Managing Complex Network Operation with Predictive Analytics

**Zhenyu Huang, Pak Chung Wong, Patrick Mackey, Yousu Chen,
Jian Ma, Kevin Schneider, and Frank L. Greitzer**

Pacific Northwest National Laboratory
902 Battelle Blvd, Richland, Washington, USA 99352
{zhenyu.huang | pak.wong | patrick.mackey | yousu.chen | jian.ma | kevin.schneider | frank.greitzer}@pnl.gov

Abstract

Complex networks play an important role in modern societies. Their failures, such as power grid blackouts, would lead to significant disruption of people's lives, industry and commercial activities, and result in massive economic losses. Reliable operation of these complex networks is an extremely challenging task. None of the complex network operations are fully automated; human-in-the-loop operation is critical. Given the complexity involved, there may be thousands of possible topological configurations at any given time. During an emergency, it is not uncommon for human operators to consider thousands of possible configurations in near real-time to choose the best option and operate the network effectively. In today's practice, network operation is largely based on experience with very limited real-time decision support, resulting in inadequate management of complex predictions and the inability to anticipate, recognize, and respond to situations caused by human errors, natural disasters, or cyber attacks.

A systematic approach is needed to manage the complex operational paradigms and choose the best option in a near-real-time manner. This paper applies predictive analytics techniques to establish a decision support system for complex network operation management and help operators predict potential network failures and adapt the network in response to adverse situations. The resultant decision support system enables continuous monitoring of network performance and turns large amounts of data into actionable information. This paper presents examples with actual power grid data to demonstrate the capability of a proposed decision support system.

Introduction

Electric power grids, gas pipeline systems, telecommunication systems, and aviation networks are just a few examples of complex networks that are important in modern society. Their failure, such as power grid blackouts, would lead to significant disruption of peoples' lives, industry and commercial activities, and result in massive economic losses (DOE 2004). Operation of these complex networks is an extremely challenging task as they all have complex structures, wide geographical coverage,

and complex data/information technology systems. The complex networks also exhibit highly dynamic and nonlinear behaviors with numerous network configurations and are affected by a number of external factors. The external factors include physical attacks, cyber threats, human errors, and natural disasters. None of the complex network operations are fully automated; human-in-the-loop operation is critical. During an emergency, it is not uncommon for human operators to consider thousands of possible configurations in near real-time to choose the best option and operate the network effectively. In today's practice, network operation is largely based on operator's experience with very limited real-time decision support. Because of the complex nature of large networks (e.g. complex structure and wide geographical coverage), large amounts of data and information have to be processed to gain adequate situational awareness and the ability to adapt to emergency situations. Managing this complexity is emerging as a critical issue in complex network operation. Lack of complexity management often results in the inability to anticipate, recognize, and respond to situations caused by human errors, natural disasters, and cyber attacks and inadequacy in predicting the effect of operational decisions.

In this paper, we apply visual analytics techniques to enhance the processing of large amounts of operational data. Previously, visual analytics has been successfully applied to process massive amounts of data and extract useful information from this data (NVAC 2008). In this paper, we adapt the visual analytics techniques to convert massive amounts of operational data into actionable information. The resultant application enables prediction of network status, enhances the response to network operational requirements and provides real-time decision support to network operators. The application has been successfully demonstrated with actual power grid models and data. In the next section, a brief overview of the power grid operation is presented to address the needs for real-time decision support. The following section defines the problems in power grid operations and the process of applying visual analytics techniques. The application aims to convert data into information and present the information in an operator-friendly manner as a contoured geographic map. With the contoured map as the basis, the

next section develops a method to predict network security trends based on graph analysis. In both sections, actual power grid examples are presented to demonstrate the contoured mapping and graph trending analysis. The paper is then concluded with closing remarks and recommendations for future work.

Overview of Power Grid Operation

Recent power grid blackouts, such as the west coast blackouts of 1996 (Kosterev, Taylor, Mittelstandt 1999) and the east coast blackout of 2003(DOE 2004), brought significant attention to the reliability of power grids. How to predict and prevent or mitigate such blackouts has been a central topic in the area of power system research, and has also become one of the primary focuses of the DOE Office of Electricity Delivery & Energy Reliability (DOE-OE) (Congress 2005). Power grid operation involves complex computational processes with advanced power grid models. Figure 1 shows a functional structure of real-time power grid operation (Huang et al. 2007). The processes investigated in this paper are State Estimation and Contingency Analysis. The "State Estimator" typically receives telemetered data from the supervisory control and data acquisition (SCADA) system every few seconds and extrapolates a full set of grid conditions for operators based on the grid's current configuration and a theoretically based engineering power flow solution. The output of the State Estimator drives other operation functions including Contingency Analysis. Contingency Analysis studies "what-if" conditions in anticipation of potential power grid failures. Contingency Analysis identifies operation violations if one or more elements fail. The violation results are then presented to operators for review and decision-making to determine remedial actions if necessary. The burden of decision-making all falls on the shoulders of the operators.
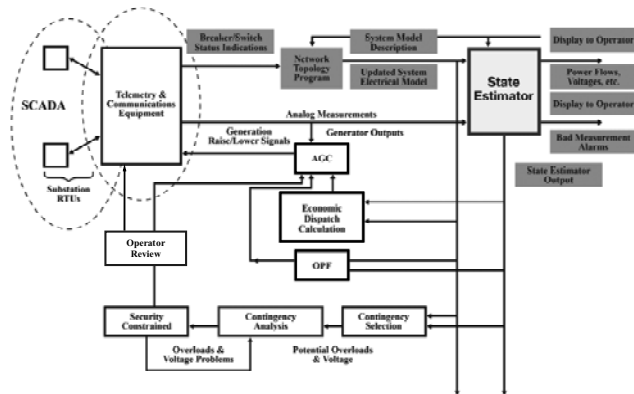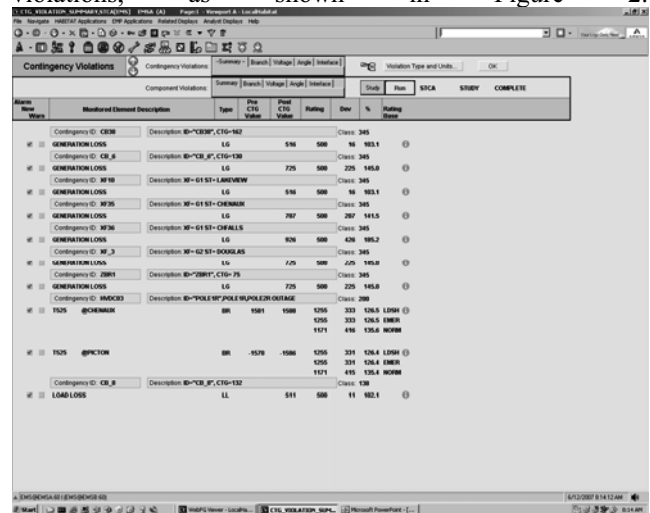

Figure 1. Functional structure of power system operations

The operator's decision-making process is the key step in ensuring power grid reliability. The operating standards of the North American Electric Reliability Corporation (NERC) require that the loss of any single element in the power grid should not cause system instabilities; this is referred to as the N-1 reliability criteria (NERC 2005). Contingency analysis is continually run in an interval of seconds to minutes to determine the impact of equipment failures. If the loss of one or more elements does not result in any limit violations, then the system is said to be secure for that contingency. The contingencies that result in violations of operating limits are flagged and placed in a list for the operators to inspect. NERC mandates that operators take actions to mitigate the situation in a timely manner when there are contingency violations.

Because it is not uncommon for several hundred contingencies to be examined, conveying this information to system operators in a meaningful and easy-to-understand way is a fundamental challenge. Because of the size of modern power grids, the number of contingencies to be studied can be very large. For example, the western North American high-voltage power grid has about 20,000 elements. Failure of any one element, i.e. N-1 contingencies, would constitute 20,000 contingency cases. "N-2" contingencies would be in the order of 10^8. Actual grid blackouts often involve the failure of multiple elements (N-x contingencies). State-of-the-art commercial tools use a tabular form to display each of the contingency violations, as shown in Figure 2.

Figure 2. Tabular representation of violation data in the state-of-the-art power grid operation tool

Each violation of operating limits is a row in this tabular form, without showing geographical information and the degree of severity. When there are only a few contingencies where the system is not N-1 secure, the method of tabular display is adequate. But when the system is heavily stressed, and there are significantly more contingencies violations, the tabular method of display is rapidly overloaded. It is then impossible for an operator to sift through the large amounts of violation data and understand the system situation within several seconds or minutes. However, it is in these situations that the

operators would most need the information when the tabular representation techniques are saturated.

Because of the above-mentioned challenges in processing large volumes of data from the contingency analysis process, it is certain that we will need a second layer of analytical tools to analyze the data and extract useful and necessary information for power grid operators. This layer of tools not only provides information about the current power grid status, but they also analyze historical data and generate system trending information to enable predictive capabilities. With this kind of real-time decision support, the operators will then have no need to review the massive amount of data but be presented with actionable information of the current status and system trends.

The next two sections present the application of visual analytics and graph trending techniques to construct such decision support tools for power grid operators. This decision support system will be able to fully utilize the contingency analysis results to predict potential problems of the power grid and adapt the power grid to adverse situations. Though this paper presents power grid examples, the decision support system can be extended for complex network operations in other industries. Examples include gas pipeline systems, telecommunication systems, and aviation networks.

Risk Assessment of Contingency Violations

Contingency violations are defined as operational parameters (i.e. power on a line or voltage at a substation) exceeding their limits. For example, the power that a transmission line can transfer has a limit due to thermal or stability constraints (NERC 1997). Exceeding the limits will result in equipment failure and/or system instability. Thus, the risk of a transmission line can be defined as the relative loading $R\%$ with respect to the limit P_{max}:

$$R_{ik}\% = \frac{P_{ik}}{P_{i\max}} \times 100\% \qquad (1)$$

where "ik" denotes the ith line of the kth contingency.

The risk of a substation can be defined similarly with the only difference being that the substation voltage has both lower and upper limits (V_{min} and V_{max}).

$$R_{ik}\% = \left| \frac{(V_{ik} - V_{i\min}) - (V_{i\max} - V_{i\min})/2}{(V_{i\max} - V_{i\min})/2} \right| \times 100\% \qquad (2)$$

where "ik" denotes the ith substation of the kth contingency.

For each contingency k, the risk of lines and substations can be categorized as:

$$R_{ik}\% \in \begin{cases} [0, R_T\%), & \text{safe} \\ [R_T\%, 100\%), & \text{alert} \\ [100\%, \infty), & \text{violation} \end{cases} \qquad (3)$$

where $R_T\%$ is the pre-specified alert risk level.

Compared with the tabular form shown in Figure 2, the improvement is that (1)-(3) convert the contingency data into quantitative risk levels, which indicate the severity if an operational parameter exceeds a limit. This conversion also goes beyond the violation data. Risk levels are defined as how close the operational parameters are to the limit, even if there are no violations, as shown in (3).

Defining the risk levels of individual elements (lines and substations) is the first step in converting contingency data into actionable information. Each contingency will generate a set of risk levels as defined in (1)-(3). If a total of K contingencies are analyzed, there will be K sets of risk levels. Across all of the contingencies, the risk level of the ith element can be defined statistically as the maximum, summation, or mean of the individual risk levels. Using the maximum as an example is shown below:

$$R_i\% = \max(R_{ik}\%), \ k = 1, 2, \ldots, K \qquad (4)$$

Two questions remain: How to present the risk levels in an easy-to-understand manner? And how to define risk levels for the whole network and for regions of interest? The next section presents an approach based on visual analytics techniques.

Visual Analytical Application

Failure of one element in a power grid could propagate into other areas of the grid. Given the different geographic locations of lines and substations and the heterogeneous structure of a power grid, the propagation would likely be different, and a same risk level of different lines or substations would have various levels of impact to the power grid. We assume that higher risk levels and risks in dense areas would have a larger impact on the reliability of the system. We further assume that the same risk level would propagate into the same radius of a geographic area, which is determined using visual analytics techniques. The result of this application is a contoured map with the color indicating the risk levels. Then it is very easy for operators to see the vulnerable areas of the grids without the need to sift through individual numbers.

Visual Analytics-Based Contoured Maps

The visualization starts with assigning the lines and substations the risk level as defined in (4) on the geographical map of the power grid. Then the propagation is visualized as fading colors from the center as shown in Figure 3. The impact area of a substation has a circular shape, while a line has an elliptical shape. Individual risk

areas are then superposed to form the collective risk areas. The same superposition is done among multiple contingencies as well.
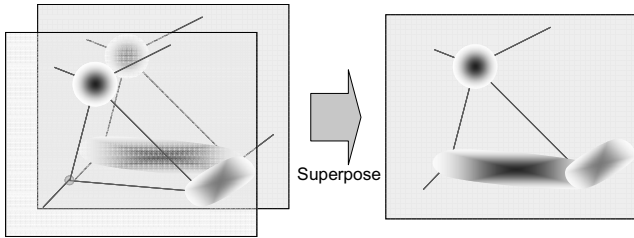

Figure 3. Collective risk area using fading colors and superposition of individual risk areas

The implementation uses a hash table to store all the pixels of the lines and substations. Each pixel has a value determined by the risk level of the line or substation. When lines are crossing, the larger value remains in the table so the highest risk is represented (Figure 4).

Figure 4. Hash table for storing pixels

The next step is to create the color filter for displaying the resultant collective risk. The filter is circular shaped with values conforming to that of a Gaussian curve (Figure 5). The Gaussian curve is normalized so that the peak height is equal to one. The radius of the filter is a parameter that is set by the user. We define the Gaussian curve to have three standard deviations within one radius. Next we iterate through all the pixel points associated with the lines and substations stored in the hash table.
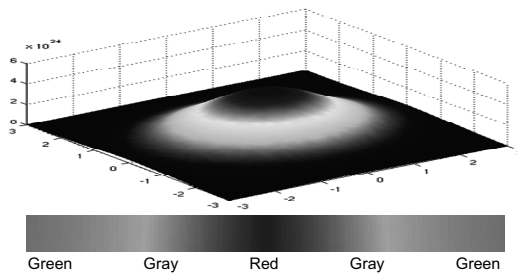

Figure 5. Gaussian color filter and green/gray/red color map.

At each one of these points, the value in the table is multiplied by the Gaussian curve. These values are then added to an output graphic matrix representing the final contour. The outcome of the Gaussian filtering is the output matrix defining each point in the map with a floating point number. Then these floating point numbers are assigned to a color map to obtain the final contour. In

order for it to be easy to interpret, a green/gray/red color map is selected. Considering (3), the color map can be understood as green, gray and red correspond to three risk categories – safe, alert and violation.

The final visual representation uses HaveGreen (Wong et al. 2006) as the application framework, which provides the interface for navigating and zooming over the power grid. The graphics is developed in C# using Managed DirectX. An example of the color contoured map is shown in Figure 6. This example uses actual model and data of the western North American power grid. 200 contingencies are analyzed, and 200 sets of risk levels are overlaid on the single map to visualize the collective risk of the contingencies on the system security. The red color (shown as darkest areas in the figure) indicates vulnerable portions of the power grid and brings attention to network operators. Compared with Figure 2, this color contoured map has the obvious advantage of bringing information rather than raw data to operators.
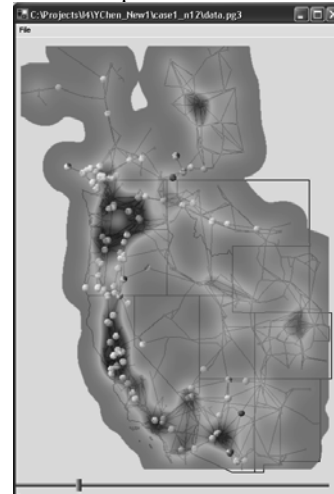

Figure 6. Western North American power grid risk map with 200 overlaid contingency analysis results

System and Regional Risk Levels

The color contoured map visually shows the risk across the network. This section employs statistical analysis methods to quantitatively calculate the risk level $R\%$ of the network and individual regions. The risk level is defined as a combination of arithmetic average and geometric average.

$$R\% = a_1\eta + a_2\gamma \qquad (5)$$

where a_1 and a_2 are weighting constants. η and γ are the arithmetic average and geometric average, respectively. The statistics is performed over all the pixel points on the map. Each pixel has a color value corresponding to the risk level at that pixel. If we categorized the pixels into M categories and there are N_m pixels in each category with the same color value $(R\%)_m$, the arithmetic and geometric averages are calculated as follows:

$$\eta = \frac{\sum_m (R\%)_m N_m}{\sum_m N_m}, \quad m = 1, 2, \ldots, M \tag{6}$$

$$\gamma = \left[\prod_m (R\%)_m^{N_m} \right]^{\frac{1}{\sum_m N_m}}, \quad m = 1, 2, \ldots, M \tag{7}$$

For regional risk levels, the same process can be applied but only the pixels in the region are considered.

Figure 7 shows the risk levels of the western North American power grid over a morning load pick-up period. When the system total power consumption is at a low level (the beginning of the period), increasing load does not increase risk levels as much as when the total load is at a higher level toward the end of the period. This is consistent with operational experience.
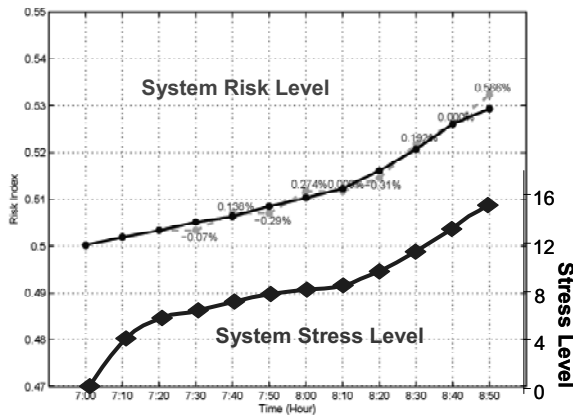
Figure 7. System risk level and stress level over time

It is worth pointing out that the same statistical analysis can be performed on the risk levels calculated from (4). The advantage of performing the analysis in the visual space is that the propagation and collective risk areas are considered, which is more reasonable and realistic for actual power grids.

Visual Trending Analysis

Converting the data into risk levels and visualizing them as contoured maps enables visual network management, which makes it easier to gain situational awareness and recognize problems. Operators would now have more time to focus on urgent issues rather than spending time analyzing unimportant data. Operators could also observe the evolving patterns of the visual maps to determine network reliability and security trends. For example, an increase in color intensity and size of the risk contour would indicate a deteriorating network situation that would raise awareness for the operator. In a simple network, evolving patterns are simple to understand and visual examination of the maps would be adequate to determine trends. However, in a complex network, evolving patterns can be complicated and the number of the patterns can be

numerous at any given time. Figure 8 shows a simpler case with a few obvious red areas and several fuzzy gray areas. All areas evolve in time. An operator may be able to recognize the pattern of areas 2 and 3 merging into one single area. But it would be very difficult to determine how the other areas are evolving and how to quantify the implications. And more importantly is to use the results of the contingency analysis to determine the trend and predict the network status in the future.
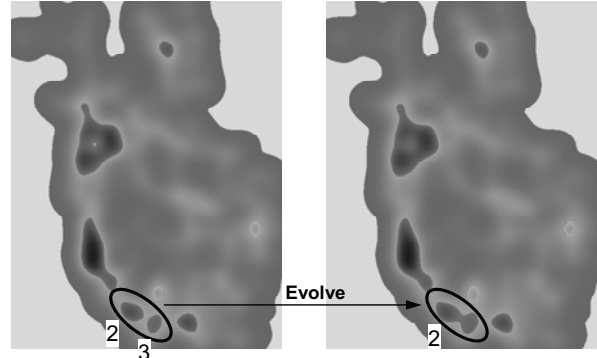

Figure 8. Power grid risk evolving patterns

To enable this predictive capability, we developed a method for visual trending analysis. The method is based on the system and regional risk levels as defined in (5). The trend is obtained by fitting a curve to historical risk levels of the network or regions, and extrapolating to predict the future system situation, as shown in Figure 9.
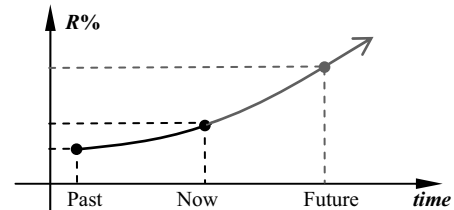

Figure 9. Illustration of visual trending analysis

Complex evolving patterns may exist in a network. Some of complex patterns are shown in Figure 10. Two areas can "merge" into one, or one can "split" into two, or one area "steals" a portion of another area.
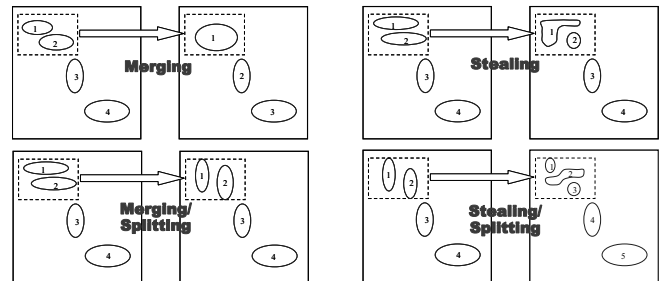

Figure 10. Complex evolving patterns of network risk impact areas

To automatically identify all the complex patterns, the actual implementation of visual trending analysis combines structural analysis and statistical analysis, as shown in

Figure 11. Statistical analysis is used to calculate the risk indices of individual areas, while the structural analysis uses a relation matrix to capture the relationship between areas, i.e. how two areas overlap or differ at the pixel level. The number in the relation matrix is the sum of the risk level for each pixel in the overlapped area, except the last row and last column are for the adjacent area. The numbers in Figure 11 correspond to the case shown in Figure 8. This trending analysis approach has been able to identify all the complex evolving patterns shown in Figure 10.
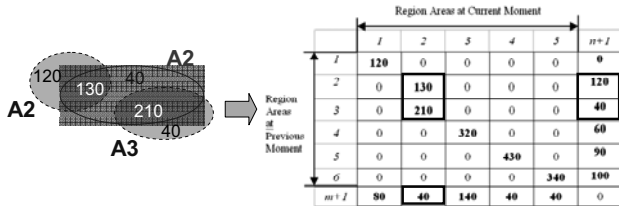

Figure 11. Combination of structural analysis and statistical analysis using a relation matrix for visual trending analysis.

The green dashed line in Figure 7 is the predicted system risk level, each point based on the three prior risk levels. It can be seen that the prediction is reasonably close to the actual system risk level (blue line). Figure 12 further shows the trends for the five most critical regions in the power grid for this example, corresponding to the same system conditions in Figure 7. The regional risk trends are more extreme than the system trend. The system trend is relatively flat as changes in different regions may cancel each other's impact. Therefore it is important to observe regional trends to recognize potential regional issues.
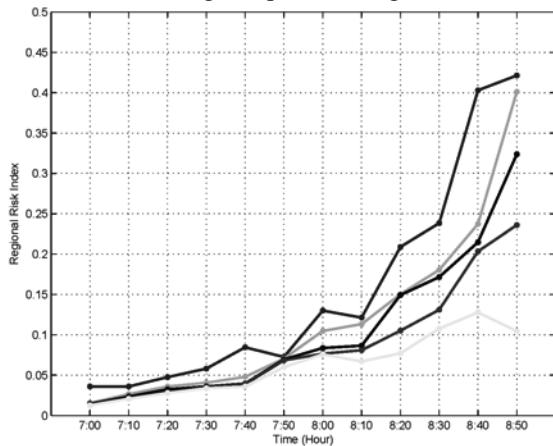

Figure 12. Example of regional risk trends of the western North American power grid

Conclusion and Future Work

This paper described visual analytics techniques that were successfully applied to complex network operations by converting large amounts of operational data into actionable information. Operational data are translated into risk levels and then visualized as a color contoured map. This application will greatly improve real-time decision support for network operations. Operators can quickly gain situational awareness of the network without sifting through large amounts of raw data. Predictive capability is established by analyzing the trend of the network risk level with an approach that combines structural analysis with statistical analysis. Examples using actual models and data of the western North American power grid demonstrate the validity of the predictive analytics.

Given that these results were obtained in a research environment and based on simulation, an important step to bring these proposed methods into practice will be demonstrating this approach in an actual power system environment and evaluating the decision support tool with experienced operators. Further work should also focus on developing methods to improve the prediction by including probability to enable better forecasting for network operations. This forecasting capability refers to multiple future paths, each of which has a certain probability. Another enhancement is helping operators decide the outcome of various remedial actions through an interactive analysis function. This interactive analysis would identify remedial actions that can turn the "red" to "green" on the contoured map. Some of this work is ongoing and results are expected to be published in the near future.

Acknowledgement

References

Congress U.S. 2005. Energy Policy Act of 2005. URL: http://www.doi.gov/iepa/EnergyPolicyActof2005.pdf.

DOE. 2004. U.S.-Canada Power System Outage Task Force: Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations.

Huang, Z., Guttromson, R., Nieplocha, J., and Pratt, R. 2007. Transforming Power Grid Operations via High-Performance Computing. *Scientific Computing* April.

Kosterev, D.N., Taylor, C.W. and Mittelstadt, W.A. 1999. Model Validation for the August 10, 1996 WSCC System Outage. *IEEE Trans. Power Syst.* 14(3): 967-979.

The National Visualization and Analytics Center (NVAC). 2008. Accessed 12/2008. URL: http://nvac.pnl.gov/.

North American Electric Reliability Corporation (NERC). 1997. NERC Planning Standards.
URL: www.nerc.com/pub/sys/all_updl/pc/pss/ps9709.pdf.

North American Electric Reliability Corporation (NERC). 2005. Transmission System Standards – Normal and Emergency Conditions. Accessed 12/2008. URL: www.nerc.com.

Wong, P.C., Chin, Jr., G., Foote, H.P., Mackey, P.S., and Thomas, J.J. 2006. Have Green – A Visual Analytics Framework for Large Semantic Graphs . In IEEE Symposium on Visual Analytics Science and Technology. Baltimore, MD.