Inferring Network Structure: the New York Power Grid

James R. Thompson[1], **Charlotte James**[2], **Ryan McGee**[3], **Harrison B. Smith**[4], **Aina O. Vila**[5], **and Mika Straka**[6]

[1]MITRE Corporation, Washington D.C., USA
[2]University of Bristol, Bristol, UK
[3]University of Washington, Seattle, WA , USA
[4]Arizona State University, Phoenix, AZ , USA
[5]Pompeu Fabra University, Barcelona, Spain
[6]IMT School of Advanced Studies, Lucca, Italy

ABSTRACT

Opening wholesale power markets to speculative trading is a deliberate attempt to incorporate the benefits of competition and price discovery to a historical monopoly. Whether those benefits are being realized is an ongoing discussion, but it begs the question: do the new policies unintentionally make the grid more vulnerable to cascading failure? More importantly can we measure that vulnerability and take steps to minimize it, making our critical infrastructure more resilient and secure? Much of the power market research to date focuses on the optimization algorithms required to dispatch power and solve for location based pricing. Agent-based models have been constructed to investigate the economic impact of market deregulation, and the financial industry has employed data analytics and machine learning algorithms to measure financial risk associated with congestion contracts and other financial instruments (e.g.[1–5]). The perspective of most of these studies are the economic benefits and improvement of normal day-to-day business practices. Few studies, if any, have taken a grid security perspective at the Independent Service Operator level, which necessarily changes the study objectives and focus. This research takes a bottom-up approach to market operations in an effort to understand the risks of cascading failures at the national level. We analyze the processes employed by market participants to deliver power, and investigate if the underlying grid network structure and vulnerabilities can be inferred from the analysis.

Introduction

In a securities exchange, the double-auction order book organizes the bids and offers into an agreed upon queuing arrangement. Generally the highest bid and lowest offer constitute the *top of the book* and the difference between them is the so called *spread*. Any incoming order that crosses the spread – i.e. a bid higher than the lowest offer or an offer lower than the highest bid – results in a trade and thereby sets the current price. Economists argue that such a process will naturally uncover the *market equilibrium price* where the value of excess demand is $0.00. In this decentralized market, the buyers and sellers are trying to maximize their own utility functions independent of other market participants. Economic theory contends that the market's "invisible hand" emerges from this process as if an implicit Walrasian clearing agent were governing price[6].

Unfortunately the physics of electrical power makes it considerably more difficult to trade in this decentralized manner. First, power cannot be stored efficiently. Generators that convert different fuels into power are thus forced to operate on a just-in-time basis. Second, the transmission of power from the point of generation to the point of demand requires an expansive network of transmission lines. The transmission grid constitutes a natural monopoly over the transport of power since it would be impractical to build competing grids. Last, the fundamental physics of electromagnetism as approximated by Kirchhoff's circuit laws state that electric charge and energy must be conserved around any circuit. Thus, excess inventory cannot simply wait on a suitable buyer, but must be consumed immediately to maintain the conservation laws throughout the grid. The net result is that power markets are designed, engineered, and operated by centralized organizations known as Independent Service Operators (ISO). ISOs function as explicit Walrasian clearing agents[6] that solve directly for the market equilibrium price at any given point in time while simultaneously ensuring that the physical engineering constraints of the grid are met[7].

Power markets consist of generators, transformers, wholesale distributors, the transmission grid, the distribution grid, and the ISO. Generators supply hundreds of megawatts per hour to the wholesale distributors via the high-voltage transmission grid. Transformers step the voltage up for efficient transfer in the transmission grid and then back down for distribution in the retail or distribution grid. Wholesale distributors purchase power from the generators and subsequently resell it to retail customers. Some retail customers are fixed-price customers that pay a fixed amount per kilowatt consumed. Other retail customers are capable of generating their own power (e.g. college campuses) and thus have a variable price with the wholesaler. Finally, the ISO manages the auction of power between the generators and the wholesale distributors while simultaneously ensuring the constraints of the high-voltage transmission grid are met.

The auction process varies from one ISO to another, but generally follows the process established by Hogan[8] and summarized by Biggar and Hesamzadeh[7]. On a minute-by-minute basis the ISO must;

- dispatch the lowest cost mix of generators that meet demand subject to system constraints
- manage demand through load shedding and other incentives
- utilize the transmission grid efficiently in terms of balancing, maintenance, etc.
- and manage billing and collection.

Additionally, the ISO provides risk management instruments that allow market participants to hedge against congestion on the transmission grid and insulate themselves from price volatility. This is done by encouraging speculative traders to enter the market and trade the financial contracts rather than the physical asset. All of this is accomplished by solving an elaborate mixed-integer linear program (MILP) that considers bids, offers, and a direct current (DC) approximation to the alternating current (AC) power flow equations, which is solved iteratively on a minute-by-minute basis to account for the stochastic nature of demand in real time. The MILP determines which generators will be dispatched to produce a given amount of power and at what price. The prices in turn drive the trade and outcomes of the risk management instruments. A simplified version of the MILP is given below:

$$\text{minimize: } \sum_i p_i^G C_i^G \qquad \text{//Minimize cost according to the generator bids } C_i^G.$$

subject to:

$$\begin{aligned}
&\sum_i p_i^G - p_i^L = 0 && \text{//Power conservation; generator offers equal consumer bids}\\
&\sum_i H_{l,i}(p_i^G - p_i^L) \leq K_l && \text{//Transmission constraints using shift factors } H_l\\
&p_i^G \leq p_i^{\max} && \text{//Maximum generating capacity of each generator } i\\
&-p_i^G \leq -p_i^{\min} && \text{//Minimum generating capacity of each generator } i\\
&p_i^L = B_i && \text{//Meet the wholesaler bids for power} B_i
\end{aligned}$$

where p_i^G is the power dispatched by generator i, p_i^L is the power consumed by a load at i, C_i^G is the price per megawatt offered by generator i, K_l is maximum number of megawatts allowed on transmission line l, and B_i is the quantity of megawatts wholesaler i would like to purchase. The integer variables (not shown above) are binary decision variables that reflect whether a generator is dispatched or not. The shift factors H_l are properties of the transmission lines and network topology that determine how power is redistributed as demand changes. The overall ISO process is shown in Figure 1.

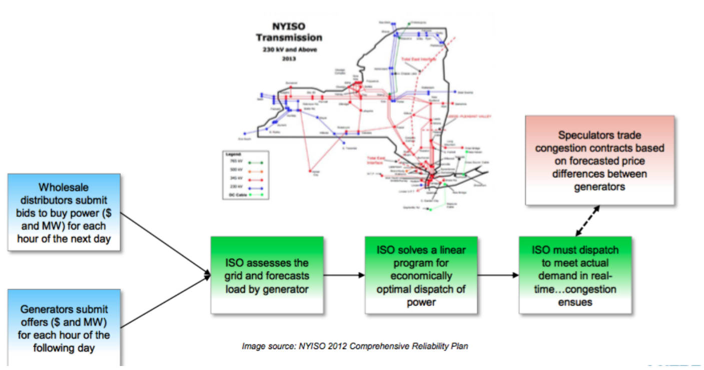


Figure 1. ISO Process and New York transmission grid

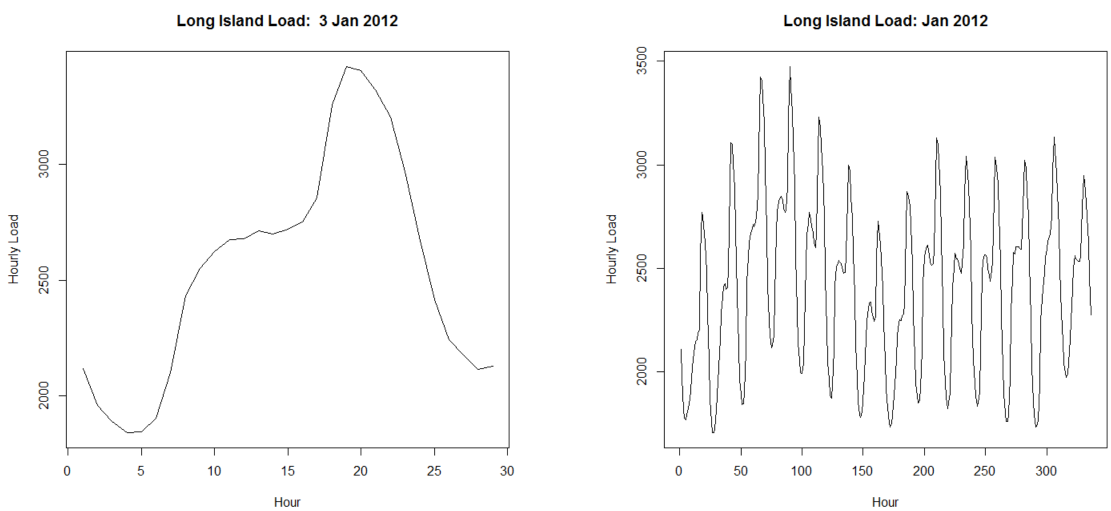
There is anecdotal evidence that misplaced incentives can lead to profiteering that runs counter to social welfare[9]. The collapse of the energy trading firm, Enron, is one extreme example. Traders within the firm used their influence to limit the flow of power to the state of California, resulting in numerous blackouts and a 500% increase in the price per megawatt in

2000 and 2001[10]. The Enron example relates to deliberate, nefarious behavior, but subtler problems may exist. The MILPs solved by ISOs employ objective functions that minimize cost or in some cases maximize dollar-surplus across all parties. Indeed it can be shown that in the absence of transmission constraints and power production limits, the optimal solution is to dispatch power in rank order from cheapest to most expensive until demand is met. This narrow view of dispatching power ignores longer term effects on critical infrastructure and engineered equipment that is being over utilized or utilized in a manner beyond its original design specifications. In this paper we take a preliminary step toward measuring the impact that market structure has on the underlying critical infrastructure. We first analyze the data produced by the New York ISO to understand the relationships between market participants and the transmission grid. The goal is to build toward a model that accurately represents the physical structure and incentive structure so that policy changes can be made or improved that lower the risk of cascading failures.

Methods

We chose to focus our analysis on the New York ISO (NYISO) for practical reasons. First, they make much of their data public via their website. Second, of the nine interconnects in the United States, NYISO is the smallest. However, it also contains New York City and therefore one concentrated area of the state demands a substantial amount of power. This gives rise to a substantial amount of congestion contracts being traded. NYISO manages 512 generators, spread across 15 zones, and connected by over 11,000 miles of transmission lines.

Our first step was to characterize the historical load data and price for that load across the 15 zones. Power load follows a relatively predictable pattern that is driven by daily human activity. There are discernible peaks and valleys that correspond to the productive and normal sleeping hours of the day respectively. Weekends and holidays generally have lower peaks than regular work days. This is not to say that load is deterministic, but by employing Fourier analysis paired with regression one



(a) Load profile - 1 day

(b) Load profile - 2 weeks

Figure 2. Periodicity of Power Demanded

can forecast the daily load with an acceptable degree of accuracy. Beyond human activity, weather – specifically temperature – is the other main driver of demand for power. Temperature extremes lead to spikes in power usage beyond what would be predicted by frequency analysis alone. Most ISOs and market participants employ sophisticated weather models when trying to predict the demand for power in the near term.

The price of power is considerably more complex. In the simplest context, one might assume that demand for power drives price and thus price would follow a similar periodic structure to load. If this were true, the same data analysis methods used for load would be immediately applicable to price. However, the interconnection of power generators via the transmission grid and the complicated process of dispatching power governed by the ISOs leads to strong interdependence in the price of power. This dependence structure is non-linear in nature and difficult to quantify as a result. As such, the price time series for load exhibits the so-called *stylized facts* that are observed in most securities' price time series. These stylized facts are generally measured in the *returns* time series, which is defined as the relative change in price from one period to the next. There is ongoing debate about how to quantify the stylized facts, but it is generally accepted that they at least include

- long memory in the absolute returns as characterized by the auto-correlation function
- volatility clustering as characterized by non-stationary behavior in the variance of the returns
- heavy-tails in the returns as characterized by a leptokurtic distribution governing returns[11].

The problem with the above measures is the non-stationary aspect of financial data. That is, the mean and variance of the time series appears to be changing constantly. Additionally, heavy-tails and long memory are imprecise terms that gloss over the fact that they refer to infinite measures that cannot be readily compared from one time series to the next. Fortunately, the presence of stylized facts also implies the presence of power laws, which can be measured. It turns out that the price of power exhibits multifractal behavior, that is fluctuations that scale in time as a spectrum of power laws rather than just one[12]. As such, we employed a method put forth by Kandelhart et. al. known as multifractal detrended fluctuation analysis (MF-DFA) that extracts the multifractal spectrum from empirical data[13]. Our implementation of MF-DFA was taken from software written by Thompson and Wilson[14].

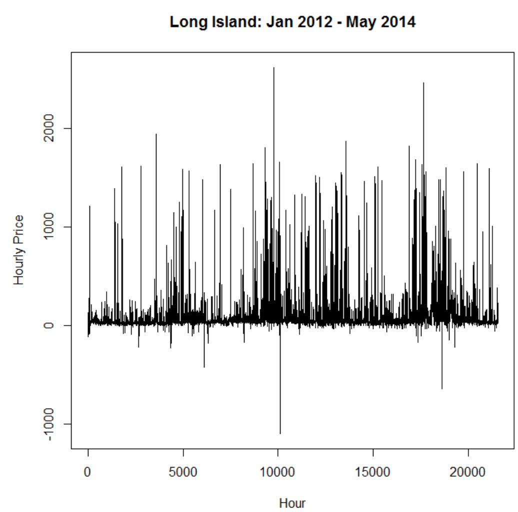


Figure 3. Price per megawatt of power in Long Island 2012-2014

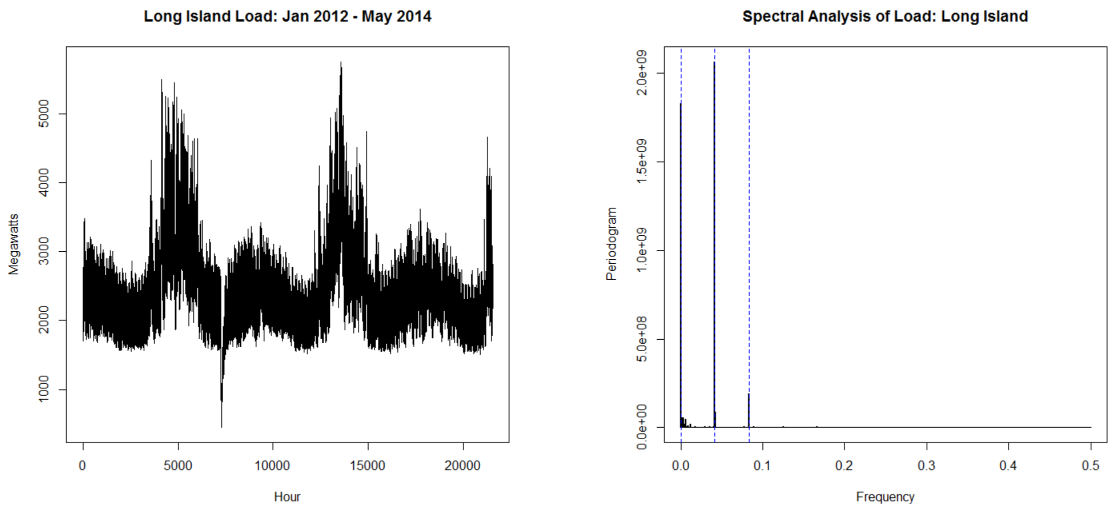
In order to characterize relationships between market participants and the underlying infrastructure, we required methods for measuring dependence between empirical datasets. The usual standard is linear correlation, but as previously mentioned power markets are exceedingly complex and thus non-linear. We therefore chose to employ techniques from information theory, specifically mutual information. Similar to correlation, mutual information measures statistical correspondence between two variables. However, mutual information does not assume linearity, continuity, or other specific properties and can therefore detect interactions that may be missed by linear correlation measures. After computing the mutual information between two variables, a method known as "Context Likelihood of Relatedness" (CLR) can calculate the statistical likelihood of each mutual information value within its network context[15]. The most probable interactions are those whose mutual information scores stand significantly above the background distribution of such scores. This CLR step removes many of the false correlations in the network by eliminating cases such as when one generator weakly covaries with a large numbers of others.

Finally, our original objective was to infer the structure of the physical grid and the associated incentives and policies that have built up around it. Our intent is to use the aforementioned techniques to quantify the relationships between market participants and the physical infrastructure and then use those relationships to construct meaningful networks that can be analyzed using the techniques of network theory[16].

Results

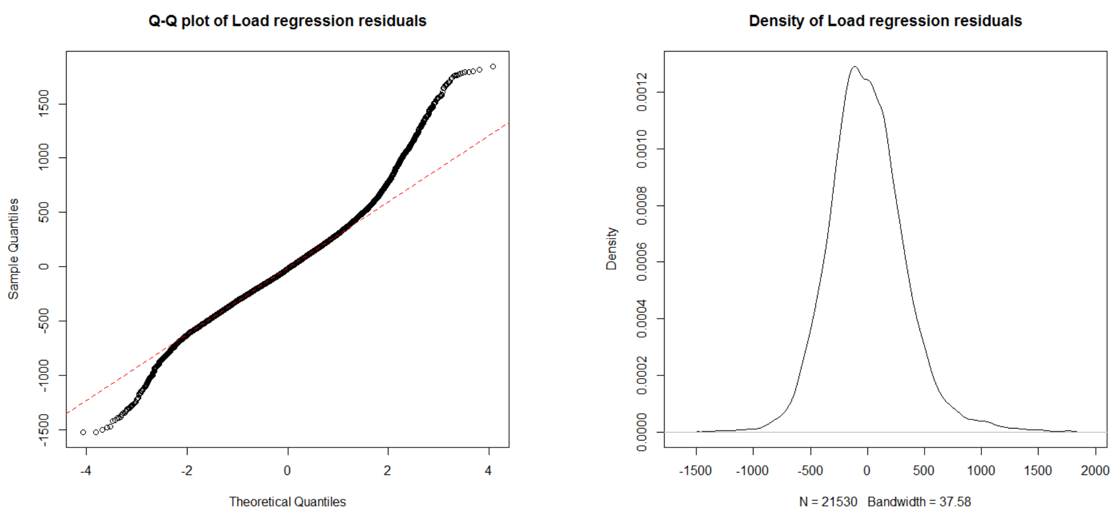
Time Series Analysis

Our first step was to analyze the time series of power load and price for that load to look for relationships between demand, price, and the process required to deliver power. As previously mentioned, the load time series follows a relatively predictable pattern based on the periodicities of human activity and weather. Using Fourier transforms, we extracted the 10 most resonate frequencies in the load data and performed a regression against the resulting trigonometric polynomial[17]. The regression analysis provided a suitable enough fit (R^2 value of 0.73) for us to conclude that power demand was relatively predictable. However, our goal was not to forecast the demand for power, but rather to characterize it statistically in order to explore its relationship with price and the ISO process. Although the residuals from the regression analysis fall reasonably well on a Gaussian quantile plot, they do not pass the Box-Ljung portmanteau test for autocorrelation. A more sophisticated model would be required to account for the correlation structure in a forecasting endeavor.



(a) Load profile - 2 years **(b)** Spectral analysis

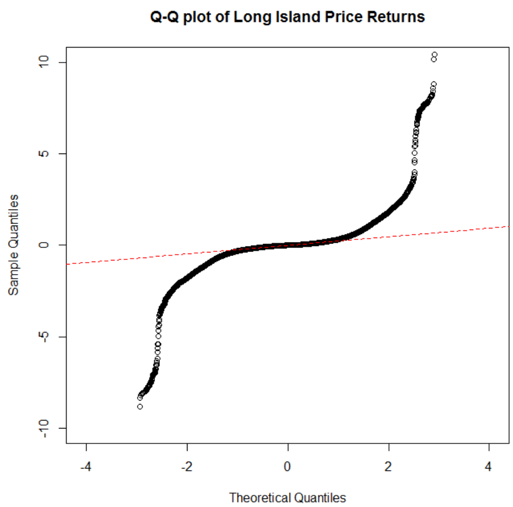
Figure 4. Periodicity of Power Demanded



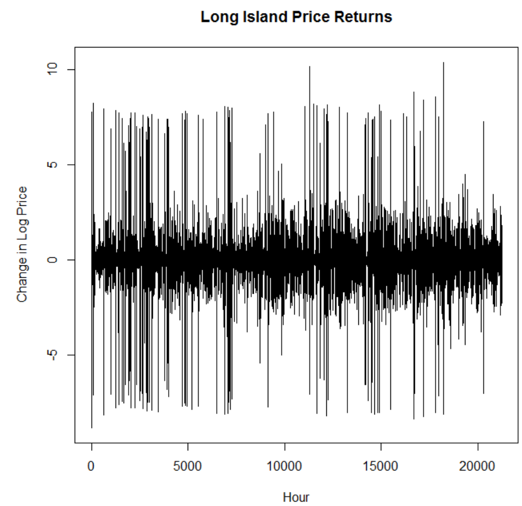
(a) Q-Q plot of regression residuals **(b)** Empirical density of residuals

Figure 5. Analysis of regression residuals

As outlined earlier, the price paid for power is determined by bids for load and offers from generators to supply megawatts over the transmission grid shared by all market participants. The process of dispatching the lowest cost mix of generators thus determines the location-based marginal price for each generator connected to the grid. The result is a complex time series that bears little resemblance to the demand driving its production. Although there is a somewhat predictable frequency of price spikes that coincides with the trends in load, the regression procedure mentioned above fails all the usual tests for statistical significance. Similarly, the log-returns – defined as the difference in the natural logarithm of the price – exhibits the stylized facts generally associated with securities prices. The plots in Figure 6 illustrate the leptokurtic nature of the returns and the volatility clustering that results from a non-stationary variance. The volatility clusters and heavy-tails suggest that the



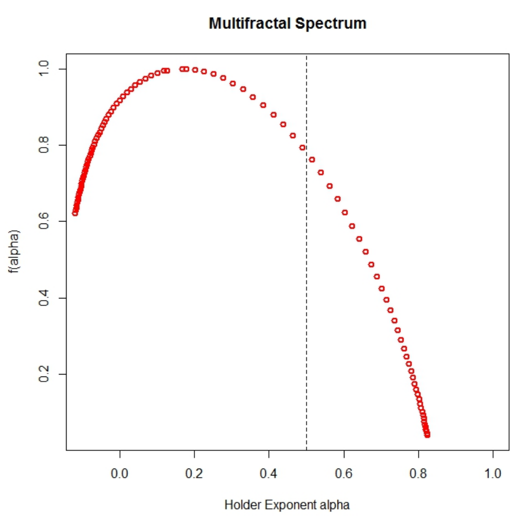
(a) Q-Q plot of regression residuals



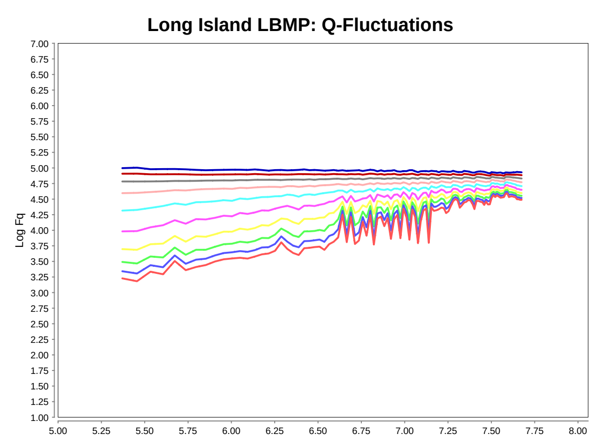
(b) Time series plot of price returns

Figure 6. Standard analysis of price time series

time series is likely to be multifractal, i.e. the fluctuations in the time series are power law distributed across different time resolution scales. We confirmed that the price for power is indeed multifractal using MF-DFA and analyzing the fit on the q^{th}-order fluctuations in log-log scale. In multifractal analysis, the exponent q acts as a filter for different size fluctuations.



(a) Multifractal spectrum of price returns



(b) q^{th}-order fluctuations versus log scale s

Figure 7. Multifractal analysis of price returns

When q is negative and large (say around -5) the large fluctuations are filtered out of the data set and the small fluctuations are accentuated. When q is positive and large (say around $+5$) the small fluctuations are filtered out of the data set and the large fluctuations are accentuated. By systematically testing for power law behavior at different values q, we can test whether the Hurst exponent is changing for different size fluctuations or remaining constant. A constant Hurst exponent indicates a monofractal process such as Brownian motion, while a changing exponent indicates a multifractal process. Last, if the multifractal spectrum is not concave, the data in question is not fractal. The advantage of multifractal analysis is that we can compare the scaling properties of different time series. Rather than simply claiming a time series exhibits stylized facts, we can compare and contrast the scaling behavior of multiple time series in a quantifiable way. Specifically in a monofractal the Hurst exponent quantifies the statistical behavior of successive increments in the process. A Hurst exponent of 0.5 indicates that successive increments are perfectly uncorrelated. Values less than 0.5 indicate that successive increments are antipersistent, meaning they are more likely to be in opposite directions. Greater than 0.5 the increments are persistent or more likely to be in the same direction[18]. In a multifractal, different size fluctuations behave differently. Thus, large fluctuations may be antipersistent while small fluctuations may be persistent or uncorrelated. In the case of the price of power, the multifractal spectrum peaks below 0.5 indicating a predominately antipersistent behavior. However, for small fluctuations (larger values q) the tail of the spectrum extends above 0.5 indicating some persistence.

The key result of our time series analysis is that the process of pricing and distributing power generates a complex signal in the form of the price time series. Simply quantifying the demand and available supply of megawatts is not enough to effectively characterize the market price for those megawatts.

Mutual Information and CLR

Given that the ISOs process creates complex dynamics in the resulting price time series of any one region, we explored the relationship between prices at the same time across multiple regions. Theory predicts that if a transmission grid contains N lines (or edges) and $m < N$ of them are at their maximum allowable limit (i.e. congested), then there are only $m+1$ independent prices in the network. The remaining prices are linear combinations of the $m+1$ independent prices. Thus, a natural first step was to analyze the pair-wise correlation between the 15 zone prices in the NYISO. Unfortunately, we found that linear correlation was insufficient for extracting meaningful relationships. In point of fact, all the 15 zones exhibited pair-wise correlation with each other above 0.75 and most were approaching 1.0 (see Figure 8). Thus, we employed mutual information

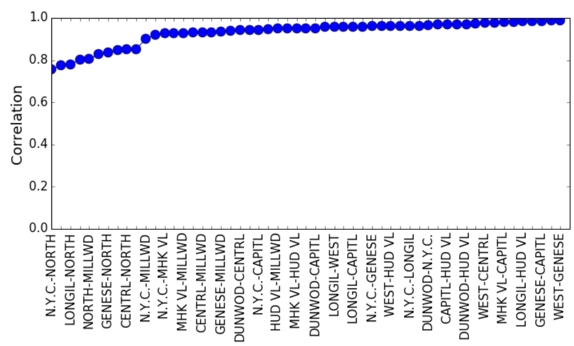
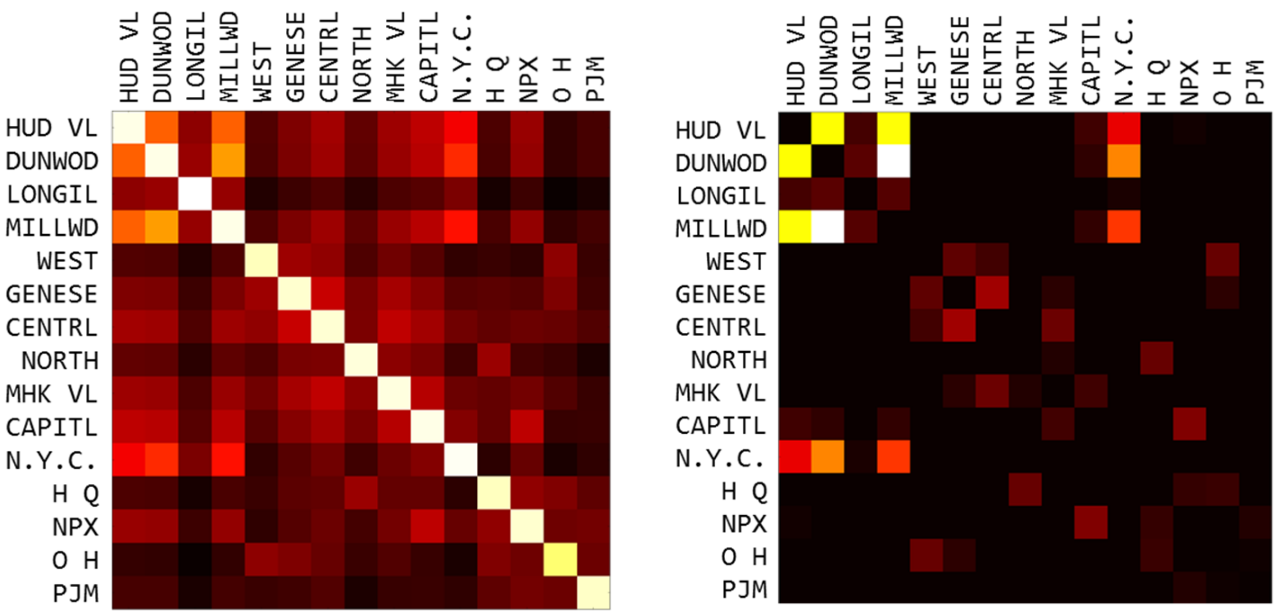


Figure 8. Linear Correlation

and CLR calculations to infer interactions between zones based on their price time series data. The pair-wise mutual information (MI) between respective price time series was calculated for all pairs of zones. Then the CLR algorithm was run to transform these pairwise MI values to pair-wise CLR scores, which ranks the relative likelihood of interaction for each pair of zones.



(a) Mutual information heat map

(b) Mutual information heat map with CLR filter applied

Figure 9. Mutual Information Analysis

Network Analysis

The final step was to convert the analysis of previous sections into a network representation that might allow us to infer vulnerabilities or stress points in the actual transmission grid. Our initial step was to use the pair-wise CLR scores to assign edges and weights between zones. The results show that this method infers a network topology that is largely consistent with geospatial relationships of zones and known high-level transmission patterns. The mutual information between pairs of generators indicates similar results to the piecewise linear correlation; zones that are nearest neighbors share higher information than those that are separated by larger distances. This suggests that this method is reasonable to use with this type of data, and can be explored through application at a more granular generator scale.

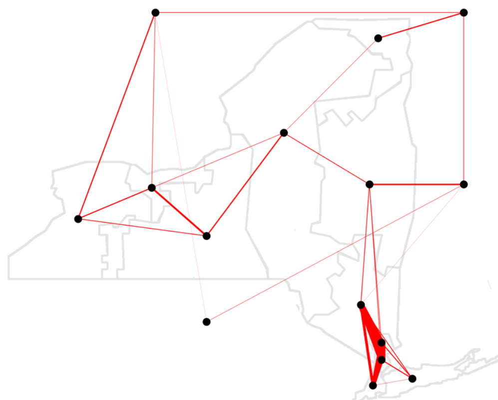


Figure 10. Zone network using mutual information

However, we also theorized that less expensive power will exert a great deal of influence on the overall flow of power throughout the grid. Again focusing on the zone scale, the difference between the prices of two locations at any given time provides information about the most desirable location from which to purchase power. This is an oversimplification of the process; Kirchoff's Laws and congestion in the grid prevent a given zone from purchasing all of its power from the most reasonably priced location. For each of the 15 zones we looked at the zones corresponding to the largest and smallest price differentials, resulting in a directed network of 15 nodes and 30 edges. In general, the most reasonable (low priced) zone is Hudson Valley; all other zones want to buy from Hudson, while the most undesirable zone to purchase from is ISO New England (ISO-NE), a generator that is out of state and therefore has additional zones it needs to supply within its separate coverage area.

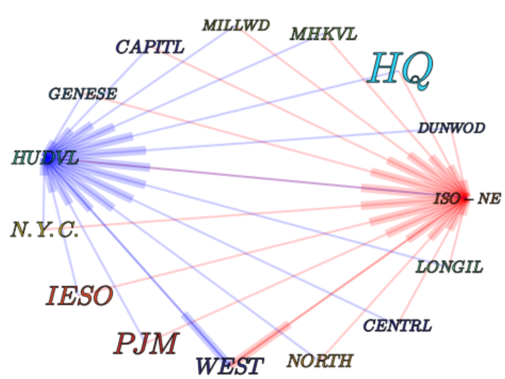


Figure 11. Pair-wise Price Gradients

Discussion

This work constitutes a preliminary step towards inferring network structure – both physical and financial – in the U.S. power markets. Our ultimate objective is to analyze and identify potential vulnerabilities in this critical infrastructure that can be mitigated by changes in policy or the ISO process. Avenues for future research include a deeper analysis of the properties of the mixed-integer linear program that would allow for greater inference on the properties of the underlying transmission grid. All electrical power grids can be characterized by the line connections and the electrical resistance of those lines. These physical properties must be incorporated into the MILP that in-turn solves for location-based prices. As such, the prices function as signals that contain information about the underlying grid structure. Additionally, agent-based models can be employed to test the deductions made through empirical analysis by simulating a grid with known properties along with agents representing the wholesalers, generators, and speculative traders.

References

1. Zhou, Z., Zhao, F. & Wang, J. Agent-based electricity market simulation with demand response from commercial buildings. *IEEE Transactions on Smart Grid* **2**, 580–588 (2011).
2. Weidlich, A. & Veit, D. A critical survey of agent-based wholesale electricity market models. *Energy Economics* **30**, 1728–1759 (2008).
3. Sun, J. & Tesfatsion, L. Dynamic testing of wholesale power market designs: An open-source agent-based framework. *Computational Economics* **30**, 291–327 (2007).
4. Garcia, R. C., Contreras, J., Van Akkeren, M. & Garcia, J. B. C. A garch forecasting model to predict day-ahead electricity prices. *IEEE transactions on power systems* **20**, 867–874 (2005).
5. Deng, S.-J. & Oren, S. S. Electricity derivatives and risk management. *Energy* **31**, 940–953 (2006).
6. Kelso Jr, A. S. & Crawford, V. P. Job matching, coalition formation, and gross substitutes. *Econometrica: Journal of the Econometric Society* 1483–1504 (1982).
7. Biggar, D. R. & Hesamzadeh, M. R. *The economics of electricity markets* (John Wiley & Sons, 2014).
8. Hogan, W. W. A market power model with strategic interaction in electricity networks. *The Energy Journal* 107–141 (1997).
9. Coffee Jr, J. C. Understanding enron:" it's about the gatekeepers, stupid". *The Business Lawyer* 1403–1420 (2002).
10. Joskow, P. L. California's electricity crisis. *Oxford Review of Economic Policy* **17**, 365–388 (2001).
11. Cont, R. Empirical properties of asset returns: stylized facts and statistical issues (2001).
12. Mandelbrot, B. B. Multifractal measures, especially for the geophysicist. In *Fractals in geophysics*, 5–42 (Springer, 1989).
13. Kantelhardt, J. W. *et al.* Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications* **316**, 87–114 (2002).
14. Thompson, J. R. & Wilson, J. R. Multifractal detrended fluctuation analysis: Practical applications to financial time series. *Mathematics and Computers in Simulation* **126**, 63–88 (2016).
15. Faith, J. J. *et al.* Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS biol* **5**, e8 (2007).
16. Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Physical review E* **70**, 066111 (2004).
17. Bloomfield, P. *Fourier analysis of time series: an introduction* (John Wiley & Sons, 2004).
18. Mandelbrot, B. B. & Van Ness, J. W. Fractional brownian motions, fractional noises and applications. *SIAM review* **10**, 422–437 (1968).

Acknowledgments

We would like to express our deepest gratitude to Sander Bais, Jelen Grujic, JP Gonzles, Juniper Lovato, Carla Shedivy and all the faculty of CSSS2016 for putting on an excellent summer program. The Summer School was truly a unique experience that none of us are soon to forget. Thank you, we can't say it enough.