Council (WECC) which is illustrated in Figure 1. It includes one year of load data for its 39 service providers at a one hour time series resolution, wind farm output data at one hour intervals over one year period for 82 existing wind farms, and solar power output at one hour intervals for one year at 9 existing sites. Closer examination of this data indicates that the data is highly non-stationary (*e.g.* the distribution changes over time). Therefore, external predictors will be essential to improving forecasts.

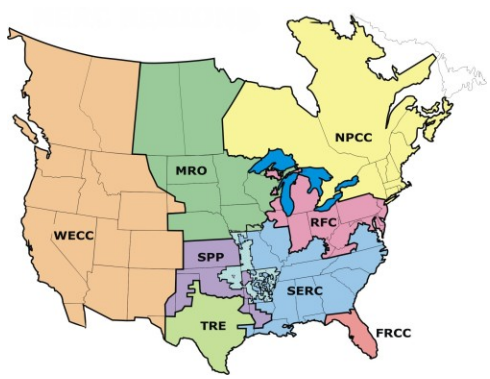

Figure 1: The Western Electricity Coordinating Council (WECC) is one of nine regional electric reliability councils under North American Electric Reliability Corporation (NERC) authority. WECC covers the Western United States and Western Canada.

Weather data is specified by longitude and latitude. While this is directly applicable to the

locations of the wind generation stations, it provides some challenges for the load domains. As illustrated in Figureẽfefservice, some service providers cover areas that are difficult to describe with one representative points. The area may contain widely varying climates. This is true of Pacific Gas & Electric in the Northern California Bay Area which covers both San Francisco and cities in the East Bay, and temperatures in these areas can vary by upwards of 20 degrees. Additionally, some service providers cover areas which vary widely with respect to population density. For example, the Canadian service provider areas correspond to territories, and the population of these territories is concentrated near the US border which is a small percentage of the total land area. Therefore, the weather data corresponding to a service provider may require range of values or a weighted metric to improve load forecasts. Also, the uncertainty of weather forecasts themselves must be a consideration. In addition to weather information, future work will improve predictions by incorporating information about the type of day (*e.g.* a weekend day, a holiday, or a normal weekday) as another external predictor.

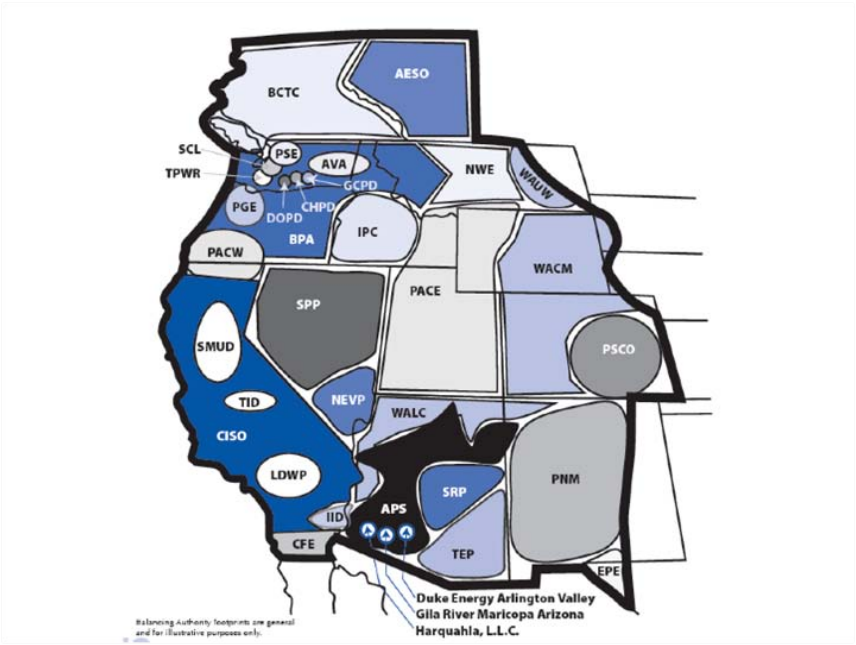

Figure 2: The WECC is composed of 39 service providers. Note that the service provider areas vary greatly.

3.2 Forecasting Models

Forecasting is the process of making predictions of future events. In this project, we use quantitative forecast as opposed to qualitative forecasting techniques. Qualitative forecasts rely on expert opinion and do not require past data. On the other hand, quantitative forecasts describe data associated with future events given past events. Two such techniques and their application to generating

a set of scenarios for use in solving the stochastic UC problem are described below.

ARIMA Model

An ARIMA (Auto-Regressive Integrated Moving Average) model is part of a general class of phenomenological time series models and is arguably the most commonly applied time series forecasting tool. It predicts future values of a time series using a combination of past observations and projected errors. ARIMA models were designed with the intent of capturing all the ways that a time series can evolve independent of seasonal terms and trends [2]. More specifically, an ARIMA model is define by the three-tuple (p,d,q) where p is the number of autoregressive terms, $d$is the number of nonseasonal differences, and q is the number of lagged forecast errors. These three parts can be further defined as:

- The Auto-Regressive (AR) part which allows future observations to depend linearly on a fixed number (p) of past iterations
- The Integrated (I) part which involves taking d-differences to potentially remove some non-stationarity.
- The Moving Average (MA) part which allows errors to accumulate linearly up to a finite lag (q)

Using this notation, some well-know examples include the $ARIMA(0,1,0)$ model which is a random walk, $ARIMA(0,1,1)$ which is the exponential smoothing model, and $ARIMA(1,1,0)$ which is the differenced first-order autoregressive model. There are a number of ARIMA model software packages available. Many of the ones in the statistical software package R [8] try to automate the selection of autoregressive terms (p) and lag (q) and decide on first differencing, via information criteria, while simultaneously estimating the known coefficients. Many packages also automate some kind of rudimentary de-seasonalization and de-trending.

In this work, we employed the basic ARIMA model in the forecast package in R [10]. In order to overcome data issues associated with frequency, we consider 30-day batches of data to predict the next 48 hours. In other words, to predict the load requirements, wind farm output, or solar farm output for days x and $x+1$, we consider the information for the days prior to x. As future work, we will consider longer time horizons, which are also important in grid integration problems. In addition, we selected outcomes that are within the 95% predictive interval. For generating scenarios from the load data, this approach was straightforward and sufficient. However, the wind and solar data provide some additional challenges. First, the outputs of the farms are bounded below by zero and above by the capacity of the farms. This must be incorporated into the model so that the results do not include infeasible scenarios. Second, the time series have numerous zeros and replicates that make standard statistical modeling difficult. In the case of the solar data, many of these zeros can be predicted as they correspond to times when the sun is down. For the wind data, this is not the case. As we continue this work, these data features must be addressed. Moreover, as this is a relatively unexplored area of statistical forecasting, any resulting modeling

process may lead to a technical contribution to the field at large. However, such an advance falls outside the scope of this LDRD and thus, in order to focus on demonstrating the usefulness of the stochastic formulation of the UC problem, we proceeded with a simplistic scenario generation process. To meet the bounds of the wind and solar data, all infeasible points of the 5% to 95% predictive interval are snapped to the feasibility boundary, and then the scenarios generated are generated so that they stay within this revised interval. Special consideration of the zeros and replicates was left for future work.

ARIMA models are powerful and widely used in practice, and using the implementation described above, we were able to generate a reasonable set of representative scenarios. However, ARIMA models are often highly automated and thus can usually be improved upon with a more hands-on approach. In fact, in our study, our specialized implementation of the ARIMA model allows the paths of the forecasted scenarios to bounce around unrealistically. Therefore, we opted to consider an alternative forecast model designed specifically to meet the characteristics of the grid-related data.

Specialized Load Forecast Model

The goals of the new forecast model were to generate scenarios more faithful to the predictive distribution, to respect the time structure by not sampling around the (moving) mean, and to significantly reduce the number of scenarios without losing variability and uncertainty information. In addition, we wanted to devise something that would more easily handle external predictors as this is critical to improving the forecasting as described earlier. This remainder of this section focuses mainly on scenario generation for the load data. However, a few comments are included to describe how this process could be applied to the wind and the solar data.

Popular forecasting models, such as ARIMA, rely on normality, so considering log values of positive data sets is reasonable and logical. Therefore, the first step in the modeling process for the load data is to take the log of the load values. This is corrected with exponentials at the end of the process. For the wind and the solar data, this step is skipped due to the many 0 data values. The next step is to describe a set of variables that appear to be useful for modeling the data. In this case, these included : autoregressive components of orders 1 and 2, autoregressive terms of order 12 and 24 to describe half-daily features, and sin and cos terms with periods of 12 and 24 to describe some smooth half-daily features. Then, an information criterion was applied to select the best subset of these 8 terms. Initial autocorrelated residual diagnostics indicated that lag-6 might be useful, so two more sin and cos terms with period 6 were added. After the model was selected (*e.g.* the variable values were selected), a simple Monte Carlo procedure was applied to simulate the time series forward and generate the scenarios. Note in the case of the wind and solar data, the information criterion can (and likely does) chose the components differently. However, the components may need to be modified to incorporate the daily features of wind and solar power generation. Moreover, the basic Monte Carlo process used here does not guarantee that the scenarios will be feasible. Therefore, as a simple fix, the results for wind and solar were linearly projected back into the interval $[0, \texttt{FM}]$ where $\texttt{FM}$ indicates the maximum capacity of the farm.

3.3 Comparing Scenario Generation Methods

Figure 3 shows a direct comparison of two scenarios of load over the next 48 hours. The one in black is generated by the ARIMA model and the one in red is generated by the specialized model. Note that the two methods forecast load profiles with the same basic trends, but the specialized model is much smoother.

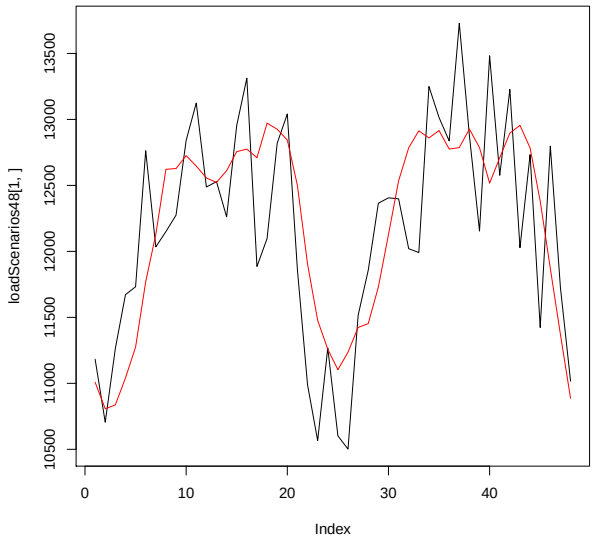

Figure 3: This plot illustrates the smoothness differences between a load scenario generated by and ARIMA model (in black) and our specilized model (in red). Note that both follow the same basic trend over the 48 hour forecast time period.

The coefficients of the specialized model are being studied to see how they can be used to determine how many regimes are present and thus how many scenarios are needed to describe the system. This is one of the advantages of using a the specialized model as opposed to the ARIMA model. For the study of the load data, we also considered how the overall variability was reduced with the addition of scenarios. A plot of this study is shown in Figure 4. Note that at 100 scenarios, the variability stops changing significantly. Therefore, we produced sets of 100 scenarios for use in some UC problems. Future work includes developing a computationally inexpensive metric to automate the process of determining an optimal number of scenarios.

Figure 5 gives some examples of how 100 scenarios generated by the ARIMA model compare with 100 scenarios generated by the specialized model. In the pictures, the ARIMA-model generated scenarios are in black and the specialized model scenarios are in red. Note that approximately 5% of the Specialized scenarios break out of the 95% confidence interval. Also, observe that there is no specific pattern for how these sets differ. In some cases (Subfigure 5a), the ARIMA scenarios are completed contained within in the bounds of the Specialized model scenarios while in other cases (Subfigure 5b), the opposite is true. Subfigure 5c shows a case where the two are very simi-

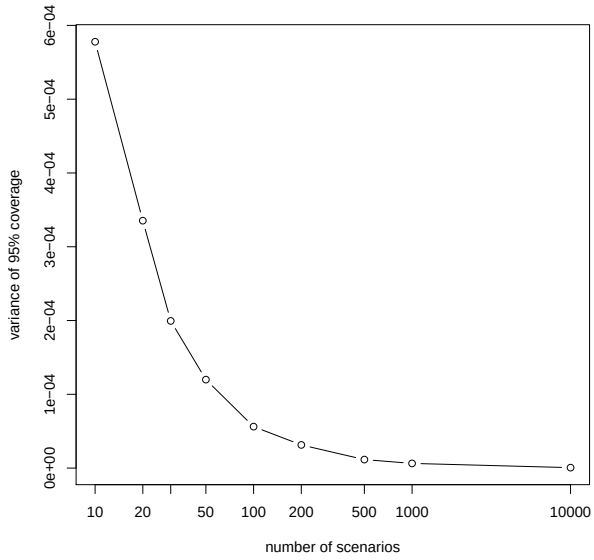

Figure 4: Using the specialized model, the variance represented by the scenarios (y-axis) is decreased as the number of scenarios in the set is increased (x-axis). This plot shows that using 100 load scenarios is sufficient for capturing the variance of interest.

lar except at the extremes of the curves, and Subfigure 5d shows a case where the ARIMA model scenarios seem to be shifted down slightly.

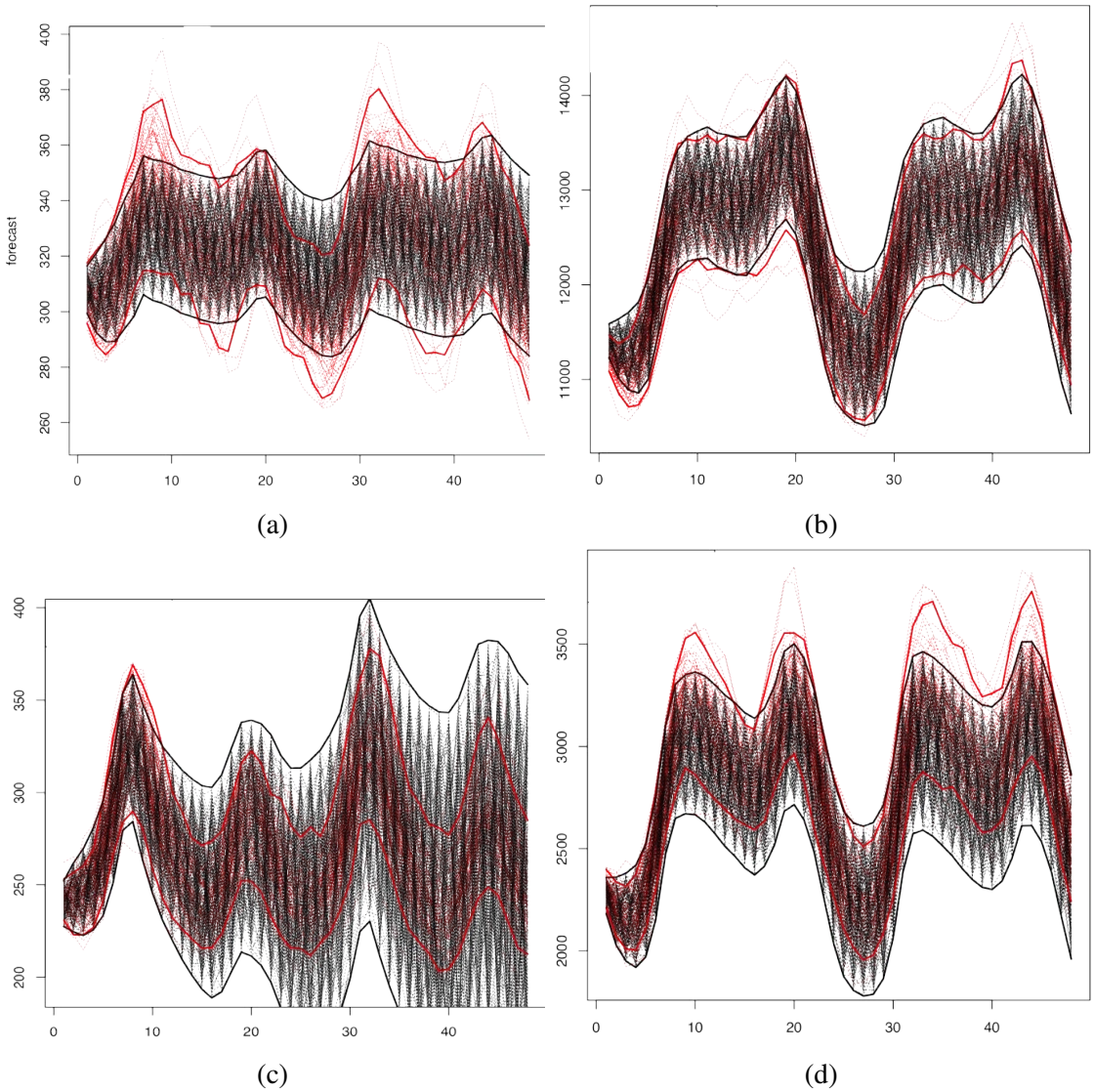

Figure 5: Some examples of how the sets of 100 scenarios generated by the ARIMA forecast model (in red) and the specialized forecast model (in black) compare with one another.

4 Expansion Planning

Expanding the grid is important for many reasons including meeting the demands of changing populations, satisfying changing energy requests of current populations (*i.e.* an increase in the use of electric cars or the building of new computer facilities), strengthening connections with outlying communities, incorporating designs that better accommodate fault tolerance, and facilitating the efficient use of renewable energy. In this project, we paid particular attention to the incorporation of renewable energy. More specifically, we studied the decision making process of erecting new solar facilities.

Decisions about financing solar projects pay particularly attention to risk assessments related to projected annual energy productions (AEPs). More specifically, AEP is used to calculate an exceedance probability [3]. For example, given an exceedance probability of P90, there is a 10% chance that this level will not be met. In other words, P90 is the AEP that can be expected in 9 out of 10 years. Determining the exceedance probability of a solar power system is highly dependent on annual insolation measurements. Therefore, understanding and quantifying uncertainties in the AEP requires correctly calculating the inter-annual variability of annual insolation. In this project, we have found that while the concept of year-to-year variability is recognized and discussed, there is no agreed upon value or methodology for calculating and using this value. To study this issue in more detail, we investigated approaches to the issue of inter-annual insolation variability throughout both the business (*i.e* companies that buy and sell components of solar power systems as well as those that consult on the design and use of such systems) and research communities.

4.1 Current Practices

This section describes how the business of risk assessment of solar power systems is currently carried out. It includes both small projects such as panel installation at single family home and larger projects.

In general, we observed that companies that sell components of solar power systems offer their customers give the high, low and average solar insolation for a year in a certain area. However, they do not indicate how they arrived at these values, and they rarely mention year-to-year variance. Instead, they tend to work with average numbers of full sun hours and geographical zones. The calculators available on the website of Wholesale Solar (`http://www.wholesalesolar.com/`) are an example of this practice.

Green Rhino Energy, a management consultant company specializing in PV, wind, and tidal power systems, specifically identifies ”natural fluctuations in annual irradiation” as a source of uncertainty in its AEP calculation. It goes on to describe the statistical process used to calculate the overall uncertainty in AEP using the basic statistical tools of standard deviation and correlation, and describes the resulting covariance matrix. However, no specific information with respect to calculating uncertainties in annual insolation is given. Moreover, the data used in these calculations is not identified. (See `http://www.greenrhinoenergy.com/` for more information)

A few management consultant companies (*i.e.* AWS True Power, GL Garrad Hassan, and Megajoule) specifically call out year-to-year variability as a source of irradiation uncertainty. AWS True Power classifies it as a solar resource estimate uncertainty that is between 5-17% and advocate on-site monitoring to mitigate risks. GL Garrad Hassan gives an example of the variability for Global Horizon Irradiation in France, but does not specifically describe how to account for this issue. Megajoule classifies resource uncertainty, which includes annual variation, as 6-12% of the overall project uncertainty and annual variability on its own as 1-4%. They advocate taking this into account when calculating exceedance probabilities but give no mathematical details. Moreover, no details are given as to how they arrived at these percentages.

Partners at the South African National Department of Energy and the Nelson Mandela Metropolitan University are designing a PV plant for the Eastern Cape of South Africa using a model based meteorological data. They use data from Meteonorm, a comprehensive climatological database. The base stochastically generates a statistical representation of a typical year, and the designers rely on this value to represent inter-annual insolation variability. Although uncertainties are specifically discussed with respect to the subsequent exceedance probability calculations, the year-to-year variability is incorporated solely by the use of the Meteonorm data. (See presentation by EE van Dyk) An article on the CSP Today Business Intelligence website in March 2011 also advocates the approach of using a typical year. The article focuses on building a better meteorological dataset to improve exceedance probability calculations [5].

4.2 Approaches in Research & Development

Note that there are many areas of science and engineering that use solar insolation data. A few examples include spacecraft design, understanding of equilibrium temperatures in planetology, the construction of climate-adapted buildings, and the runoff of fresh water from the snow pack to the water system. The literature describing these fields is rich and comprehensive. However, they suffer from the same issues found in the energy sciences, namely that there is no definitive methodology for calculating the uncertainty associated with annual insolation nor a general method for propagating the result through a predictive model that uses the mean value of annual insolation. There are publications that correlate annual insolation variability to annual water supply, annual precipitation, and annual atmospheric carbon levels. Other studies try to identify spatial and temporal patterns in year-to-year variations or to prove or disprove theories a bout the effects of climate on the environment. Many of the studies conclude that so-called "larger than expected" changes to environmental patterns were indeed observed. Such results are used to motivate future research and to implement risk-reduction projects. In general, this literature showed more results related to identifying insolation anomalies rather than calculating insolation uncertainties.

Much of the discussion related to inter-annual variability relates to the formation of a typical meteorological year (TMY). A TMY gives a full year of hourly weather data that is consistent with the long-term averages of the location of interest. The National Renewable Energy Laboratory has created the follow three such sets:

- TMY- based on data collected at 229 US locations between 1948 & 1980

- TMY2- based on data collected at 229 US locations between 1961 & 1990
- TMY3- based on data collected and 1020 US locations, both continental and territory (*i.e.* Guam, Puerto Rico & the US Virgin Islands), between 1976 & 2005 wherever available or from 1991 to 2005 otherwise [20]

Other TMY sets are available for specific locations provided by vendors of planning and assessment software. There are some calls by researchers to produce improved TMYs and decrease uncertainty in solar radiation averages [5]. Others use a TMY in combination with the statistical probability that an extreme event might occur. For example, in [6], the probability of a volcanic eruption is incorporated into to the typical data set.

The pros of using a TMY include its usefulness in comparisons of system configurations and locations in typical situations. Moreover, if a TMY is utilized, no additional data collection is required which can represent a significant savings. However, there is no strict standard for a TMY. Moreover, a TMY is merely an average and does not represent extreme or worst-case scenario ([3] and references therein). It does not guarantee average values will closely represent averages over time [18] and may only be available for a nearby location of interest instead of the actual location of interest which ignores microclimate behavior.

Given the negatives of using TMY, researchers have suggested creating multi-year historical models using existing data or with new data. This approach invites a number of questions including where to conduct the measurements and over what time frame, how to quality control the data, and how to handle missing or obviously erroneous readings. The most complete discussion of this topic is in [6]. In that work, data from three solar stations in Oregon and one in Colorado are used since they have the longest data records and consistent solar radiation measurements. To study inter-annual variability, the data was sorted from the best years to the average years and from the average years to the worst years. Among the conclusions of these studies were that the significant inter-annual variability of solar irradiation precludes the use of TMYs as appropriate to simulate AEP and that in the US, the highest variations of inter-annual variability were observed in the Pacific Northwest and lowest variations in the Southwest. Other studies on inter-annual variability in climate related analysis that do not use TMY include [16, 11, 12, 17, 4].

5 Conclusions and Future Work

This project elucidated a number of the issues associated with the uncertainty and variability associated with secure power grid operation, management and evolution. More specifically, we investigated how variabilities in weather and uncertainties in weather predictions can be mitigated in planning problems in order to improve management and expansion plans. We can draw a number of conclusions and make suggestions for future work.

First, we observed that forecast models play a critical role in addressing the uncertainties associated with unit commitment. Stochastic methods rely on scenarios to demonstrate the range of possible situations. There is certainty an art to creating and implementing forecast models, and these models can be improved with specialized components. Although the specialized model was shown to work well for the UC problem of interest, it can be improved. Specifically, we suggest the following continuations of this work:

- Include external predictors in the forecast model.
- Study the affect of the uncertainty of the external predictors on the overall system variability.
- Refine the Specialized model to meet the specific traits of the wind and the solar power generation scenarios.
- Consider sets of scenarios for longer time horizons.
- Create a general method that is both computationally inexpensive and easy to execute that determines an appropriate number of scenarios.

The scenarios play a key role in showing the computational tractability of the stochastic unit commitment methodology being developed in an ongoing ARPA-E project.

Second, we observed that the variability of weather from year-to-year adds a significant amount of uncertainty to the decision-making process of whether or not to build renewable energy plants. Currently, there is no proven process for defining the variability of annual irradiation. Moreover, the uncertainty in some of the currently used percentages may be reduced with additional studies and data. Given these results, the following continuations of this study seem appropriate:

- Identify and distinguish between the uncertainties in the inter-annual variability of the weather, the measurements of irradiance, and the model of translating irradiance to power.
- Compare inter-annual irradiance variabilities for some well defined data sets.
- Create an improved forecast model of solar plant output for use in the stochastic UC problem.
- Compare results for TMY data versus other data sources.

It should be noted that some of these tasks are will be carried out as part of a DOE project.

Finally, we note that all the problems discussed in this report are particularly difficult in a security-conscious, cost-constrained world where there is a need to quantify and enumerate the tradeoff between uncertain risks (e.g., grid stability, renewable penetration, and security threats) and the costs to mitigate those risks. This need motivates the development of multi-objective optimization of stochastic systems, but it also introduces a number of new challenges including quantification of the tradeoff curve itself, managing the potentially astronomical computational budget and educating decision makers in how to interpret and use the results.

References

[1] J. R. Birge and F. V. Louveaux. *Introduction to Stochastic Programming*. Springer Verlag, New York, 1997.

[2] G. E. P. Box and G. C. Tiao. Intervention analysis with applications to economic and environmental problems. *JASA*, 70:70–79, 1975.

[3] A. Dobos, P. Gilman, and M. Kasberg. P50/p90 analysis for solar energy systems using the system model advisor. Technical Report CP-6A20-5448, National Renewable Energy Laboratory, June 2012.

[4] K. Eerme. *Solar Radiation*, chapter Interannual and Intraseasonal Variations of the Available Solar Radiation. InTech, 2012.

[5] B. Gallego. Tmy typical meteorological year - what should it represent and what means p50, p70 and p90. *CSP Today*, March 2011.

[6] C. A. Gueymard and S. M. Wilcox. Assessment of spatial and temporal variability in the us solar resource from radiometric measurements and predictions from models using ground-based or satellite data. *Solar Energy*, 85:1068–1084, 2011.

[7] J. C. Helton. Treatment of uncertainty in performance assessments for complex systems. *Risk Analysis*, 14:483–511, 1994.

[8] Kurt Hornik. The R FAQ, 2013.

[9] R. A. Howard and J. E. Matheson. *Readings on the Prinicples and Applications of Decision Analysis (Volumes I and II)*. Strategic Decisions Group, Menlo Park, CA, 1984.

[10] R. J. Hyndman. Forecasting functions for time series and linear models, June 2013.

[11] P. Inechien. Interannual variability and global irradiance variability. Technical report, University of Geneva, May 2010.

[12] G. A. Meehl. The annual cycle and interannual variability in the tropical pacific and indian ocean regions. *Mon. Wea. Rev.*, 115:27–50, 1987.

[13] W. L. Oberkampf et al. Error and uncertainty in modeling and simulation. *Reliability Eng. Sys. Safety*, 75:333–357, 2002.

[14] United States Department of Energy, editor. *Wind Power in America's Future*. Courier Dover Publications, 2010.

[15] D. Savenije. The 10 greatest challenges the utility industry faces today, July 2013.

[16] G Stanhill and S. Moreshift. Global radiation climate changes: The world network. *Climate Change*, 21:57–75, 1992.

[17] D. P. Vasquez et al. Interannual variability of long time series of dni and ghi at psa, spain. Report of Solar Consulting Services, 2012.

[18] F. Vignola, C. Grover, N. Lemon, and A. McMahn. Building a bankable solar radiation data set. *Solar Energy*, 86:2218–2229, 2012.

[19] J.P. Watson and D. L. Woodruff. Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Sciences*, 8(4):355–370, 2011.

[20] S. Wilcox and W. Marion. Users manual for tmy3 data sets. Technical Report TP-581-43156, National Renewable Energy Laboratory, May 2008.

[21] D. Xiu. Fast numerical methods for stochastic computations: a review. *Comm. Comput. Phys.*, 5:242–272, 2009.

[22] A. Zobra and R. Bansal, editors. *Handbook of Renewable Energy*. World Scientific, 2011.

DISTRIBUTION:

1	MS 1327	Bill Hart, 1465
1	MS 1326	Jean-Paul Watson, 1465
1	MS 1140	Ceasar Augusto Silva Monroy, 6113
1	MS 1140	Ross Guttromson, 6113
1	MS 9159	Rich Chen, 8954
1	MS 9159	Genetha Gray, 8954
1	MS 9159	Patty Hough, 8954
1	MS 9159	Jerry McNiesh, 8954
1	MS 9159	Ali Pinar, 8954
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories