Toward an Analytic Framework for the Electrical Power Grid

Scott Alfeld
Department of Computer Sciences
University of Wisconsin
salfeld@cs.wisc.edu

Carol Barford
Nelson Institute Center for Sustainability and the Global Environment
University of Wisconsin
clbarford@wisc.edu

Paul Barford
Department of Computer Sciences
University of Wisconsin
pb@cs.wisc.edu

ABSTRACT

The large majority of electrical power in the United States today is generated from fossil feedstocks. While renewable energy sources offer compelling alternatives, there are many challenges and complexities that currently limit their use. The high-level objective of our work is to create an analytic framework to provide decision support for renewable energy use in electrical power generation in the US. For security reasons, many of the details of the infrastructure that would facilitate our work are not openly available. Thus, we seek to infer key properties of the power generation and transmission infrastructures, using alternative data sources and recognizing that grid dynamics are constrained by federal regulation and the laws of physics. In this discussion paper we describe the design space for our study and our initial analyses of energy pricing data. These data are openly available from Regional Transmission Organizations and Independent System Operators. Our results highlight the complexities and dynamics of the relationships between locations in the power grid, and set the stage for inferring physical and behavioral properties of the power grid.

Categories and Subject Descriptors

I.6.5 [**Model Development**]: Modeling methodologies

General Terms

Smart Grid, Power Generation, Energy Markets

Keywords

Renewable Energy, Locational Marginal Price, LMP Landscape

1. INTRODUCTION

Fossil fuel-based electrical power generation poses risks to climate, environment and health, but fossil sources such as coal and natural gas are still the primary feedstocks for power generation in the United States. Renewable sources of energy such as solar, wind, and biomass feedstocks offer compelling alternatives to fossil fuels as they can lower net greenhouse emissions and help to alleviate our dependence on foreign fuels. Over the past decade, significant efforts in academia, industry and government have attempted to overcome the impediments to renewable energy use. An important example is the renewable portfolio standards (RPS) that require electricity producers to generate a minimum percentage of power from renewable sources by specified dates [7]. RPS have been adopted in 24 states. However, from the perspective of power producers, utilizing renewable energy sources complicates business processes.

The goal of our work is to develop analytic capabilities that clarify our understanding of the electrical power ecosystem and provide decision support for power producers, including when, where and how renewable energy sources should be used for electricity production. We seek answers to questions such as (1) what are the characteristics and behaviors of the electrical generation and transmission infrastructure?, (2) what are the relative cost implications for the use of sustainable energy sources in local and regional markets?, and (3) where and how might infrastructure changes make renewable energy use more attractive? The challenges in this effort stem from the scale, complexity and dynamics of the electrical power ecosystem and from the fact that many aspects of the physical infrastructure are obfuscated for business or national security reasons.

In this discussion paper, we describe our initial efforts to characterize power generation and transmission infrastructure in the US. We use openly available energy pricing data (from Regional Transmission Organizations and Independent Transmission System Operators that manage regional energy markets) to infer properties of the infrastructure. Our hypothesis is that this data will indirectly reflect both operating norms and physical constraints of the infrastructure. To investigate this hypothesis, our tasks are to *(i)* specify and evaluate models for behavior of the infrastructure, and *(ii)* validate and enhance our analyses with additional data sets. Here, we use real-time energy pricing in the midwest US market to examine the time-variations and relative prices between power producers and thus identify certain properties of the grid infrastructure. Preliminary results of our analysis reveal that the pricing landscape is a dynamic entity that changes dramatically over the course of a day. We also find unanticipated phenomena related to the physical grid that will be explored more fully in future work.

2. RELATED WORK

Our work relates most closely to three main areas of prior study; the geometry of the power grid, distributed generation (DG) placement, and machine learning (ML) techniques used to investigate the power grid.

Electricity market models and simulators are widely used for analysis and forecasting [2]. These models project bidding strategies based, among other things, on characteristics of the transmission network and estimated loads. Previous work explores the topology of the electric grid in terms of both geographical and

e-Energy 2012, May 9-11 2012, Madrid, Spain.

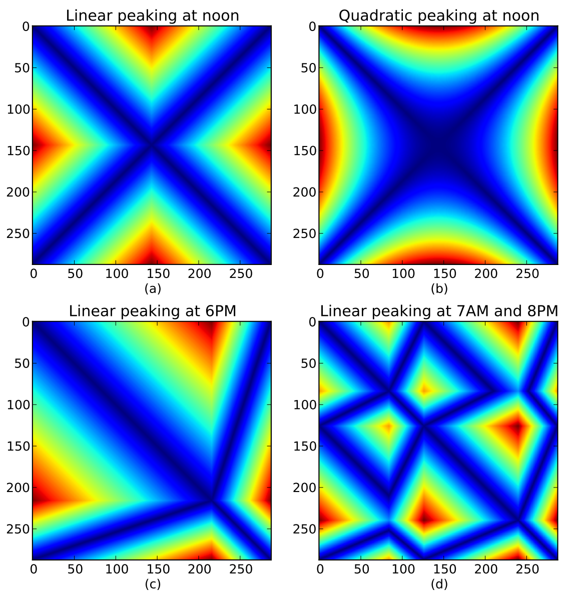

Figure 1: Synthetically generated example $Z^{(f)}$ matrices. If the LMP landscape changes gradually until noon, and then returns to its midnight configuration, $Z^{(f)}$ would look much like (a). If instead the landscape changes smoothly, but more quickly at night than during the day, we see a matrix more similar to (b). Subfigure (c) shows a peak at 6:00PM rather than noon, and (d) shows peaks at 7:00AM and 8:00PM (and a local minimum at 10:30AM).

electrical connections (see [5] and the references therein). Our work can potentially improve the capabilities of forecasting models by providing a more detailed and accurate characterization of the grid's electrical structure.

Placing small-scale power generation facilities at locations near areas that would otherwise be difficult to serve (*i.e.,* Distributed Generation (DG) placement) is becoming increasingly important [1]. This approach yields several advantages over the more traditional large power plants due to the modular design of DG technologies [3]. Finding physical locations for DG placement is a well studied problem and many methods of determining optimal locations have been explored (see [3, 1] and references therein). A better understanding of the market behavior and its relation to geographical location should aid in future DG placement.

ML techniques have been used to analyze power systems for over a decade (*e.g.,* [4]). For example, ML methods have been applied to historical data to predict future power failures [6]. ML techniques have also been applied to the problem of preventative maintenance [8]. This prior work informs our efforts at developing ML-based methods for grid analysis, which will be a critical part of our larger analytic framework for understanding the power grid and markets.

3. DATA

To promote the public interest, regional authorities organize and oversee activities in the electric power grid. Among other things, they manage a bidding process for power production and consumption that results in a Locational Marginal Price (LMP) or *clearing price* for each location or *bus*. There are two distinct markets for energy bidding. On the Day-Ahead market, an LMP for every hour of the day for each location is released the previous day, and on the

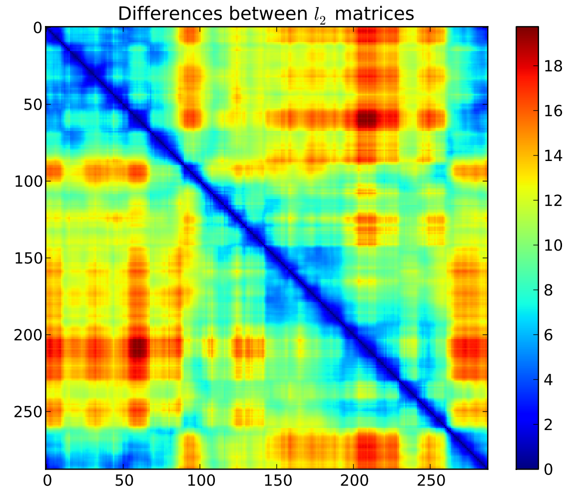

Figure 2: The matrix $Z^{(l2)}$ for Real-Time LMP values in 2010. A low value in matrix location (i, j) indicates that the LMP landscape (as defined by l_2 distance in LMP) during the hour T_i is similar to that during T_j. A high value indicates that the LMP landscapes differ.

Real-Time market, the LMP changes every 5 minutes.

The Real-Time LMP values used in this work are publicly available from the Midwest Independent Transmission System Operator (MISO). The Midwest was chosen arbitrarily but is representative of other regional energy markets in the US. However, geographical locations of the buses are not available from MISO. We were able to determine the locations of buses in 78 cities by hand. One location proved to be an outlier in terms of LMP values during 2010, and was removed. We used data from one bus from each city, yielding a total of 77 buses. For each we have a latitude, longitude, and 105,120 LMP values – one for every five minute period in 2010.

Offers (referred to as bids in the literature) made by power companies reflect the perceived best interest of the controlling company. Power companies typically use proprietary models (based on *e*.g., weather and history) to facilitate bidding. Our hypothesis, however, is that LMP not only reflects the business objectives of power plants, but that it also encodes information about the grid (*e.g.,* congested lines) and other factors. Our challenge is to develop analytic methods that enable these characteristics to be decoded.

Throughout this work we discuss the *LMP landscape* – a term we use to describe the dynamic relationships between LMPs of different buses. Let $G = \{V, E\}$ be a weighted complete graph where the vertices are the buses and an edge exists between all pairs. For each edge (u, v), we define the weight of that edge as $d(u, v)$ where d is a function describing the difference (or similarity) between the LMP patterns of u and v. We explore two specific choices of d. This graph (with its weights), along with the physical locations of the buses, defines the LMP landscape. Our analytic focus is on the dependence of LMP on time and location for the Real-Time market.

4. ANALYTIC METHODS

To quantify the difference between LMP patterns of two buses u and v we use two different functions. The first is the l_2 distance in LMP and the second is the correlation between u and v's LMP patterns. Formally:

$$d_{l2}^{(T)}(u, v) = \sqrt{\sum_{t \in T} \left(u_{\text{LMP}}(t) - v_{\text{LMP}}(t)\right)^2}$$

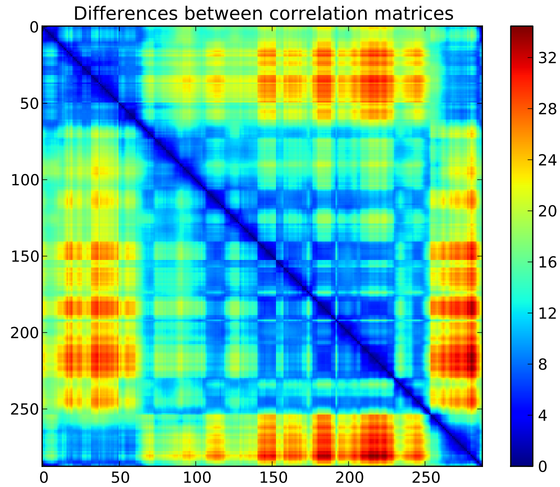

Figure 3: $Z^{(\text{corr})}$ for Real-Time LMP values in 2010. As with $Z^{(l2)}$, a low value indicates similarity (this time as defined by correlation between LMP patterns) in LMP landscapes, and a high value indicates that the LMP landscapes differ.

and

$$d_{\text{corr}}^{(T)}(u,v) = \frac{\sum_{t \in T}\left(u_{\text{LMP}}(t) - \mu_u^{(T)}\right)\left(v_{\text{LMP}}(t) - \mu_v^{(T)}\right)}{|T|\sigma_u^{(T)}\sigma_v^{(T)}}$$

where $u_{\text{LMP}}(t)$ denotes the LMP for the bus u at time t, T is the set of times being considered, $\mu_u^{(T)}$ denotes the average LMP (over T) of u, and $\sigma_u^{(T)}$ denotes the standard deviation of u's LMPs (and similar notation for v). We use correlation because it captures the dependence between the buses' LMP behaviors. The l_2 norm is a standard measure of the difference between vectors. We use it as a measure of error when using one bus's LMP pattern as the prediction of another's[1].

Due to the time-variant nature of sustainable energy sources different times of day have different significance. For example, a solar power plant generates electricity only during daylight hours, and thus the LMP landscape at night has little relevance. For our analysis, we vary the set T, and all times in T have equal weight. However, the methods used easily extend to varying weights over times.

Let N denote the number of buses (in our case, $N = 77$). For a difference function d_f and set of times T, we construct an $N \times N$ matrix $D^{(f,T)}$ where $D_{i,j}^{(f,T)} = d_f^{(T)}(i,j)$. This matrix holds the relations between buses (in terms of LMP patterns) during the times in T. We consider overlapping sets T_i for $i = 1 \ldots 288$ constructed as follows: T_1 contains all times in 2010 between 12:00AM and 1:00AM, T_2 contains all times between 12:05AM and 1:05AM, T_3 between 12:10AM and 1:10AM, and so on. This yields a total of $12 \times 24 = 288$ matrices. To measure the relative rather than absolute differences between these matrices, as we do later, we normalize each $D^{(f,T_i)}$ such that its maximum entry is 1.

These matrices show the relations between buses based on time of day. If, for example, a power plant p can only operate between 12:00AM and 1:00AM every day of the year, the matrix $D^{(f,T_1)}$ describes how the other buses relate to each other during those hours, and thus describes the LMP landscape as it pertains to p. If all D matrices were (roughly) equal, it would indicate that the LMP landscape is static across times of day. As we will see in Section 5, this is not the case.

[1]We use l_2 instead of mean-squared-error only for plotting purposes.

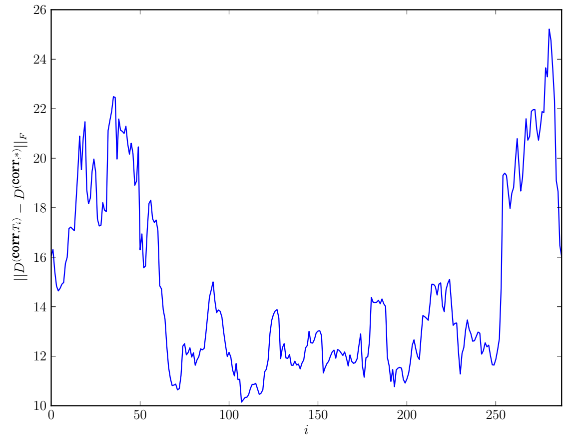

Figure 4: Difference between the one-hour and average (all hours) LMP landscapes, where LMP is defined using d_{corr} (y-axis) vs. time period of day (x-axis).

We now describe how the LMP landscape changes when different times are considered. To this end, we construct a matrix $Z^{(f)}$ such that:

$$Z_{i,j}^{(f)} = ||D^{(f,T_i)} - D^{(f,T_j)}||_F$$

where $i, j \in (1, 288)$ and $|| \cdot ||_F$ denotes the Frobenius norm.

$$||A||_F = \sqrt{\sum_i \sum_j A_{i,j}^2}$$

A large value of $Z_{i,j}^{(f)}$ indicates a substantially different mode of operation in the network, whereas a zero value indicates that there is no difference between the LMP landscapes.

To provide intuition about $Z^{(f)}$, we note the following. For all $i = j$, $Z_{ij}^{(f)} = 0$, and we expect entries near the diagonal to be near zero. For example, we expect $D^{(f,T_1)}$ to be similar to $D^{(f,T_2)}$ because 12:00AM-1:00AM should yield essentially the same relations as 12:05AM-1:05AM. We also expect low values in the off-diagonal corners of the matrix due to the 24-hour periodic nature of the data ($Z_{1,288}^{(f)}$, for example, relates 12:00AM-1:00AM with 11:55PM-12:55AM). If the LMP landscape (or more precisely, the Frobenius norm of $D^{(f,T)}$) increases linearly until noon, and then similarly decreased back to its midnight value, we would see a matrix similar to that seen in Figure 1 (a).

5. RESULTS

Given space limitations, our objective here in reporting results is to demonstrate the scope and efficacy of our methods. We begin by examining the basic relationships between buses during the year 2010 by aggregating LMPs of overlapping one-hour periods of the day. The matrices $Z^{(f)}$ that we obtained demonstrate that the relations between buses vary dramatically depending on which hour of the day is considered. The matrices $Z^{(\text{corr})}$ and $Z^{(l2)}$ show similar structure, but also key differences. It is clear that the true matrices (Figures 2 and 3) exhibit much more complicated structure than the artificial examples shown in Figure 1. Both show some expected structure, such as low values near the diagonal in general, but also unanticipated structure that we believe relates to the physical constraints of the infrastructure.

At times around T_{80} and T_{255} (around 7:00AM and 9:45PM) we see a sharp shift in the LMP landscape, indicated by comparatively high values near the diagonal. This means that the LMP landscape changes drastically around 9:45PM compared to its fairly gradual changes throughout the rest of the day. The evening shift is prominent in both $Z^{(\text{corr})}$ and $Z^{(l2)}$, but the morning shift is far more apparent in $Z^{(l2)}$, indicating a difference (beyond just magnitude and timing) between these two shifts. Our conjecture is that these are caused by the typical working day schedule determining power usage, which in turn affects electricity prices. Further investigation will determine other underlying causes of the shifts and their precise nature.

In addition, we note the periodic nature shown in both $Z^{(f)}$s. The period of 24 hours is expected by the nature of the construction, however there are shorter periods visible in the data. This indicates that, over the hours of the day considered, the structure of bus relations shows a cyclical behavior. To investigate this behavior further, we consider the matrix $D^{(f,*)}$, constructed with T being the set of all times in 2010. This matrix holds the relations between buses when all times of day are weighted equally – intuitively this is the average LMP landscape. For each $D^{(f,T_i)}$ we compute $||D^{(f,*)} - D^{(f,T_i)}||_F$. Results are shown in Figures 4 and 5.

Consider first Figure 4, when $d_f = d_{\text{corr}}$. We note the same cyclical behavior seen in Z, indicating that the LMP landscape departs from the average behavior periodically. The sharp changes in the morning and evening remain present as well. We also see a sudden yet brief dip just after midnight. In addition, we see relatively low values during the day, and higher values at night. This indicates that, during the day, the correlation of LMP patterns between buses better matches the more global view when all times are considered.

When $d_f = d_{l2}$ (Figure 5) this behavior is not clearly evident, nor is the abrupt dip after midnight. The periodic nature, however, persists. While related, d_{l2} and d_{corr} differ from each other, and further investigation is needed to determine what differences cause the observed behavior. We predict an underlying cause is the process of normalizing the D matrices. This causes a loss of information in $Z^{(l2)}$ but $Z^{(\text{corr})}$ is unaffected (all correlation values are in [-1, 1]).

Finally, we see additional structure in the Z matrices. The blocks of low values (as well as those of high values) are axis-aligned, indicating they occur in specific time periods. We also see periodic discrete changes shown more prominently in $Z^{(l2)}$. Further analysis will reveal the causes of these phenomena as well as more subtle effects.

6. CONCLUSIONS AND FUTURE WORK

The objective of our work is to develop decision support tools for electrical power generation that can help to accelerate adoption of green energy sources, aid in the deployment of new infrastructure and improve our general understanding of power grid structure and behavior. While the energy sector and power grid are the focus of a large body of research, empirical study in this domain is limited by the lack of data on the infrastructure and its behavior (due to privacy and security considerations). Our work is based on the hypothesis that publicly available LMP pricing data from regional energy markets can be used to infer and extract information about the power grid.

In this discussion paper, we introduce the general notion of the *LMP landscape*. We describe an initial set of analytic methods to investigate the LMP landscape and apply them to a year-long Real-Time LMP record from a regional US market. We show that the LMP landscape varies dramatically depending on the hour of day, and that there are subtle correlations of behaviors between different locations. We also discovered a periodic relationship with the average LMP landscape that further highlights the efficacy of our approach.

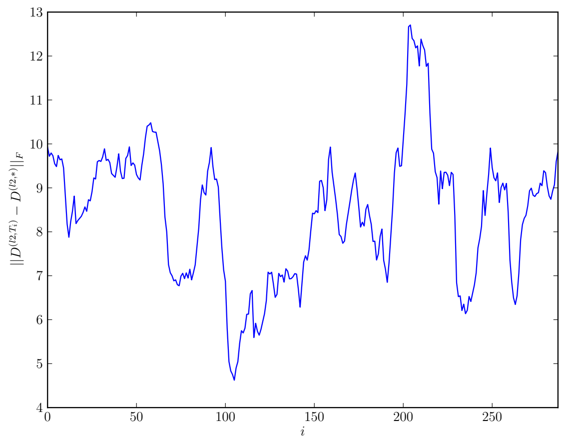

Figure 5: Difference between the one-hour and average (all hours) LMP landscapes, where LMP is defined using d_{l2} (y-axis) vs. time period of day (x-axis).

Our ongoing work will connect the observed phenomena to the specifics of the physical grid as well as the LMP market's behavior. To that end, we are refining and expanding our analytic methods and broadening the data sets that we will consider in our evaluations (*e.g.,* weather characteristics, population densities). Finally, as physical and behavioral characteristics of the grid come into focus, we will expand our analytic framework to include specific characteristics of biomass production, transportation, use and impacts (including geographic implications) so that we can close the loop in our decision support framework.

7. ACKNOWLEDGMENTS

The authors would like to thank Professor Bernard Lesieutre of the University of Wisconsin – Madison's Electrical Engineering Department for his insightful comments during extended conversations. This work is funded by the US Department of Energy Genomics: GTL and SciDAC Programs (DE-FG02-04ER25627).

8. REFERENCES

[1] W. El-Khattam, K. Bhattacharya, Y. Hegazay, and M. M. A. Salama. Optimal investment planning for distributed generation in a competitive electricity market. *IEEE Transactions on Power Systems*, 2004.

[2] R. Frowd and A. Papalexopoulos. Market simulation for lmp forecasting. *Power and Energy Society General Meeting, PES '09*, 2009.

[3] D. Gautam and N. Mithulananthan. Optimal DG placement in deregulated electricity market. *Electric Power Systems Research*, 2007.

[4] N. Hatziargyriou. Machine learning applications to power systems. *Machine Learning and its Applications*, 2001.

[5] P. Hines, S. Blumsack, E. Cotillia Sanchez, and C. Barrows. The topological and electrical structure of power grids. *System Sciences (HICSS)*, 2010.

[6] C. Rudin, D. Waltz, R. N. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. N. Gross, B. Huang, S. Ierome, D. F. Isaac, A. Kressner, R. J. Passonneau, A. Radeva, and L. Wu. Machine learning for the New York City power grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[7] U.S. Department of Energy. EERE State Activities and Partnerships. http://apps1.eere.energy.gov/states/maps/renewable_portfolio_states.cfm, 2012.

[8] L. Wu, G. Kaiser, C. Rudin, D. Waltz, R. Anderson, A. Boulanger, A. Salleb-Aouissi, H. Dutta, and M. Pooleery. Evaluating machine learning for improving power grid reliability. *ICML 2011 Workshop on Machine Learning for global Challenges*, 2011.