

UC BERKELEY



CE 263N - SCALABLE SPATIAL ANALYTICS

Electricity Price Prediction and Nodal Price Analysis in California ISO

Authors:

Mélanie MANGUIN

Emily PORTER

Bertrand TRAVACCA

Professor:

Alexey POZDNUKHOV

December 8, 2016

1 Introduction

In the United States, an independent system operator (ISO) controls and monitors the operation of the electrical power system within a given perimeter. In California, CAISO also acts as the electricity market operator. In this project we focus our attention on the day ahead market (DAM) for electricity. This market takes place one day prior to the operating day and consists of scheduling the quantities and marginal electricity prices (wholesale) for each hour. We will denote $p \in \mathbb{R}^{24}$ the DAM price for a given day.

The marginal clearing price corresponds to the intersection between the supply and demand curve: these curves are formed by CAISO who receive bids from all the market participants (essentially: electricity retailers, generators, and virtual bidders). Figure 1 gives an overview of the DAM market process, while Figure 2 illustrates the obtainment of the market clearing price for a given hour.

CAISO has the specificity of having different clearing prices based on location, this is done with the purpose of including externalities of electricity losses (around 6% on the transmission grid [6]), and electricity congestion in the final price [7]. All in all, the clearing price in the DAM is specific to a location (or node), and is referred to as Local Marginal Price (LMP).

It is essential for a market participant to be able to predict the DAM price before submitting it's bid to CAISO. That is why in a first step we decided to build a prediction model for the DAM price (average price in the system, not the LMP). It is also important for the market participant to understand the underlying risk when making a bid: that is why we studied the error between our prediction and the realized price.

In a second step, we examined and visualized the LMP behavior, which is particularly interesting if we want to understand the local market power an agent can have.

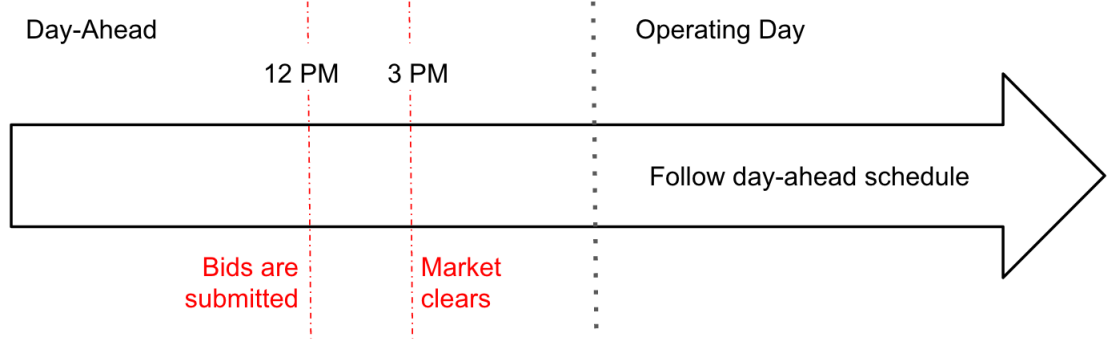


Figure 1: Day-Ahead Market overview.

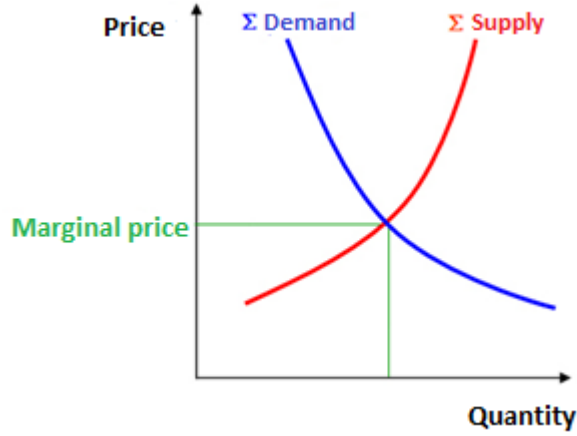


Figure 2: Clearing price: intersection between the demand and supply curves.

2 Price prediction, simulation, and risk management

We gathered 3 years of data (Jan 2013- Dec 2015) from CAISO (PG&E region only) to build a prediction model $\hat{p} \in \mathbb{R}^{24}$ and produce a covariance matrix estimate $\hat{\Sigma}_p$, such that $\mathbf{p} \sim \mathcal{N}(\hat{p}, \hat{\Sigma}_p)$. The data was taken from [4]. Note that this data corresponds to an average in the PG&E electric region

(this is not an LMP as in Section 3 or 4).

2.1 DAM price prediction

We started by producing a prediction model for DAM prices using random forest regression [2]. The input/features for predicting the price are:

- demand forecast (CAISO)
- year, month, day, and hour

Method 1: We started by randomly picking 80% of the days in the dataset to train our prediction model, and we then tested our model on the remaining 20%. We found a RMSE around 3.5\$.

Method 2: In order to build a more realistic model we implemented an 'online version': we started by training our model with the data of 2013, and then recursively updated our model and prediction as follow. Let us denote $\overline{p_d} = \{p_1, \dots, p_d\}$ the price history we have at a given day. Let us denote \mathcal{M}_d the random forest regression model corresponding to this historical data. In order to predict the next day price $\hat{p}_{d+1} \in \mathbb{R}^{24}$ we use the model \mathcal{M}_d we have. For the next day, we build an updated model \mathcal{M}_{d+1} using historical data $\overline{p_{d+1}}$, and we use this new model to predict the price \hat{p}_{d+2} , etc. With this model we obtained a RMSE of 4.5 to 5\$.

2.2 Risk management for a market participant (motivation)

Let us consider a market participant who wants to bid a quantity $Q \in \mathbb{R}^{24}$ in the DAM. This participant does not know yet what is the clearing price (cf. Introduction). If this agent wants to chose the optimal bid Q^* with respect to its portfolio, only considering minimizing its cost $\hat{p}^T Q := \sum_{h=1}^{24} \hat{p}_h^T Q_h$ would correspond to a risk neutral attitude.

Most of the time, an agent is risk averse regarding the bid he makes. The choice of the amount Q_h for $h \in \{1, \dots, 24\}$ can be considered as a classic portfolio problem where the assets are the DAM prices [1]. In a way, the risk averse agent does not want to "put all his eggs in the same basket" and looks

for a trade-off between the expected DAM price \hat{p} and its variance $Q^T \Sigma_p Q$, where $\Sigma_p \in \mathbb{R}^{24 \times 24}$ denotes the covariance of the error on prediction. **This digression motivates the following.**

2.3 Covariance matrix estimation and simulation using Cholesky decomposition

In a second step, we looked at the error ϵ between the predicted and realized price. $\epsilon_d \in \mathbb{R}^{24}$ refers to the error vector for the day d . We denote,

$$\delta = \frac{1}{N} \sum_{k=1}^{N_d} \epsilon_k \epsilon_k^T \in \mathbb{R}^{24 \times 24}$$

We assume that ϵ is a multivariate Gaussian random variable with zero mean (note: we empirically have an error ~ 0) and covariance matrix Σ_p . We want to estimate this covariance matrix using the samples we have. We consider each ϵ_d to be statistically independent. The density function for ϵ is given by:

$$f(x) = (2\pi)^{-12} \det(\Sigma_p)^{-1/2} \exp(-x^T \Sigma_p^{-1} x / 2)$$

The log-likelihood function has the form,

$$L = \log f(\Pi \epsilon_d) = -(12N_d) \log(2\pi) + (N_d/2) \log \det \Sigma_p^{-1} - (N_d/2) \text{trace}(\Sigma_p^{-1} \delta)$$

Note that L is a concave function of Σ_p^{-1} (often referred to as precision matrix). We can have an estimate of the precision matrix solving the following strictly convex optimization problem:

$$\hat{\Sigma}_p^{-1} := \arg \max_{\Theta \succ 0} \log \det(\Theta) - \text{trace}(\Theta \delta)$$

Or equivalently,

$$\hat{\Sigma}_p^{-1} := \arg \min_{\Theta \succ 0} \text{trace}(\Theta \delta) - \log \det(\Theta)$$

Adding a L_1 norm regularization term (lasso regularization) to this convex optimization problem allows to obtain a sparse precision matrix [5], which is (heuristically) useful if we want to have deeper insight on the error correlation between different hours and erase 'weak' correlations. The sparse covariance is obtained solving the following convex optimization problem:

$$\hat{\Sigma}_p^{-1} = \arg \min_{\Theta \succ 0} \text{trace} \Theta \delta - \log(\det(\Theta)) + \beta \|\Theta\|_1$$

Nevertheless, we found that β should be kept very small if we do not want to deteriorate the covariance matrix model. This tends to prove that, in our case, only using the empirical estimation might be reasonable. We used scikit learn [2] to obtain an estimate for the covariance matrix (we did this with the two prediction methods in 2.1). Figure 3 and 4 on page 6 display a heat-map of the covariance matrix estimate we have for the DAM price respectively with Method 1 and Method 2 ('online learning'). We can observe that the online learning method results are higher valued on the diagonal but also show high correlation for the error from one hour to the other. This is really different for Method 1 where the covariance matrix is mostly diagonal.

Figure 5 on page 6. shows some simulations of a priori possible DAM prices: $p_{simu} = Lu + \hat{p}$ with $u \sim \mathcal{N}(0, I)$ and L cholesky decomposition of the covariance matrix (from Method 2). We can see that the covariance matrix estimate is not able to capture all the uncertainty on the prediction we make: for instance we can observe that all the simulations tend to expect a higher price between midnight and 3AM. Figure 6 confirms this phenomenon, with simulations unable to capture high prices between 10AM and 5PM. This observation motivates section 2.4.

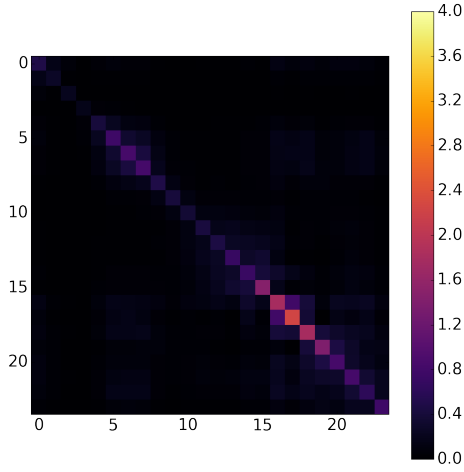


Figure 3: Heat map of the covariance matrix estimate $\hat{\Sigma}_p$ using method 1 of Section 2.1.

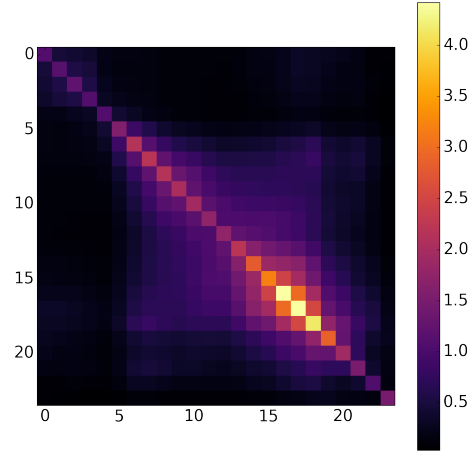


Figure 4: Heat map of the covariance matrix estimate $\hat{\Sigma}_p$ using method 2 of Section 2.1.

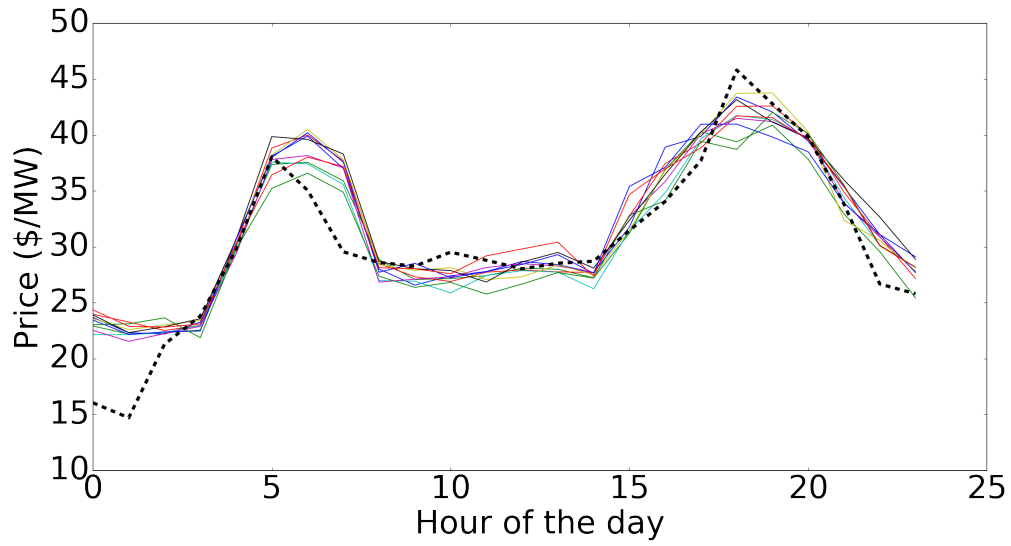


Figure 5: Day Ahead price simulations (colored lines) and real price (black dashed line).

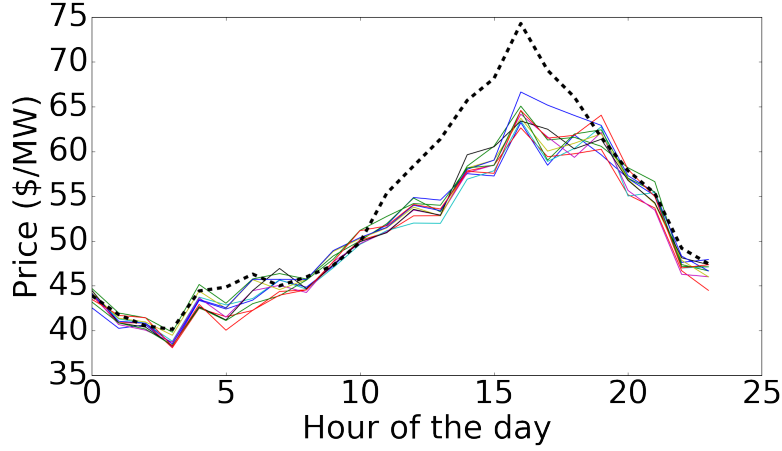


Figure 6: Day Ahead price simulations (colored lines) and real price (black dashed line).

2.4 Gaussian Mixture Model on the error

This last observation made us consider the fact that the error should be modeled as a Gaussian Mixture Model (GMM) rather than a simple Gaussian model (or a GMM with one component). Indeed, it is possible to imagine that the error can be clustered with respect to the day: for instance, for a given day our estimate might be below/above the true price all day. We used sklearn mixture library to fit a GMM to the error using a BIC criterion to define the number of components for the GMM. Figure 7. shows the BIC score for a different number of components. We found that the optimal number of components was one. This means that we cannot improve our model on the error with GMM.

This result is still interesting: if we go back to the digression in section 2.2, as the covariance matrix does not incorporate all the uncertainty (and we cannot find a model that would), it would be interesting for the risk averse agent who wants to optimize Q^* to include a linear matrix inequality constraint on the covariance matrix estimate to make it's choice more pertinent/robust (this would result in a SDP: Semi-Definite Program).

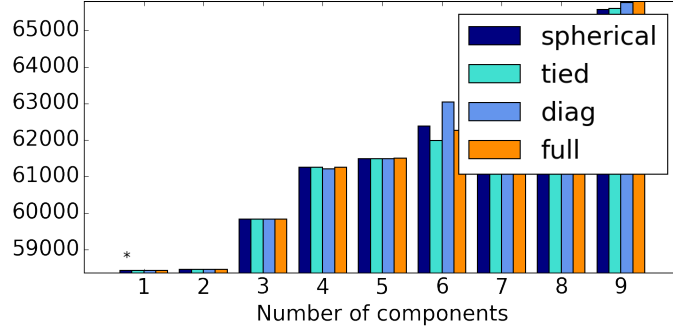


Figure 7: BIC score obtained for different number of components in the GMM for the error.

3 Nodal clustering

Data clustering is a useful tool to reduce the complexity of large datasets. It is helpful to group similar data so that the overall dataset can be thought of as a conglomeration of a few clusters, instead of several hundred / thousand / million individual data points. Although the concept of clustering is easily graspable, the actual methods used to cluster data vary widely based on the data being considered. Further, it is often difficult to know how many clusters should be used to ideally represent the data. In the case considered here, the data consists of hourly local marginal prices corresponding to nodes in the CAISO network. The data is grouped into days, so that each data point contains 24 numbers that represent the marginal price of electricity at that node over the course of the day. Figure 8 demonstrates this concept.

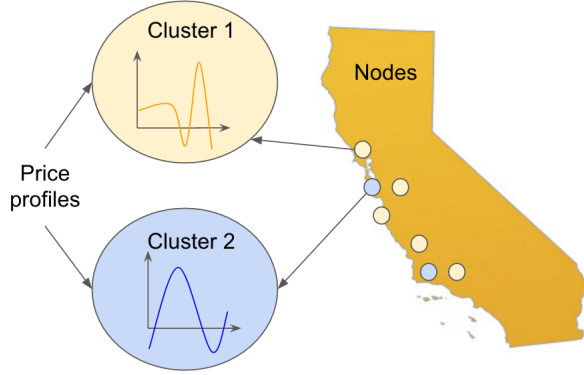


Figure 8: Visualization of data clustering.

For this project, two clustering methods were considered:

1. Density Based Spectral Clustering of Applications with Noise (DBSCAN)
2. K-Means (and Mini Batch K-means)

For both clustering methods, scikit implementations were utilized [9] [8]. The applications of these two clustering methods are described further below.

3.1 DBSCAN

The DBSCAN method has two necessary input parameters: 1) the maximum distance between two points that are in the same cluster, and 2) the minimum number of data points in a single cluster. These input parameters are difficult to optimize for several reasons. The notion of distance between the data points is not easily graspable because the distance is between hourly prices, not geographical distance. The minimum number of data points in a single sample is strongly related to the maximum distance between points. Furthermore, the DBSCAN method does not necessarily place all data samples into a cluster. Samples can be given a label of -1 meaning that they do not fall into any of the created clusters. Ultimately, if the maximum distance between samples in the same cluster is small, then the cluster itself will be composed of relatively few data points meaning there will be many clusters to represent the full data set. If the minimum number of samples in a cluster

is large then the maximum distance must be selected to allow so many data points to be in the same cluster, otherwise the majority of the data will be given the -1 label. Due to the difficulty of selecting these two parameters to work well with each other and accurately represent the data-set, the K-means clustering method was preferred.

3.2 K-means and Mini Batch K-means

K-means clustering has only one necessary input parameter: the number of desired clusters (mini batch k-means also requires the size of mini batches to be used). A brief literature review shows that there are several suggested performance metrics that can be used to determine the optimal number of clusters. These metrics include:

- The squared error between each point in a cluster and the cluster's center, defined below.

$$\sum_{i=1}^K \sum_{p \in C_i} (p - m_i)^2$$

- The entropy of the cluster, defined below [10].

$$- \sum_{i=1}^K P_n(C_i) \log(P_n(C_i))$$

Figure 9 shows the entropy and error calculated for different numbers of clusters using the mini batch k-means clustering method with a batch size of 50.

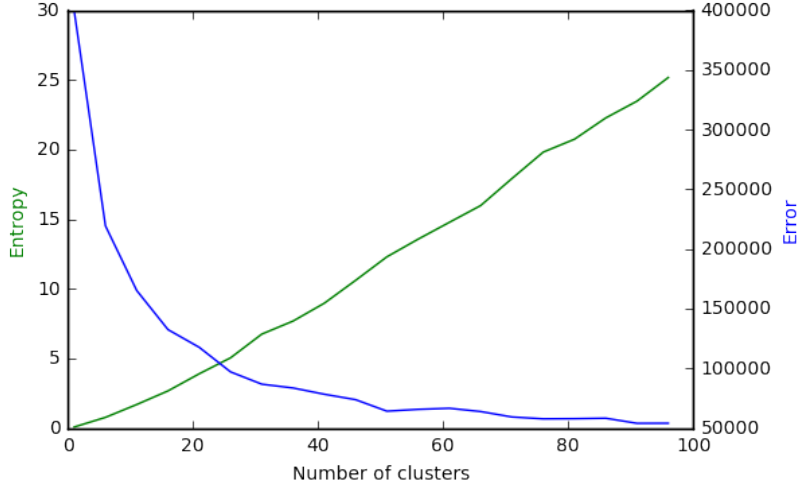


Figure 9: Variation in entropy and error with number of clusters.

As shown, the entropy increases as the number of clusters increases and the error decreases as the number of clusters increases. Both of these trends are expected. The entropy can be thought of as a measure of disorder, so as the number of clusters increases, the disorder/entropy increases. On the other hand, since the error is essentially measuring the distance between each point and its cluster center, this value decreases as the number of clusters increases because the clusters more closely represent the data within them. Ultimately, determining the best number of clusters to use is an unsolved problem. The ideal number of clusters is heavily dependent on the dataset. For the visualizations of the data presented in the following section, 10 clusters were chosen to represent the data. This number was used because it is large enough to capture some of the diversity in the data but small enough to be easily captured through visualization.

4 Spatial visualization of the nodes

In an electric power network, it is useful to examine local characteristics of the different nodes to understand macro tendencies. The marginal local price of the electricity (**LMP**) is the cost of supplying electric demand at a particular location (node). This LMP takes into account both supply bids and demand offers, as well as physical aspects of the transmission system.

Hence, it is composed of three components : (i) the marginal cost of energy (**MCE**), (ii) the marginal cost of congestion (**MCC**), and (iii) the marginal cost of losses (**MCL**). For instance, there may be some high-congestion areas that explain an increase in the average price, or we could detect that a whole portion of the transmission network is obsolete because of high losses in the nodes of this portion. Hence, we used visualization tools from Mapbox [11] to get better intuition on the CAISO network’s price patterns.

4.1 Local values

We started by scraping data from CAISO’s website for the whole month of November 2015 using [12]. We extracted the local marginal price (LMP) along with its three components MCC, MCE, MCL. We also extracted the geographical information associated with each node. A pre-cleaning of the data was necessary as there were some outlier values. In brief, steps we performed to get data in a readable format by Mapbox were the following:

- Clean outliers from dataset.
- Link price values to geographical information.
- Import this file as a GeoJSON to Mapbox and store it in a layer.
- Customize color chart by modifying the layer in Javascript.
- Display the map locally: we built 4 different maps, one with the LMP, and one for each component (MCC, MCE, MCL).

Below are some visualization maps and insights we could get from them (see Appendix for more visualizations):

4.1.1 Local marginal price (LMP)

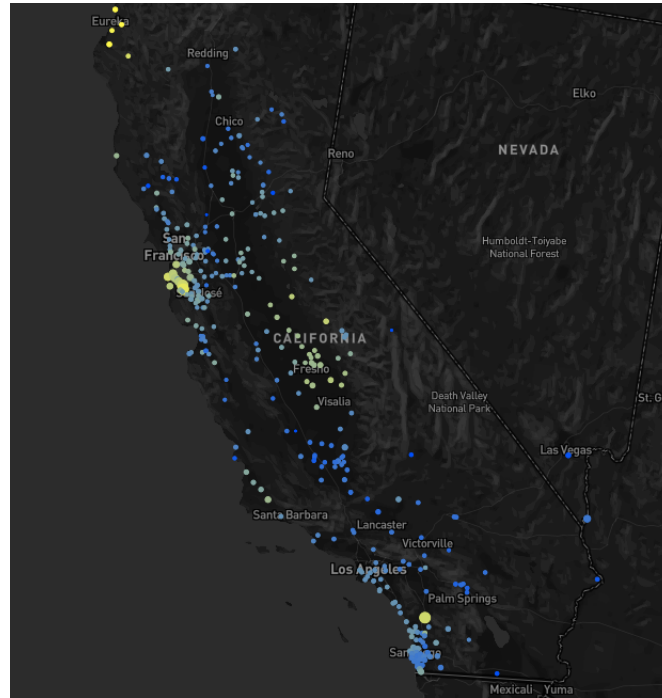


Figure 10: Local marginal price in California for November 2015.

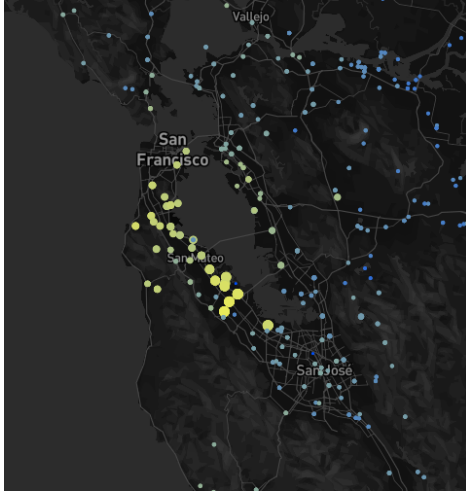


Figure 11: Zoom over San Francisco.

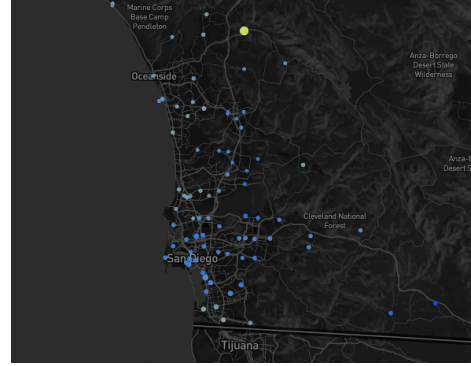


Figure 12: Zoom over San Diego.

The color represents the average price for November 2015, while the size of the dots represents its variance. The more yellow the higher the price is, and the bigger the dots, the more the price fluctuates. We can observe generally bigger dots around big cities like San Francisco, Los Angeles, and San Diego, which indicates the price varies more there. This may be due to higher demand peaks in cities. Also, it can be noted that prices are higher in San Francisco than in San Diego, which can be explained by higher demand for heat in Northern California.

4.1.2 Marginal Cost of Congestion (MCC)

The MCC designates the marginal cost of transmission system congestion due to binding constraints, e.g., binding transmission line constraints. Below is one visualization of the MCC component of LMP:

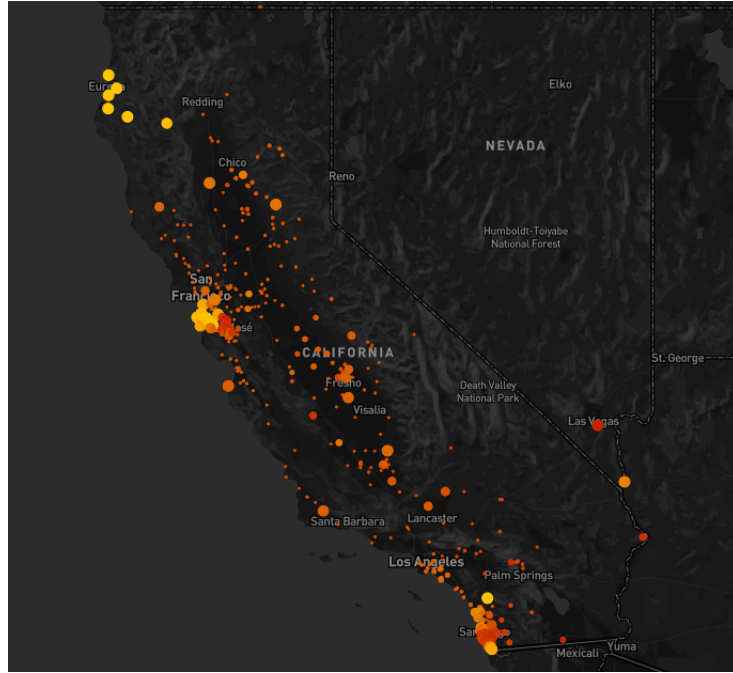


Figure 13: Marginal cost of congestion in California for November 2015.

As before, a color gradient was used to represent the average cost of congestion (the more red the higher the cost), and the size of the dots indicates the variance of this cost. A first takeaway is that this cost fluctuates more in urban areas. However, it is worth noticing that this cost stays lower than average in San Francisco (see Appendix for a zoomed in view).

4.1.3 Marginal Cost of Losses (MCL)

The MCL stems from thermal losses in the transmission lines. Below is one visualization of the MCL component of LMP over California:

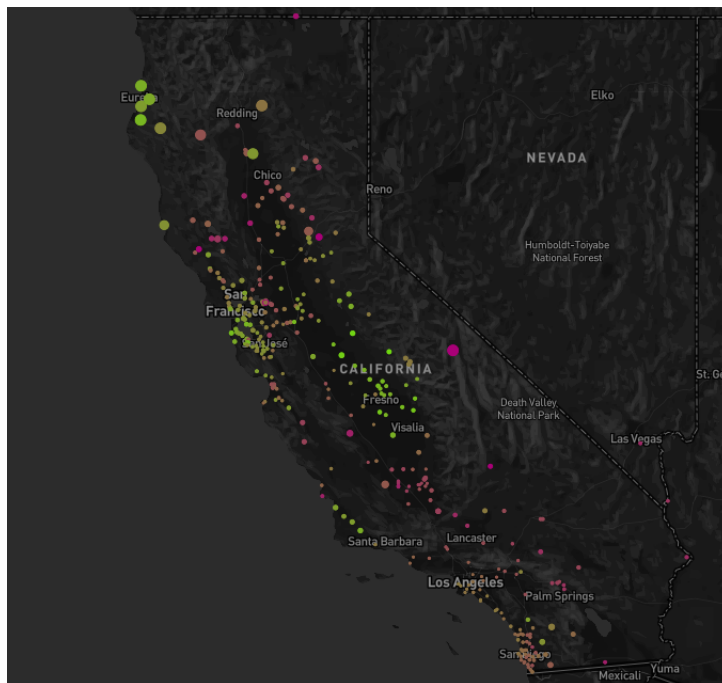


Figure 14: Marginal cost of losses in California for November 2015.

The color gradient here is such that the greener the dot is, the lower the loss is. This map shows that the losses are very low in the Bay Area, and seem to get higher inland, especially in farming areas between San Francisco and Los Angeles.

4.2 Visualization of K-Means clusters

We also made a web map of the clustering done with K-Means (only made on the LMP, not on marginal components). The web map is available at [this link](#), and each color maps to one particular cluster, with 10 clusters total.

5 Conclusion and future works

As explained above, the day-ahead market for electricity is well suited to prediction models, such as random forest regression. Accurately predicting the prices of electricity at specific nodes is of high importance to participants in the electricity market. Further, the clustering and visualization of

the electricity price data reveal interesting information such as the higher variance of electricity prices in urban areas, and the lower losses in the SF Bay area compared with places further inland. Further analysis regarding the best number of clusters to represent the data, using metrics other than the error and entropy, may be useful. Additionally, efforts to improve the predictive capabilities of the regression model would be interesting.

References

- [1] Boyd, Stephen, and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.
- [2] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of Machine Learning Research 12.Oct (2011): 2825-2830.
- [3] CAISO. Interface Specification for OASIS.
- [4] CAISO: Average Price - LCG Consulting :: EnergyOnline
- [5] Mazumder, Rahul, and Trevor Hastie. "The graphical lasso: New insights and alternatives." Electronic journal of statistics 6 (2012): 2125.
- [6] Wikipedia contributors. "Electric power transmission." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 9 Dec. 2016. Web. 9 Dec. 2016.
- [7] Singh, Harry, Shangyou Hao, and Alex Papalexopoulos. "Transmission congestion management in competitive electricity markets." IEEE Transactions on power systems 13.2 (1998): 672-680.
- [8] "k-means++: The advantages of careful seeding" Arthur, David, and Sergei Vassilvitskii, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics (2007)
- [9] "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" Ester, M., H. P. Kriegel, J. Sander, and X. Xu, In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, AAAI Press, pp. 226–231. 1996

- [10] J. Kwac, J. Flora, R. Rajagopal, Household energy consumption segmentation using hourly data, *IEEE Transactions on Smart Grid* 5 (1) (2014) 420–430
- [11] www.mapbox.com, 2016
- [12] Munsing, Eric, CAISO-Scrapers, (2014), GitHub repository, <https://github.com/emunsing/CAISO-Scrapers/>

6 Appendix

6.1 Zoom on particular areas for Marginal Cost of Congestion (MCC)

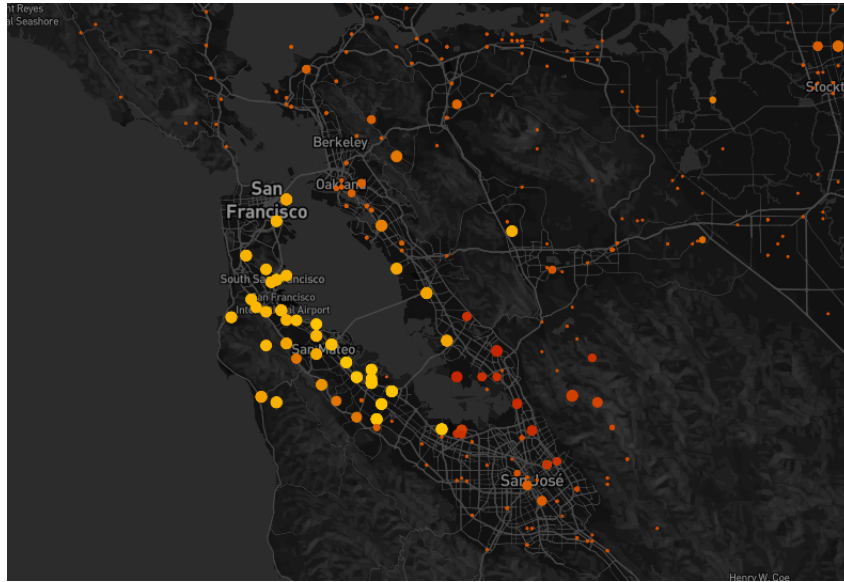


Figure 15: Marginal cost of losses in the Bay Area for November 2015.

6.2 Zoom on particular areas for Marginal Cost of Losses (MCL)

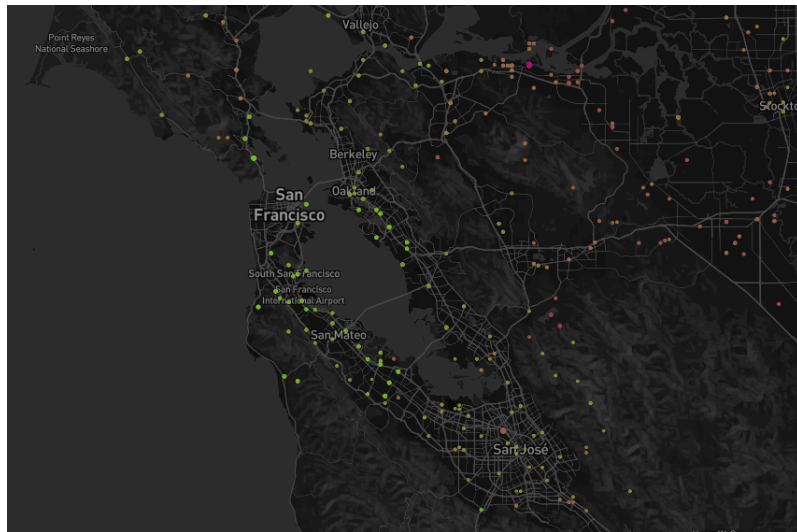


Figure 16: Zoom of MCL over San Francisco.

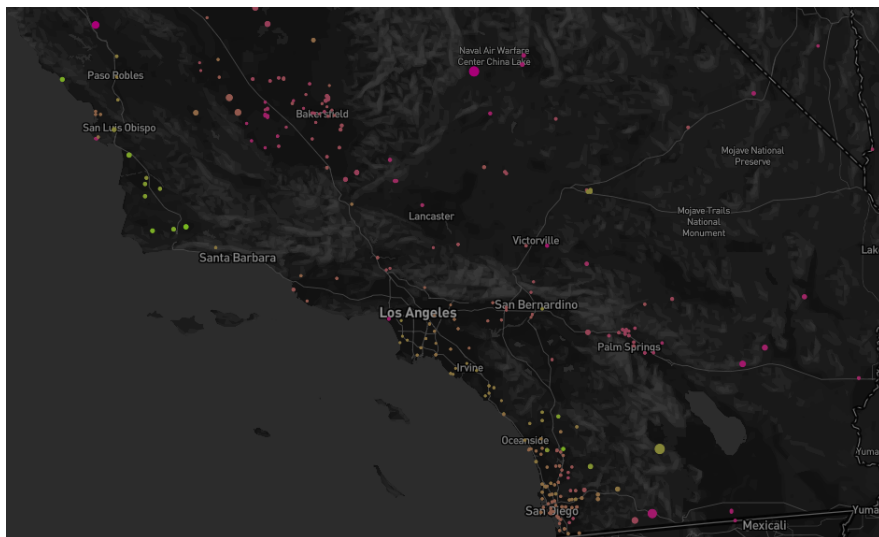


Figure 17: Zoom over South California.