Smart Grid

Rachel Aronow, Aaron da Silva, Rose Dennis, George Kevrekidis,
Richard Touret, Lance Wrobel

Department of Mathematics - University of Massachusetts Amherst

Abstract. *Global depletion of resources and increased demand on our electric grid have resulted in a need for innovation in support of an intelligent utility network — the Smart Grid. Thus, the enhancement of electrical power interactions for both utilities and customers has become a major focus for current research. At a microscale level, we use gradient boosting models to predict electrical power consumption and solar power generation for a single home. We also develop a dashboard as a visualization tool for this analysis. At a macroscale level, we explore the cost optimization of power dispatch between various energy sources in a simplified model of the German electric grid. Through these levels of study, we aim to establish a deeper understanding of the potential requirements of a modern Smart Grid.*

1. Introduction

As technology advances, current societal trends have increased the need for efficient electrical production. This increased demand has developed a highly intricate problem in the field of data analytics. In light of this, scientists, mathematicians, and statisticians have designed, constructed, and improved a smart grid -- a multidimensional network that monitors, measures, and manages the transport of electricity. The original smart grid only included the exchange of coal-generated electricity between individual homes, corporate buildings, utilities, and power plants. However, solely relying on these primary sources of energy has become unsustainable. Thus, there is a need for efforts towards more innovative and renewable energy methods. Wind power, biomass, hydroelectric, and solar power are just a few of these kinds of renewable energy approaches that are naturally replenished on a human timescale. We focus on energy generated from solar power for forecasting analysis.

With these advancements in obtaining energy, it has become imperative to analyze their input to the overall smart grid. The relationship between how much electricity a house/building/unit needs and how much solar energy the respective building can produce proves to be vital. The forecasting of solar generation and power consumption and the effective dispersion of electricity are two important factors that add value to the whole smart grid. The way renewable energies are integrated into the grid affects the overall performance of the smart grid. Thus, this project's goals were two-fold. On a smaller scale, we look at forecasting the power consumption and solar power generation of a single home in the Amherst area. We build a dashboard to visualize these predictions. On the larger scale, we consider a simplified version of the german electric grid, extracted from OpenStreetMap, and estimated average values of power generation and demand to simulate the optimal dispatch of power among various energy sources. Lastly, we consider a preliminary stochastic model for setting a renewable energy policy.

2. Single Home Forecasting

2.1. The Data

We used a single home's electricity consumption and solar generation data. We also used weather data from around the home to complement the electrical data. Data was obtained from a collection of datasets called the Smart* Data Set for Sustainability which is part of the Umass Trace Repository.

2.2. Model Predictor Selection

We began with a full model of 14 predictors. Then we ran a multiple linear regression model. We performed forward, backward, forward stepwise and ANOVA variable selection using an AIC criterion of α =.05 to choose which variables to keep in the model. We chose the method that resulted in the greatest adjusted R-squared value in the model results.

2.3. Model - EWMA & Gradient Boosting Models

In order to establish a model of these forecasting methods, we incorporated an exponentially weighted moving average (EWMA) term as well as a Gradient Boosting Model (GBM). An EWMA is a first-order infinite impulse response filter that applies weighting factors which decrease exponentially. This metric was a strong predictor in both our use and gen gradient boosting models. We included this metric to capture recent trend information in the time series of use and gen. Mathematically, the formula for EWMA is as follows:

$$S_t = \alpha[Y_t + (1-\alpha)Y_{t-1} + 1 - \alpha + (1-\alpha)^k Y_{t-k}] + (1-\alpha)^{k+1} S_{t-(k+1)} \quad (1)$$

EWMA (at time t) S_t is a weighted sum of the datum power values. Y_t is the weight of the initial datum point Y_{t-i} is $\alpha(1-\alpha)^i$. A GBM is a machine learning model which builds an ensemble of decision trees in a stepwise fashion so that the latter models learn from the mistake of the prior models.

2.4. Dashboard

We developed a interactive analytics and visualization tool for assessing how variations in a set of parameters affect the predictive accuracy of solar power generation and power consumption forecasting models applied to a single home. The tool was created using RShiny and a variety of other R packages. There are four main tabs on this dashboard: instructional, summary statistics, power consumption and solar power generation. The power consumption and solar power generation pages both include a plot of the test data, a table of predictions on the test data which includes the actual values for comparison, the mean squared error of the model on the test data, and a control panel used to adjust various parameters (listed below). The summary statistics page contains basic information and visualizations of the dataset being used.

The adjustable parameters on the power consumption and solar power generation tabs include:

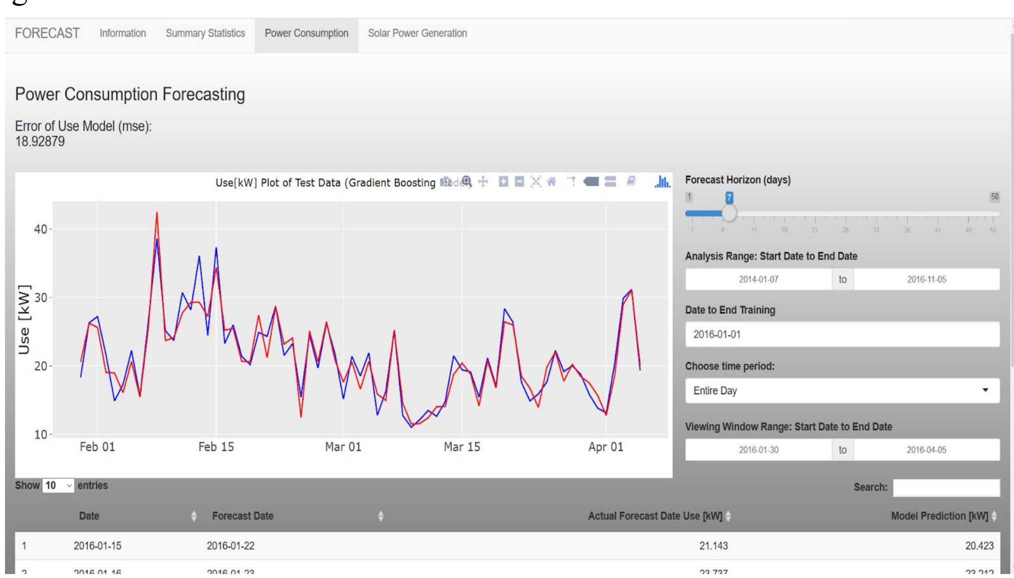
- **Forecast Horizon:** length of time into the future where the prediction is made
- **Analysis Range:** any window of time from January 1st, 2014 - December 15th, 2016 to use for training and testing data
- **Date to End Training:** date which splits analysis range into training and testing data
- **Time Period:** range of hours during the day for which the predictions are made
- Viewing Window Range: any window of time within the test data range

The forecasting models use a fixed number of predictors variables which are related to weather and electricity. The variables are summarised below.

- **Predictor variables used in demand prediction:** current day's power use (over specified time period), exponentially-weighted moving average of use over previous n days (n is forecast horizon), current day's humidity and temperature, day of the week, season and month.
- **Predictor variables used in solar prediction:** current day's power generation (over specified time period), exponentially-weighted moving average of generation over previous n days (n is forecast horizon), season month, and current day's humidity, temperature, visibility, wind bearing, pressure, precipitation and cloud-cover.

A screenshot of the dashboard is given below:

Figure 1: Screenshot of the dashboard



3. Electric Grid Simulation

3.1. OSM power graph generation

As we transition into a macroscale problem, a fundamental first step is obtaining a full featured power grid that is compatible with our power system analysis libraries. In order to yield meaningful results, we established two mains goals that guided us in this step:

1. The power grid data must describe an actual grid, designed and currently used in a developed country
2. The grid will be simplified in graph form while retaining every parameters of the power grid.

In order to achieve these two goals, we used Open Street Map, Osmosis, and SciGrid in this respective order. Their purpose is as follows:

-Open Street Map (OSM) allows the user to export precise geographical information in a given zone.

-Osmosis will process OSM data and filter elements belonging to the power grid.

-SciGrid takes filtered OSM data and abstracts it into a graph that showcases "relations", when two stations are linked by a line. SciGrid also populates a postgreSQL database with this graph that we can subsequently call in power system analysis functions. The following example showcases the importance of this process: if we observe power lines in a city, we can easily see that two stations are linked by segmented lines to circumvent topography or other obstacles. OSM preserves this complexity, and the abstraction process is necessary to represent the line as an edge instead of multiple segments. Power analysis packages are unable to work with segmented links.

We decided to use the German grid since it is by far the most descriptive OSM data available as of 2019.

3.2. Non-linear optimal power flow

For the next phase of our project, we looked at using this simplified German electric grid to do an analysis of the optimal dispatch of power throughout a typical 24 hour day. At each of our hourly snapshots, we used a python package called Pypsa[1] to solve the linearized power flow equations at each node of the OSM graph, and find the solution that minimized our cost function.

The cost function used in this analysis is given below:

$$\sum_{n,s} c_{n,s}\overline{g}_{n,s} + \sum_{n,s} c_{n,s}\overline{h}_{n,s} + \sum_{l} c_l F_l + \sum_{t} w_t[\sum_{n,s} o_{n,s,t} g_{n,s,t} + \sum_{n,s} o_{n,s,t} h_{n,s,t}] + \sum_{t}[suc_{n,s,t} + sdc_{n,s,t}] \qquad (2)$$

The indices n, s, and t refer to the buses, generators of each energy type located at each bus, and time step, respectively. While the term "bus" typically refers to an electric bus, in this study we refer to any graph node for which we want to evaluate power flow at as a "bus". Therefore

[1] https://pypsa.org/

we can attach multiple components (generators, loads, transformers, etc) to a single bus. The types of energy that can exist at each node are solar power, gas, offshore wind, onshore wind, run of river hydroelectricity, brown coal, hard coal, and nuclear power. A full description of the variables in the equation is given in table 1 below. Essentially, the function we wish to minimize describes the sum of the cost of extending nominal power of any generators in the network, the cost of increasing the nominal level of power in storage, the cost of increasing a line's capacity for power flow, the operational costs for dispatch of power generation and storage, and the shutdown and startup costs for each power source. The data used for calculating such values were found in Brown et al. 2018. These include rough estimates for the power demand, range of achievable power generation for each energy type, and marginal costs of power generation and storage, broken down by location and time of day (hourly over a 24 hour period).

- Table 1: variables considered for cost minimization

Notation	Definition
$c_{n,s}$	Cost of extending nominal power of generator s, at bus n, by one MW
$\overline{g}_{n,s}$	Nominal power of generator s located at bus n
$\overline{h}_{n,s}$	Nominal storage of generator s located at bus n
c_l	Cost of operating line
F_l	Capacity at line 1
$o_{n,s,t}$	Marginal cost of dispatching generator s, at bus n, for one MWH at time t
$g_{n,s,t}$	Dispatch of power at generator s, at bus n, at time t
$h_{n,s,t}$	Dispatch of storage at generator s, at bus n, at time t
w_t	Weight for time t. Used to vary marginal costs of power by time of day
$suc_{n,s,t}$	Startup cost of generator s at bus n and time t
$sdc_{n,s,t}$	Shutdown cost of generator s at bus n and time t
K_{nl}	Incidence matrix. K_{nl} takes a value in $\{-1, 0, 1\}$, depending on whether the line l starts or ends at the node n
$d_{n,s,t}$	Excess load generated by power source s at node n at time t

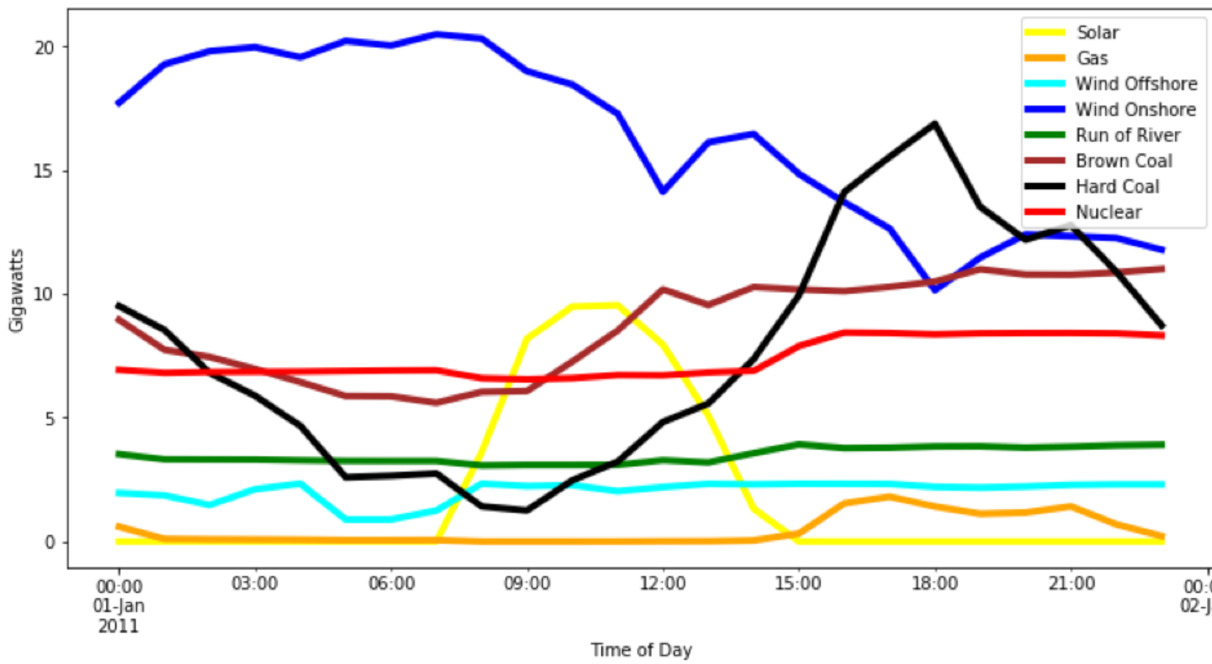
Throughout this optimization, we must also obey the law of conservation of energy (Equation 3). This equation ensures that at every time step, any energy generated or pulled from storage at a node, plus/minus the power flowing in/out of connected nodes, has to sum up to the excess load generated to match the demand at any given node.

$$\sum_{s} g_{n,s,t} + \sum_{s} h_{n,s,t} - \sum_{s} f_{n,s,t} - \sum_{l} K_{nl} f_{l,t} = \sum_{s} d_{n,s,t} \qquad (3)$$

Lastly, the current formulation of the cost optimization function in pypsa allows for any given line in the network to reach its full capacity for carrying electric power. However in practice we do not want an electric grid operating at or near full capacity, because this risks the lines overheating. If even one line fails, this can lead to cascading failures and blackouts throughout the grid. To reduce the risk of our optimization finding a solution that would push the network close to these limits, we used an approximation as recommended in the pypsa tutorial where we set each line's capacity to be only 70% of its true value.

After solving for the optimal solution at each time step, we can plot the generated power as a function of the time of day, per type of energy source (Figure 2) as in Brown et al. 2018. The most apparent trend is the resulting bell-curve of solar power. We see that most of the solar energy is being produced during the mid-day hours, when we'd expect the production of solar energy to be most efficient. During these hours, the production of non-renewable sources such as coal decreases. However we see an uptick in the use of hard coal around evening hours. This makes sense, because as people are returning home from work, this creates a peak time for home electricity use. Since solar power cannot be generated in the evening, the generation of non-renewable power has to increase in order to pick up the excess demand, and it becomes cost-optimal to produce more electricity via coal. Overall, this optimization produces predictable results, but there is one odd trend in the decline of on-shore wind production throughout the day. Since this 24-hour period should in theory be cyclical, we would expect the 12 GW of energy produced during the last hour of the day not to differ greatly from the initial 17.5 GW produced starting at midnight of the previous day. This can likely be explained as a stochastic effect - this data comes from a single day of observations, where it was likely much windier at the start of the day than the end of the day, changing the range of achievable power generation for onshore wind.

Figure 2: Power dispatch throughout a 24-hour period, by energy source (below)



4. Policy Optimization using Stochastic Programming

As seen so far, the presence of renewable energy sources in our smart grid is beneficial in several counts, and therefore likely to increase substantially over the next years. This incorporation is gradual, however, due to the large-scale and multifaceted infrastructure it requires, and also subject to constraints (e.g. supply, budget, geography). Guaranteeing that this happens in an optimal manner would therefore allow us to minimize costs or maximize the impact of decisions made in the earlier stages of incorporation.

One way to tackle this problem mathematically is using a two-stage stochastic programming approach. We treat renewable energy production as a limited resource that is stochastically produced at certain locations and then distributed throughout a consumption network so that energy demand is satisfied, while accounting for uncertain energy production and energy loss due to transportation. Many variations in the model formulation are possible, but the very general case of the problem is a Multi-Commodity Optimal Flow problem, where after weather realization (energy production) at certain nodes, energy is distributed in an optimal (cost-minimizing) manner throughout nodes of a graph.

The idea to use this type of model, which is in some way inherit to the nature of the problem, is inspired by Barbasoglu and Arda[2], who develop a two-stage stochastic model (superposed with an optimal flow) to address earthquake aid distribution (in personnel and commodities) in Turkey.

[2] Barbarosoglu, G. and Arda, Y. (2004) A Two-Stage Stochastic Programming Framework for Transportation Planning in Disaster Response. Journal of the Operational Research Society, 55, 43-53.

It is possible to solve this problem efficiently using linear programming, but the model itself is too complicated to include here. A complete formal representation of the model will be included in future work, but here we present a simplified version as a proof of concept:

- First Stage Decision Variables:
  - l_i number of solar panels located at node i
- Second Stage Decision Variables:
  - k_i conventional energy consumed at node i
  - w_{ij} amount of renewable energy transported from node i to j
- Constraints
  - $\sum l_i \leq [\textit{number of solar panels available}] -$ (supply)
  - $s_i l_i \geq \sum w_{ij} -$ (production is greater than distribution)
  - $\sum_j c_{ji} w_{ji} + k_i \geq d_i -$ (demand fulfillment)
  - $k_{i,} l_i, w_{ij} \geq 0 \;\forall\; i,j$
- Objective
  - $V = \sum c_i k_i$

Where c_i, c_{ij}, s_i are appropriate cost and yield coefficients. This is a simplified problem that assumes only one commodity distributed over an undirected graph, with weather conditions realized independently at each node.

While more detailed data sets are available even for this simplified model presentation, we solved it using freely available annual-average data (sunshine, electricity consumption and cost), aggregated on government websites. When limited to a six-node model for New England, as represented on the contiguous graph of the US, the solution is rather disappointing: robustly, under reasonable assumptions, all solar energy production and consumption is suggested to be done in CT, with all other states consuming more expensive conventional electricity (from an infinite supply). After taking a closer look at the dataset, this is a reasonable suggestion due to the extraordinarily high price of electricity in the state compared to others…

Overall, this result shows that annual solar energy production at a state-level does not have the appropriate scale for the proposed model to be effective. Variations are too small and because energy transfer is expensive (big losses over large distances), the production will almost always be local. We expect that applying the same idea to a more localized wind-generation model will have more success, given the larger variation of wind patterns over a landscape (compared to sunlight exposure), and the smaller distances that would be inherent in the problem.

5. Conclusions and future work

A deeper understanding of Smart Grid technology and methods was achieved through various analyses at a smaller scale. Gradient boosting models also proved to be an efficient way to predict both power consumption and solar power generation for a single home. In order to perform more comprehensive and rigorous studies, data at a larger scale would be necessary. Further analysis would include scaling up forecasting for larger areas of society. This would include forecasting solar power generation and power consumption for cities, towns, states, and regions. Grid operation analysis would also be enhanced through the study of more relevant regions, such as New England. The study of renewable energy and costs of power would also benefit by acquiring more relevant data.

Works Cited

Barbarosoglu, G. and Arda, Y. (2004) A Two-Stage Stochastic Programming Framework for Transportation Planning in Disaster Response. Journal of the Operational Research Society, 55, 43-53.

Medjroubi, Wided, et al. “Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models.” Energy Reports, vol. 3, 2017, pp. 14–21., doi:10.1016/j.egyr.2016.12.001

T. Brown, J. Hörsch, D. Schlachtberger, PyPSA: Python for Power System Analysis, 2018, Journal of Open Research Software, 6(1), arXiv:1707.09913, DOI: 10.5334/jors.188