NAME: Micah Botkin-Levy
DATE: 4/2/19
Homework: 06
STAT/CS 387

1 Preliminary Data Analysis

The main dataset that I plan on using is historical Real-Time Five-Minute LMPs from ISO New England. ISO New England (ISO NE) is an independent, non-profit grid operator for all the New England states. ISO NE ensures the day-to-day reliable operation of New England's bulk power generation and transmission system as well as oversees the administration of the region's wholesale electricity markets. One of these markets is the Real-Time Energy Market which lets market participants buy and sell wholesale electricity during the course of the operating day. Locational marginal prices (LMPs) are prices that determine the cost of supplying one additional unit of energy at a particular spot on the grid.

ISO NE publishes the final version of the LMPs here https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/lmps-rt-five-minute-final which can be downloaded in 6 day increments. As of now I have downloaded five minute LMPs for all of February and March of 2019. This data set includes 16980 observations of five minute LMPs for 1186 different nodes in the system. Supplemental to the LMP dataset is metadata provided by ISO NE on the 1186 different nodes in the system. For now, I have selected to use prices associated with Location ID 363 which is Burlington, VT. The dataset read in as a pandas DataFrame is just over 270,000 bytes in the python memory with each price stored as a 64 bit float.

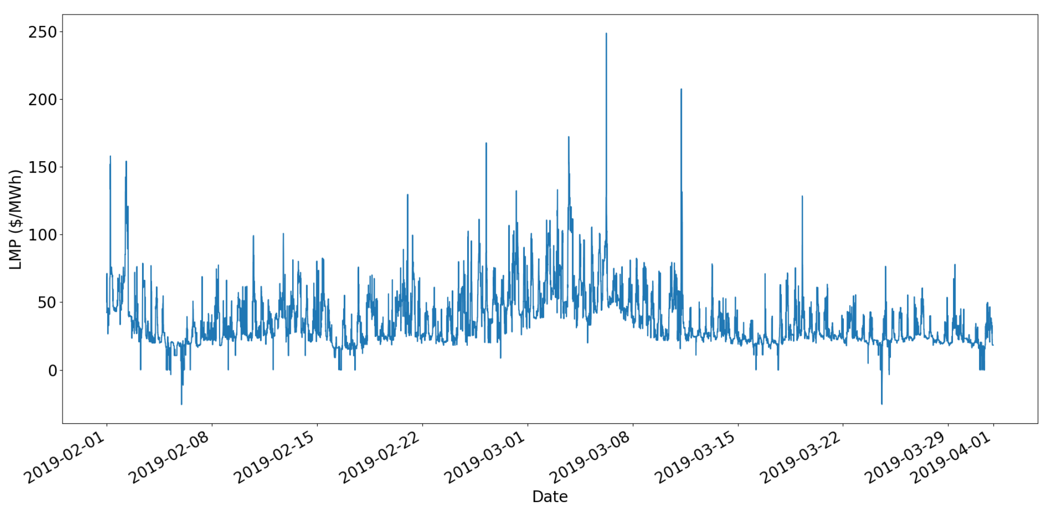

Figure 1: Timeseries Plot of ISO NE Five-Minute LMP 363 for Feb-Mar 2019

A timeseries plot of the five minute data for February and March can be seen in Figure 1. This view gives us a sense of the range of prices but since there are so many observations it does not give detailed intuition into the distribution. Next, we look at a histogram of the data in Figure 2. Here we can gain some knowledge that the majority of prices are between 20 and 50 $\frac{\$}{MWh}$. A similar intuition can be gained from the probability density function estimation plot seen in Figure 3.

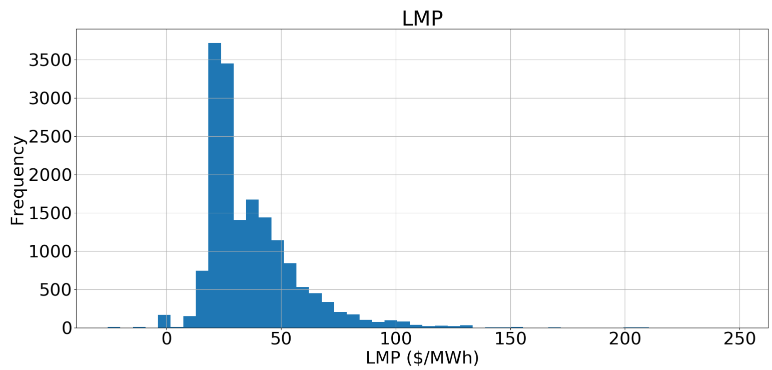

Figure 2: Histogram Plot of ISO NE Five-Minute LMP 363 for Feb-Mar 2019

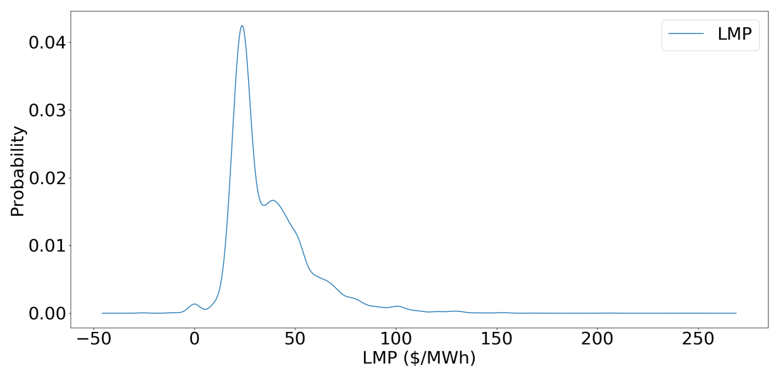

Figure 3: Probability Density Function Plot of ISO NE Five-Minute LMP 363 for Feb-Mar 2019

However, the graphics in Figure 2 and Figure 3 do not give good knowledge about the magnitude and occurrence of any outliers in the data. In order to see this better, we recreate the histogram plot with a log Y axis scale as seen in Figure 4. With the current data set the lowest prices are around -15 and highest prices are 200-250 but these only occur a handful of times in the two months. Our intuitions of the data can be confirmed by looking at some statistical measures in Table 1.

Since I am getting my data from a single reliable source and I have only downloaded two months of data, I have not run into any data quality issues yet. Some potential issues that I could run into would be missing timestamps, empty values, or erroneous values. If I need to download more data and end up running into some of these data quality issues I can use linear interpolation to replace or add the erroneous data.

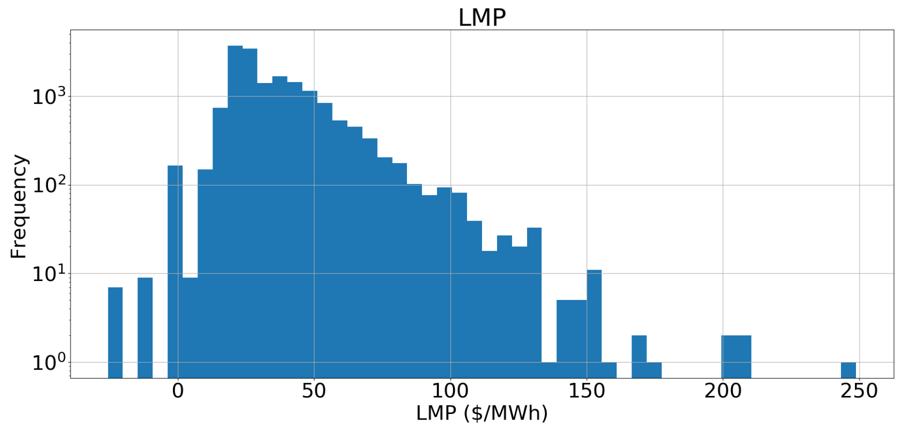

Figure 4: Log Histogram Plot of ISO NE Five-Minute LMP 363 Dataset for Feb-Mar 2019

count	mean	std	min	25 %	50%	75%	max
16980	36.7	20.0	-25.63	23.19	30.16	45.58	248.84

Table 1: Summary Statistics of ISO NE Five-Minute LMP Dataset for Feb-Mar 2019

I plan on using the LMP data collected as the simulation data for testing the Q-learning methods of control for an energy storage system. The LMP data will determine the reward for either charging or discharging at each time step. The main paper I am basing the project on used two years of hourly data for training and testing the Q-learning algorithm. This equates to 17,520 observations which is very close to the 16,980 data points I currently have. However, if this is not enough data to properly assess the performance of Q-learning it will be easy to gather more data from the ISO NE website.

So far, I do not see the scope of my project changing since my proposal. During the literature review, I have discovered a lot of interesting extensions or more advanced methods that I would like to try however, this will depend on how much time I have. The next steps will be to program the basic Q-learning algorithm and test that it does perform as expected. The following is a general workplan for the rest of the project:
Week of April 8th: Implement Q-learning Algorithm
Week of April 15th: Debug and tune algorithm performance
Week of April 22nd: Compare Q-learning algorithm to baseline methods, such as full forecast optimization and random actions, and extend methods/application, time permitting
Week of April 29th: Prepare and record presentation, write scientific article and technical report

2 Initial Literature Review

1. C. J. C. H. Watkins, “Learning from delayed rewards,” 1989.

   Summary Watkin’s PhD thesis presents some early work on reinforcement learning. He frames the methods in terms of animal behavior and presents Markov decision process as a good model of animal behavior. He states that stocastic dynamic programming could be used to calculate an animals behavior’s but an animal would never use it to make decisions or learn optimal policies. Instead, he introduces an incremental Monte-Carlo method for learning optimal values of actions that does not require a model of the environment. This method is latter dubbed Q-Learning.

   Relevance Watkins produces a truly novel piece of work in his PhD thesis. The method he introduces later becomes a prominent statistical learning method which enables learning the value of taking certain actions without a model. This method is perfect for energy storage control using real time prices as formulating a model for these prices is very difficult. His thesis presents a useful method that will be the basic brains of my project.

   Further references Most of the referenced work are journal papers for early statistical learning work. There is a pretty wide spread of ages of referenced sources starting in 1955 to when the PhD was published in 1989. The most relevant works in the references are about Markov chains such as Mandl, P “Estimation and Control in Markov Chains” (1974) and building on early work in reinforcement learning like the method of temporal differences from Sutton’s “Temporal Credit Assignment in Reinforcement Learning” (1984) and “Learning to Predict by the Methods of Temporal Differences” (1988)

2. C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, pp. 279-292, May 1992.

   Summary Walkin and Dayan’s work expands on the idea from Watkin’s thesis. They present a convergence theorem for Q-learning. The authors state that the Q-learning method will always converge to the optimum action-values as long as all actions are repeatedly sampled in all states and action-values are represented discretely. Additional extensions of the methods are also examined for different cases of Markov environments.

   Relevance This piece of work is extremely relevant for the final project. They take ideas from thesis which weren’t fully fleshed out and prove that it can work. The convergence guarantee allows us to state that we are taking optimal actions if formulating the problem correctly. This work enables us to be more confident in using the Q-learning method in the project.

Further references The authors include a small reference section mostly overlapping with Watkin's thesis references. They cited works are mostly conference or journal papers from the 70s and 80s. However, they include some new reference from after the thesis. These are the most relevant such as Lin. L "Self-improving reactive agents based on reinforcement learning, planning and teaching" (1992) and Mahadevan and Connell "Automatic programming of behavior-based robots using reinforcement learning" (1991) which are examples of researchers using Q-learning and furthering its development.

3. G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*, vol. 37. University of Cambridge, Department of Engineering Cambridge, England, 1994.

   Summary This paper from Rummery and Niranjan push the boundaries of the early Q-learning literature by comparing modified Q-learning algorithms. The authors argue that the discrete nature of the classical Q-learning is too restrictive and therefore must be expanded to high dimensional continuous state space. The authors compare some recent modifications to Q-learning from Sutton, Peng, Lin as well as introduce a new algorithm, Modified Connectionist Q-Learning. Their work is an early predecessor to deep Q-learning and an important conclusions about robustness and performance are made.

   Relevance This paper provides useful extensions of the Q-learning algorithm for energy storage project. While the energy storage application can be treated as discrete, in the real world it would be much more realistic to use continuous control actions. This work cites some modified Q-learning algorithm to pursue this if time permits.

   Further references This early work on Q-learning on cites 12 other works. They are a mix of PhD thesis, conference papers, and publications. Since the study of Q-learning is so new all referenced works are within ten years of the publication.

4. R. S. Sutton and A. G. Barto, *Reinforcement learning - an introduction*. MIT Press, 1998

   Summary The book by Sutton and Barto serves as an introduction to reinforcement learning. After an introduction chapter the book is split into three parts. The first part covers tabular solution methods such as multi-armed bandits, Markov Decision Processes, dynamic programming and Q-learning. The second section looks at approximate solution methods such as on-policy prediction and control. The book finishes with a deeper dive into the psychology and case studies of reinforcement learning.

   Relevance The book serves as a good refernce of the basics of reinforcement learning. It covers temporal-difference learning and specifically Q-learning which is the algorithm I will use in the project. It also covers many of the motivations, challenges, and tradeoffs with reinforcement learning.

Further references As expected for a book of this size there are quite a lot of references. The work cited are from a vast array of sources and publish years. The most relevant references are those related to Q-learning such as Watkin and Dayan (1992), Peng and William (1996), and Duff (1995).

5. H. Wang and B. Zhang,"Energy storage arbitrage in real-time markets via reinforcement learning," *CoRR*, abs/1711.03127, 2017.

   Summary The authors of this work perform real-time temporal arbitrage on energy storage using reinforcement learning. The main idea is by repeatedly performing charge and discharge actions under different real-time prices, the proposed reinforcement learning-based policy learns the best strategy that maximizes the cumulative reward. Their strategy uses the Q learning algorithm and applies two reward schemes to the problem for hourly realtime LMP prices for 2 years in ISO New England. They compare the performance of the Q-learning scheme to a baseline algorithm.

   Relevance The paper by Wang and Zhang is extremely relevant to the project as it was my inspiration and forms the basis of what I plan on working on. It provides a useful starting point for applying the Q learning algorithm on a energy storage system. I plan on also using real time market prices as the driving factor of the reward but will be using 5 minute ISO NE data instead of hourly data.

   Further references The reference for this paper is a combination of energy focused works from IEEE transactions and Applied Energy and others references such as energy policy and statistical learning focused. Most of the references cited are around 2011-2015, however, the Q learning scheme is much older. The authors base the energy storage problem off recent work but use methods that have been around for awhile. The most relevant references are the original Q-learning work such as Sutton and Watkins, "Online modified greedy algorithm for storage control under uncertainty" (2016) which is the baseline algorithm for the comparison and another approach to a similar problem in "Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming" (2015).

6. J. Eyer, G. Corey, and S. N. Laboratories,*Energy Storage for the Electricity Grid: Benefits and Market Potential Assessment Guide : a Study for the DOE Energy Storage Systems Program.* SAND (Series) (Albuquerque, N.M.), Sandia National Laboratories,2010.

   Summary This study which is performed by Sandia National Laboratories for the Department of Energy assesses the potential benefits and economic market potential for energy storage. There is quite a lot of information covered in this work including different energy storage technologies and the models used to represent them. The work also looks at all the different storage applications and benefits that can be gained from them. Also included are the challenges that energy storage faces and areas for further research.

Relevance While there is certainly more information than needed in this work it is a very useful reference for this project. The first important contribution is serving as a reference for the energy storage technology and discharging and charging models to be used. Secondly, it lists the possible energy storage applications including real-time arbitrage which will shape the reward function for Q-learning.

Further references This work cites quite a lot of references. The majority of which are other publications from department of energy sponsered laboratories. These publications are from a variety if years but mostly from after 2000. The most relevant references are those that give further background on energy storage systems like "Dynamic Operating Benefits of Energy Storage" (1986) and their role in electricity grid "EPRI-DOE Handbook of Energy Storage for Transmission and Distribution Applications" (2003).

7. R. H. Byrne and C. A. Silva-Monroy, "Estimating the maximum potential revenue for grid connected electricity storage: Arbitrage and regulation," Sandia National Laboratories, 2012.

   Summary This study which is performed by Sandia National Laboratories for the Department of Energy looks in depth into two potenetial sources of income for energy storage: energy arbitrage and frequency regulation. Energy arbitrage referes to using control to buy energy when prices are low and sell when prices are high. Frequency regulation is related to an ancillary services market which attempts to maintain the system frequency. The work includes explanations of energy storage dynamics, optimization formulation, and case study results.

   Relevance This study is quite important for the project. First, it validates the linear dynamic model of energy storage which includes the charging and discharging power at each time step. Secondly, it provides a formulation for a linear optimization problem to determine the maximum the potential revenue. This could come in useful to provide a comparison to the Q-learning approach.

   Further references There are a surprising low number of references for this study. The works cited are a wide variety of mathematical and model background knowledge and other studies on energy storage. Most of the references are from the late 2000s. Useful references include previous cited "Energy storage for the electricity grid: Benefits and market potential assessment guide" (2010) as well as other references on the economics and modeling: "Power System Economics : Designing Markets for Electricity" (2011), "An Introduction to Optimization" (2001).

8. J. Qin, Y. Chow, J. Yang, and R. Rajagopal, “Online modified greedy algorithm for storage control under uncertainty,” *IEEE Transactions on Power Systems*, vol. 31, pp. 1729-1743, May 2016.

   Summary The goal of this paper was to find a method to operate energy storage under uncertainty. The uncertainty they introduced was stochastic price and demand for a system. The authors propose a simple online deterministic optimization with an offline component to determine the optimal control of the energy storage system. The paper also includes a derivation of performance guarantees.

   Relevance This work is relevant as it serves an example of a non-statistical learning method for solving a similar problem to the one I would like to solve. I will not have time to actually program and run their algorithm but it is useful to know the advantages and downfalls of other approaches.

   Further references Most references of the paper are IEEE journal papers from transactions on smart grid or transactions on power systems. The majority of works cited are from within 6 year prior to posting but some of the fundamental methods are from 80s and 90s. The most relevant works cited are other solutions to similar energy storage problems such as “Fast MPC-Based Coordination of Wind Power and Battery Energy Storage Systems” (2012) and “Optimal Electric Energy Storage Operation” (2012).

9. G. Henri and N. Lu, “A multi-agent shared machine learning approach for real-time battery operation mode prediction and control,” in *2018 IEEE Power Energy Society General Meeting (PESGM)*, pp. 15, Aug 2018.

   Summary The paper by Henri and Lu aims to perform real time battery operation across multiple energy storage devices using a machine learning approach. The general idea is that all devices share data to a learning aggregator which trains algorithms for the devices to use. The learning aggregator uses a neural network with one layer and 20 neurons. The paper compares this approach with an economic model predictive controller method and finds the new approach was faster, more accurate, and returned higher savings.

   Relevance The paper is relevant to the final project because it shows an alternative statistical learning method to a similar problem. A slightly different application and data set are used but some pieces of information are still helpful. They define incremental energy storage modes besides just idle, charge, and discharge which I could use in my model.

   Further references Most references for this paper are IEEE transactions journals with transaction on smart grid being the most common. The works cited tended to be slightly older around the years 2013-2014 but all after 2007. Relevant cited work to the project include "A Machine Learning Approach for Real-time Battery Optimal Operation Mode Prediction and Control"" (2018) which was a precursor to the paper and "Modeling for residential electricity optimization in dynamic pricing environments" (2012) which helps motivate the energy storage objective and control goals.

10. A. S. Zamzam, B. Yang, and N. D. Sidiropoulos,"Energy Storage Management via Deep Q-Networks," *arXiv e-prints*, p. arXiv:1903.11107, Mar 2019.

    Summary The work published just last month looks at how deep Q-networks can serve to operate energy storage unit in a microgrid with renewable generation. The paper enhances Q learning by using neural network to approximate the action-value function for continuous states. This approach avoids the discretization needed for classic Q learning. The authors compare results to an optimal solver approach with full knowledge of future prices and energy demand and generation as well as compare to a model predictive controller approach with knowledge of the short future.

    Relevance The work provides a extension of the Q-learning method. While it is likely be too complicated an adaptation for the final project, it is still a useful reference. I liked the two baseline algorithms that they used as comparison and will try and use these as a comparison to the Q-learning algorithm I choose if there is time.

    Further references Most of the references are IEEE either transactions on smart grid, power systems or CDC. The rest of the cited works are reinforcement learning texts or applications of them. The majority of reference are from the past 6 years with statistical learning reference coming from much earlier. The most relevant references are "Optimal operation and economic value of energy storage at consumer locations"" (2017) and "Optimal operation and economic value of energy storage at consumer locations" (2015) which are referenced as other approaches to a similar problem.

11. C. Guan, X. Lin, Y. Wang, , S. Nazarian, and M. Pedram, "Reinforcement learning-based control of residential energy storage systems for electric bill minimization," in *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 637-642, Jan 2015.

    Summary This work studies energy storage control using TD(λ)-learning algorithm under renewable generation uncertainty. The model of the system includes an energy storage system and a photovoltaic system connected to the grid. The authors introduce the TD(λ)-learning as a reinforcement learning algorithm to perform cost savings. They perform a case study and compare the results to a baseline strategy which charges the energy storage during off-peak hours and discharges during peak hours.

Relevance This work provides another slightly variation to the Q-learning style reinforcement learning algorithms. The algorithm and results will be a useful comparison the classic Q-learning algorithm and can be implemented if time permits.

Further references The work only cites 13 references some of which are more applied IEEE journal papers while others are references to mathematical and stastical theory papers. The references are mostly from the 2000s. The most relevant works include Sutton's "Learning to predict by the methods of temporal differences" (1988) which introduced the TD(λ)-learning algorithm and the Baltimore Gas and Electric Company load data which could be useful to have for the project.

12. E. Kuznetsova, Y.-F. Li, C. Ruiz, E. Zio, G. Ault, and K. Bell, "Reinforcement learning for microgrid energy management," *Energy*, vol. 59, pp. 133 146, 2013

    Summary The authors in this work attempt to solve the optimal control of a microgrid using reinforcement learning. They model a consumer with load, renewable gen, and storage under market prices. They base the control strategy on Q-learning but extend the method to 2 time steps look-ahead. Additionally, they include a sensitivity analysis of learning parameters. A further contribution their approach has is that the model can account for uncertainty such as forecast error in renewable generation.

    Relevance The work of Kuznetsova et al. expands the ability of Q learning. This paper uses a more complicated model and calculations with multiple time steps that likely will not be incorporated in my final project. However, the paper provides further evidence of Q learning application and usefulness in the control of energy storage systems.

    Further references This paper included lots of references as they need to validate multiple models. There are many references from IEEE journals but also background information on electricity system and modeling wind power. The cited works are from a wide range of years but mostly after 2005. The most relevant papers are very similar applications and propose different methods for the energy management of microgrids such as Mohamed FA, Koivo HN. "System modelling and online optimal management of MicroGrid with battery storage" (2010) and Colson et al. "Towards real-time microgrid power management using computational intelligence methods" (2010).

13. L. Xiao, X. Xiao, C. Dai, M. Peng, L. Wang, and H. V. Poor, "Reinforcement learning-based energy trading for microgrids," *CoRR*, vol. abs/1801.06285, 2018.

    Summary The paper covers deep Q-learning for connected microgrids. The authors propose a stochastic energy trading game to determine the optimal energy trading schedule. Each microgrid can choose amount of energy to sell or purchase from other microgrids or from powerplants. They authors also provide the Nash equilibrium for the game which alludes to the performance bound of the scheme used. Finally, the paper preseents results from simulations and shows how the scheme improves performance over a benchmark strategy, "hotbooting Q".

Relevance The paper shows an extension of Q learning where netural network is used. It also expands on the decision of one entity to a network of entities. The exact algorithm will not be used but the work presents a good reference and potential data acquisition strategies.

Further references The majority of references are from IEEE Transactions on Smart Grid. They are mostly recent references cerca 2015. The most relevant sources include similar application of Q learning such as "Energy trading game for microgrids using reinforcement learning (2017), "Reinforcement learning-based control of residential energy storage systems for electric bill minimization" (2015), and "Dynamic pricing and energy consumption scheduling with reinforcement learning" (2016).

14. S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, p. 2010, 08 2018.

    Summary The paper presents an application of Q learning for smart energy building actions to minimize cost. The work uses a model which includes utility prices, vehicle to grid technology, photovoltaic generation and an energy storage system. A slight adaptation of the classical Q-learning scheme is used to reduce convergence time by initializing the Q-table. The authors then show simulation result which even outperforms hourly optimization.

    Relevance The paper is a very relevant application of Q learning. It is a clearly written paper and a great reference for defining actions, states, and rewards. While the setup and model will be slightly different it was the only other reference that is using 5 minute datasets. The adaptation to improve convergence and their data sources are particularly useful for the project.

    Further references The references are mostly IEEE transactions on Smart Grid journal papers. The works cited are mostly later than 2013, so are recent to the work published in 2019. The most relevant sources are data sources such as ISO New England and KEPCO as well as classic reinforcement learning references such as Sutton's "Reinforcement Learning: An Introduction 1998".

15. T. Orfanogianni and G. Gross, "A general formulation for lmp evaluation," *IEEE Transactions on Power Systems*, vol. 22, pp. 1163-1173, Aug 2007.

    Summary The goal of the paper is to give a general formulation for locational marginal prices. Locational marginal prices (LMPs) are defined as the least cost to supply an additional unit of load at specific location. The paper describes that LMPs are used in congestion management and hedging and their decomposition is needed for market settlement purposes and hedging. The formulation used breaks the LMPs into energy, loss, and congestion terms. The work then gives some examples of the decomposition on a six-bus network.

Relevance While not directly related to Q-learning or energy storage systems, this paper gives good insight into LMPs. The paper provides some intuition into what factors make up a LMP and why they are an important number to use as a price. This work gives some justification for using LMPs as real-time energy market prices and to define our rewards for the energy storage system.

Further references Most of the cited works in the bibliography are IEEE Transaction papers with some more focused on the fundamentals of mathematics and some on the background of energy systems. The references are a wide spread of publishing years starting all the way back in 1920. The most relevant references are others that give insight into LMPs such as ISO NE "Locational Marginal Pricing" and "Spot Pricing of Electricity" (1988).

3 Prior Work

I certify that to the best of my knowledge no prior University work is relevant to my proposed project for this course.