

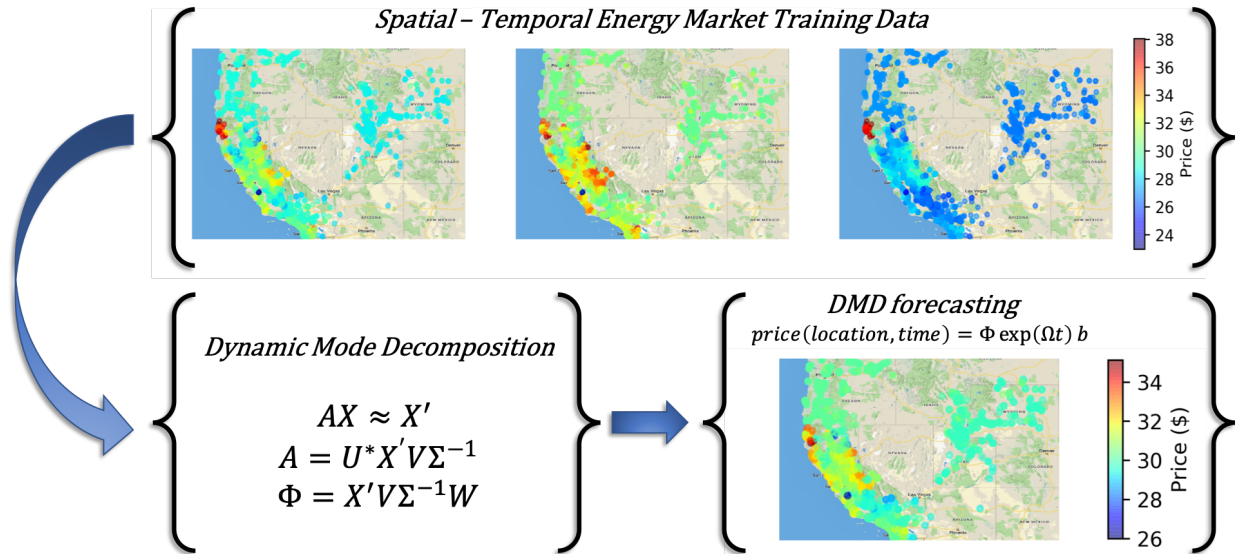
Learning Spatiotemporal Dynamics in Wholesale Energy Markets with Dynamic Mode Decomposition

Clay T. Elmore and Alexander W. Dowling

October 1, 2020

Abstract

Energy markets facilitate the balancing of electricity generation (supply) and demand while ensuring non-discriminatory access. Understanding energy market dynamics is essential to improving grid efficiency and resilience and optimizing the development of new energy conversion and storage technologies. Accurate energy price forecasts are essential for many energy storage technologies to be profitable from price arbitrage. In this paper, we apply the novel spatial-temporal dimensionality reduction method of Dynamic Mode Decomposition (DMD) to forecast 6587 locational marginal prices in the California Independent System Operator (CAISO) on the Day-Ahead Market (DAM). DMD is a promising equation-free modeling technique in systems with inherent periodic tendencies in time such as financial markets and fluid dynamics. Yet we show, for the first time, that DMD cannot reliably forecast day-ahead energy prices due to the so-called standing wave problem. Instead, we find Augmented DMD (ADMD) overcomes these limitations is a fast and accurate price forecaster. We benchmark DMD, ADMD, and backcasting forecasting methods for optimal price arbitrage with an energy storage system. We find, using ADMD, a market-connected energy storage system can capture up to 92% of allowable revenues in rolling horizon simulations. Lastly, we show ADMD is up to 1000-times faster than time-series forecasting methods (i.e., ARIMA) which requires orders of magnitude less data than deep/machine learning techniques.



1. Introduction

Wholesale energy markets are ubiquitous across the globe and provide a competitive mechanism to balance electrical energy supply and demand and ensure reliable operation of large-scale power systems.[41] Energy prices encapsulate complex trade-offs between costs and flexibility of diverse generation technologies (e.g., coal, nuclear, natural gas, solar, wind, hydro), demand elasticity (e.g., demand response), and operational limits of high-voltage transmission systems. These constraints interact with uncertainties from demand, renewable availability, unplanned equipment outages to create strong spatial-temporal variations in energy prices.[12, 40] Figure 1 illustrates stochastic and periodic variations in wholesale prices for six nodes in the day-ahead market run by the California Independent System Operator (CAISO). In this paper, we analyze hourly time-series data obtained from <http://oasis.caiso.com/> for all 6587 nodes (locations) during the calendar year 2015. Descriptive statistics for these data are reported in [12, 13].

The economics of modern electricity generation and consumption is driven by these price variations. For example, the economic case for energy storage systems is based, in part, on energy arbitrage. Through “buy low, sell high” operating policies, storage systems can monetize time fluctuations in wholesale energy prices. In their seminal work, Walawalkar et al. [48] use mathematical optimization to estimate the maximum possible revenue from historical market data and compare energy storage technology options. This now-standard approach has been applied by many technologies including batteries[12, 43, 16, 30, 44], and concentrated solar thermal with energy storage[13, 25, 22, 27, 14]. Likewise, time-varying energy prices incentivize large energy consumers to either directly participate in wholesale markets or demand response programs.[1, 34]. Early work focused on optimizing the schedule of large energy consumers such as flour mills, air separation, and combined heat and power plants subject to time-varying electricity prices.[mitra2013optimal, 4, 21, 5, 29] Recent work has focused on providing ancillary services with aluminum smelters, air separation plants, heat ventilation and air conditioning systems, and chemical plants.[dowling2018economic, 50, 49, 32] Yet many of these approaches assume market forecasts are available.

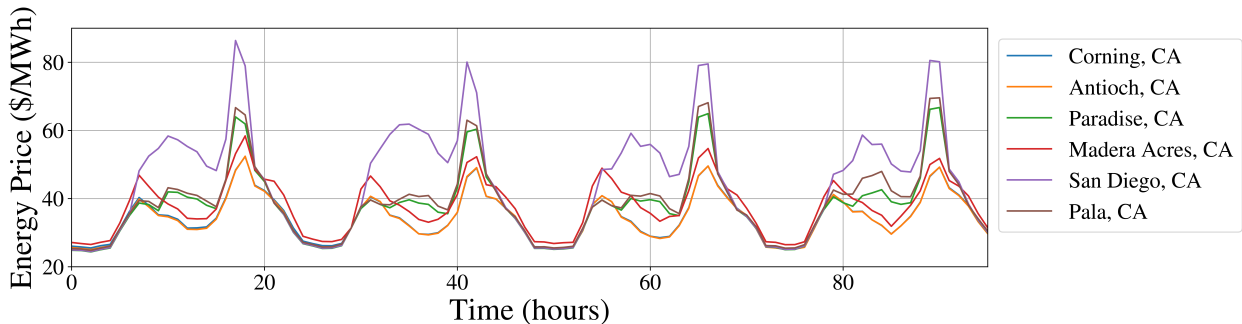


Figure 1: Day-ahead market (DAM) electricity prices for six nodes (locations) in the California from January 25th 28th, 2015. DAM prices exhibit strong daily trends typified with valleys (minimums) at midday and at night and peaks (maximums) in the morning and evening.

Both time-series and machine learning techniques have been used to extract trends from massive market datasets and make accurate forecasts. For example, Conejo et al. demonstrate autoregressive integrated moving average (ARIMA) models can predict 2002 Spanish DAM market prices with 10% mean squared error (MSE).[9] Likewise, Garcia et al. used generalized autoregressive conditional heteroskedasticity (GARCH) methods to predict Spanish and California DAM market prices from 2000 with 10% MSE.[18] These classical time-series models require tedious hypertuning (e.g., feature selection, residual analysis), especially to capture periodic trends. As such, recent work focuses on applying machine learning algorithms for DAM forecasting [2, 17, 45, 6, 33, 35, 39, 3, 35, 26]. For example, Lora et al. demonstrated 8% MSE for forecasting 2001 Spanish DAM prices using a weighted nearest neighbor approach. Similarly, Catalão et al. used a feed forward neural network to forecast 2002 Spanish and 2000 California day-ahead market prices, achieving 8% and 3% MSE, respectively. They find that the feedforward neural network is less computationally intensive than an ARIMA forecast. However, neural network methods are known to struggle with noise and recapitulating multiscale dynamics[39, 46]. To address these issues, Shafie-khah et al. combined wavelet transformations, ARIMA,

and neural networks to forecast DAM prices in Spain from 2002 [39] and demonstrated 5% MSE. Similarly, Tan et al. combined ARIMA, GARCH, and wavelet transformations to achieve near 1% MSE in forecasts for the Spanish and PJM DAM [46]. Wavelet transformations add robustness to noise and provide a systematic framework to capture diverse timescale [46, 3, 8]. All of these frameworks build individual forecasting models for each node in the electricity market, which can cause high computational times, necessitate model tuning at each node, and offers no fundamental insights about the aggregated dynamics of energy markets.

In this paper, we propose the first application of Dynamic Mode Decomposition (DMD) to forecasting day-ahead energy market prices. DMD is a so-called equation-free dimensionality reduction technique to extract complex, nonlinear, periodic behavior from experimental and computational simulations. [37, 38, 31, 47, 24, 38, 36] DMD is best-known for its ability to discover low-dimension spatial and temporal insights from computational fluid dynamics, although emerging applications include video processing and financial (stock) markets [24]. For example, Mann and Kutz [28] proposed a stock trading algorithm build on spatiotemporal information encoded in DMD modes and demonstrate 5 times higher growth than the S&P 500 index via retrospective analysis. Likewise, Cui and Long [10] showed how DMD-based trading beat a benchmark moving-average strategy in a bull market for Chinese A-stocks. Hua et al. [20] identified four new reproducible modes (cycles) by analyzing stock prices between 1992 and 2014 for major companies. All of these works underscore that DMD can successfully capture the underlying nonlinear dynamics in stock markets from data alone. In this paper, we show that Augmented DMD is an accurate short-term forecaster energy prices in wholesale electricity markets. ADMD requires only a modest amount of training data, much less than machine learning methods, and is orders of magnitude faster than ARIMA methods.

The paper is organized as follows: Section 2 provides a mathematical primer on DMD and Augmented DMD (ADMD). Section 3 explores singular value decomposition (SVD) rank reduction, interprets dominant DMD modes, shows how DMD suffers from the so-called standing wave problem. Section 4 demonstrates the superior performance of ADMD. Section 5 demonstrates how ADMD forecasting can capture up to 92% available revenue in a simple energy arbitrage optimize case study. Finally, Section 7 summarizes conclusions and future work.

2. Mathematical Primer on DMD

In this section, we review the mathematical foundations of DMD in the context of energy market forecasts. See Kutz et al. [24] and Schmid et al. [38] for additional details.

2.1. DMD Algorithm

DMD starts with an input matrix $\mathbf{X}_t \in \mathbb{R}^{m \times n}$ where the rows (m) of \mathbf{X}_t are the spatial data features and the columns (n) are discrete time measurements. For day-ahead energy market analysis, each row is a node (location in the transmission network), each column is an hour, and each element is a price. The price matrix \mathbf{X}_t is then partitioned into two matrices $\mathbf{X} \in \mathbb{R}^{m \times (n-1)}$ and $\mathbf{X}' \in \mathbb{R}^{m \times (n-1)}$, as shown in Eqs. (2) and (3). Here each $\mathbf{x}_i \in \mathbb{R}^{m \times 1}$ is a single time step of spatial data in price matrix \mathbf{X}_t .

$$\mathbf{X}_t = [\mathbf{x}_1 \cdots \mathbf{x}_n] \quad (1)$$

$$\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_{n-1}] \quad (2)$$

$$\mathbf{X}' = [\mathbf{x}_2 \cdots \mathbf{x}_n] \quad (3)$$

The overall goal of DMD is to compute a best fit linear operator, $\mathbf{A} \in \mathbb{R}^{m \times m}$ in Eq. (4), which advances the energy price dynamics by one timestep (one hour).

$$\mathbf{A}\mathbf{X} \approx \mathbf{X}' \quad (4)$$

\mathbf{A} is a finite dimensional approximation to the *infinite dimensional* Koopman operator, \mathcal{K} , which exactly maps any measurement function f from one state vector, \mathbf{x}_i , to another state vector, \mathbf{x}_{i+1} :

$$\mathcal{H}f(\mathbf{x}_i) = f(\mathbf{x}_{i+1}) \quad (5)$$

For practical analysis, we desire a low rank approximation to the linear operator \mathbf{A} . This can be accomplished via singular value decomposition (SVD) of the data matrix $\mathbf{X} \in \mathbb{R}^{m \times (n-1)}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \quad (6)$$

Here $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{(n-1) \times (n-1)}$ are unitary matrices and $*$ is the conjugate transpose. Matrix $\mathbf{\Sigma} \in \mathbb{R}^{m \times (n-1)}$ has singular values σ on the diagonal ordered $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{n-1}$. When $m > n - 1$, which is often the case in DMD analysis, the bottom of $\mathbf{\Sigma}$ are rows of zeros:

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m], \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{n-1} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \quad \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}] \quad (7)$$

Retaining only the r largest singular values and vectors gives a rank r approximation to \mathbf{X} . The choice of r is an important tuning parameter in DMD workflows (see Section 3). The maximum number of singular values is $r_{max} = \min(m, n - 1)$.

$$\mathbf{X} \approx \mathbf{X}_r = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^* = \sum_{i=1}^r \mathbf{u}_i \sigma_i \mathbf{v}_i^* \quad (8)$$

$$\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r], \quad \mathbf{\Sigma}_r = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix}, \quad \mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$$

Next, we calculate \mathbf{A} using Eq. (4) and the pseudoinverse \mathbf{X}^{-1} . Substituting Eqs. (6) and (8) gives:

$$\mathbf{A} \approx \mathbf{X}'\mathbf{X}^{-1} = \mathbf{X}'(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^*)^{-1} = \mathbf{X}'\mathbf{V}^*\mathbf{\Sigma}^{-1}\mathbf{U} \approx \mathbf{X}'\mathbf{V}_r\mathbf{\Sigma}_r^{-1}\mathbf{U}_r \quad (9)$$

Often it is more convenient to use the low-rank projection $\tilde{\mathbf{A}} \in \mathbb{R}^{r \times r}$:

$$\tilde{\mathbf{A}} = \mathbf{U}_r^* \mathbf{A} \mathbf{U}_r = \mathbf{U}_r^* \mathbf{X}' \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \quad (10)$$

which advances the low-dimensional projected states, $\tilde{\mathbf{x}} \in \mathbb{R}^{r \times 1}$, from timestep k to $k + 1$:

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\mathbf{A}} \tilde{\mathbf{x}}_k \quad (11)$$

High-dimensional states are reconstructed via projection with $\mathbf{U}_r \in \mathbb{R}^{m \times r}$:

$$\mathbf{x}_k = \mathbf{U}_r \tilde{\mathbf{x}}_k \quad (12)$$

Because $\tilde{\mathbf{A}}$ is a linear operator, the price trajectories can be calculated analytically using an eigendecomposition:

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda} \quad (13)$$

Here the columns of $\mathbf{W} \in \mathbb{C}^{r \times r}$ are eigenvectors and the diagonal of $\mathbf{\Lambda}$ are the corresponding eigenvalues. The spatial modes, $\mathbf{\Phi} = [\phi_1 \dots \phi_r] \in \mathbb{C}^{m \times r}$, are then calculated via Eq. (14), and represent how prices depend on geographic location in the electric transmission network. Next, the discrete time eigenvalues, $\mathbf{\Lambda}$, are converted to continuous time eigenvalues, $\mathbf{\Omega} \in \mathbb{C}^{r \times r}$, via Eq. (15). Here i and j are row and column indices, respectively, such that $i = j$ are the diagonals of $\mathbf{\Lambda}$ and $\mathbf{\Omega}$. Thus $\mathbf{\Omega}$ encodes price time dynamics that are independent of spatial behavior.

$$\Phi = \mathbf{X}' \mathbf{V}_r \Sigma_r^{-1} \mathbf{W} \quad (14)$$

$$\Omega_{ij} = \begin{cases} \frac{\ln(\Lambda_{ij})}{\Delta t} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

The low-dimensional time dynamics, $\mathbf{D}(t) \in \mathbb{C}^{r \times 1}$, are defined in Eq. (16). The vector $\mathbf{b} \in \mathbb{C}^{r \times 1}$ is the initial condition of the low-dimensional system, which is obtained by solving Eq. (17), often via least-squares.

$$\mathbf{D}(t) = e^{\Omega t} \mathbf{b} \quad (16)$$

$$\Phi \mathbf{b} = \mathbf{x}_1 \quad (17)$$

This concludes the decomposition of the original price data. A key advantage of DMD is that forecasting for any time and every node (location) can be quickly computed via Eq. (18).

$$\mathbf{x}(t) = \Phi \mathbf{D}(t) \quad (18)$$

In DMD literature, *reconstruction* of the original data refers to evaluating Eq. (18) for all of the times in the original original \mathbf{X}_t input matrix. Recall that DMD assumes a low-dimensional linear system governs the data. When this assumption is violated, the reconstructed prices can differ dramatically from the actual (training) prices. We further discuss the relationship between *reconstruction error* and *forecasting error* in Section 3.

2.2. Augmented DMD Algorithm

A well-documented challenge for standard DMD is modeling simple dynamics of standing waves, e.g., $x(t) = \sin(t)$ [24]. The fundamental problem for this example is that the input matrix \mathbf{X} is rank one, and as such DMD identifies a single mode with a positive real eigenvalue. But a sine wave is the solution to the differential equation $\ddot{x} + x = 0$; a conjugate pair of complex eigenvalues are needed to correctly recapitulate a sine wave. For typical days, the energy price closely resembles a simple oscillating wave (Fig. 1), which offers one explanation for the mediocre reconstruction and forecasting errors with standard DMD.

Augmented DMD (ADMD) overcomes the so-called standing wave problem by stacking time-shifted copies of input data to form the *augmented input matrix* which has higher rank than the original input matrix. This is analogous to applying a finite difference rules to estimate unobserved derivatives needed to model higher order systems (e.g., sine wave) and differencing data to achieve stationary for ARIMA time-series modeling [7]. To implement ADMD, we stack $l < n$ time-shifted copies of the data to form the augmented input matrix. The augmented matrix is then partitioned into two matrices, similar to Eqs. (2) and (3):

$$\mathbf{X}_{\text{aug}} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{n-l} \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_{n-l+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_l & \mathbf{x}_{l+1} & \cdots & \mathbf{x}_{n-1} \end{bmatrix}, \quad \mathbf{X}'_{\text{aug}} = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_{n-l+1} \\ \mathbf{x}_3 & \mathbf{x}_4 & \cdots & \mathbf{x}_{n-l+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{l+1} & \mathbf{x}_{l+2} & \cdots & \mathbf{x}_n \end{bmatrix} \quad (19)$$

The DMD algorithm prescribed in Eqs. (2) - (18) is applied to augmented matrices $\mathbf{X}_{\text{aug}} \in \mathbb{R}^{(m \cdot l) \times (n-l)}$ and $\mathbf{X}'_{\text{aug}} \in \mathbb{R}^{(m \cdot l) \times (n-l)}$ in place of \mathbf{X} and \mathbf{X}' , giving eigenvalues Λ_{aug} and modes Φ_{aug} . The first m rows of Φ_{aug} correspond to the current (not shifted) time and are used to forecast $\mathbf{x}(t)$. For ADMD, the maximum number of singular values is $r_{\text{max}} = \min(m \times l, n - l)$

3. DMD Results and Discussion

We now apply DMD to CAISO energy price data. Our analysis focuses on truncation level, window size, consistency, forecasting, interpretation of DMD modes. Throughout this paper, we measure percent error using the ℓ_2 norm:

$$\% \text{ error} = \frac{\|\mathbf{X}_t - \mathbf{X}_{\text{DMD}}\|_2}{\|\mathbf{X}_t\|_2} \times 100 \quad (20)$$

where \mathbf{X}_{DMD} are either reconstructions or forecasts calculated with Eq. (18).

3.1. Optimal Singular Value Truncation to Minimize Reconstruction Error

An essential tuning parameter in DMD analysis is the level of singular value truncation, especially for noisy data[24]. We begin by systematically exploring the impact r , the number of singular values retained in Eqs. (8) - (10) on reconstruction and forecasting error for DAM energy price data. It is well-known that retaining too many singular values, especially for noisy datasets, can corrupt DMD results.[24] Fig. 2 shows this trade-off holds for DAM price data for all nodes for three days ($\mathbf{X}_t \in \mathbb{R}^{m=6587 \times n=96}$). Too much truncation leads to decaying reconstruction with damped dynamics, whereas not enough truncation leads to an exponentially growing reconstruction (diverges to ∞) with exaggerated dynamics. The optimal truncation level was determined via sensitivity analysis and corresponds with the minimum reconstruction error of 12%, computed via Eq. (20). We reiterate these findings are consistent with DMD literature for noisy datasets.

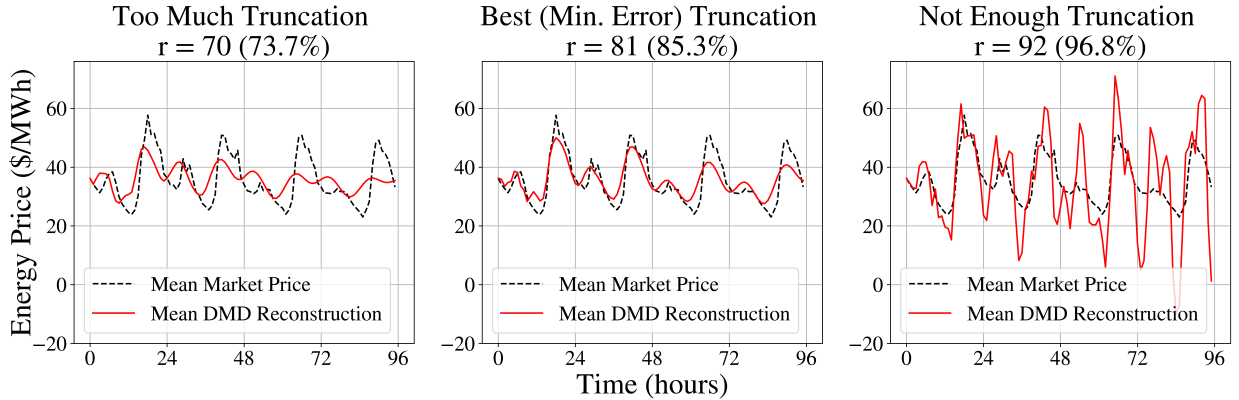


Figure 2: Three DMD reconstructions for DAM energy prices from all nodes for January 1st to 4th, 2015 such that $\mathbf{X}_t \in \mathbb{R}^{m=6587 \times n=96}$. The mean price for all nodes is shown in black (dashed) and the mean DMD reconstruction is shown in red (solid). Too much truncation, $r = 70$ ($70/95 = 73.7\%$), causes the DMD reconstruction to decay to constant value as $t \rightarrow \infty$. Too little truncation, $r = 91$ ($91/95 = 96.8\%$) causes the DMD reconstruction to diverge as $t \rightarrow \infty$. But with the best truncation, $r = 82$ ($82/95 = 82.3\%$), the DMD reconstructions recapitulates the dominant dynamics of the energy price signals.

We now explore how the optimal singular value truncation level, r , changes for different datasets. Fig. 3 shows the sensitivity analysis of reconstruction error versus r for three four-day windows considering all nodes ($X_t \in \mathbb{R}^{m=6587 \times n=96}$). For all three datasets, we clearly see the trade-off between not enough and too much truncation. Moreover, the optimal truncation level (lowest reconstruction error averaged over all nodes) varies between $r = 81$ and $r = 91$, which is $81/96 = 84\%$ and $91/96 = 95\%$ of the full rank number of singular values. This means that standard DMD fails to extract coherent low-rank structures from the DAM energy price data. Moreover, *each individual training window requires a different r to recover the optimal reconstruction*. We will revisit these two observations in Section 4 when we discuss ADMD.

Next, we quantify how DMD performance varies for individual nodes. Fig. 4 shows the distribution of reconstruction errors across all nodes for three 4-day training windows. These errors are approximately normally distributed for the January 1st - 4th dataset, but skewed-left for the January 11th - 14th dataset and bimodal for January 6th - 9th dataset. All three datasets show statistical outliers, which are likely due to random phenomena in the market clearing (price setting) process. Next, we compute average reconstruction errors for individual nodes using the optimal r (determined via sensitivity analysis) for all eight-day training windows ($X_t \in \mathbb{R}^{m=6587 \times n=192}$) spanning January 1st through 30th. Fig. 5 visualizes means reconstruction errors in a histogram (all nodes) and overlaid on a map (≈ 2000 nodes with known latitude and longitude coordinates). We find 75% of nodes have an average reconstruction error between 12% and 13% for January 2015. Reconstruction errors follow a right-skewed distribution, with 1.9% of nodes with 20% to 30% error. From these results, we six geographically separated nodes in California such that nodes 1 and 2 have *great* (less than 11%), nodes 3 and 4 have *average* ($\approx 12\%$), and nodes 5 and 6 have *poor* (greater than 16%) reconstruction errors. Table 1 highlights official CAISO node names, location, and mean reconstruction

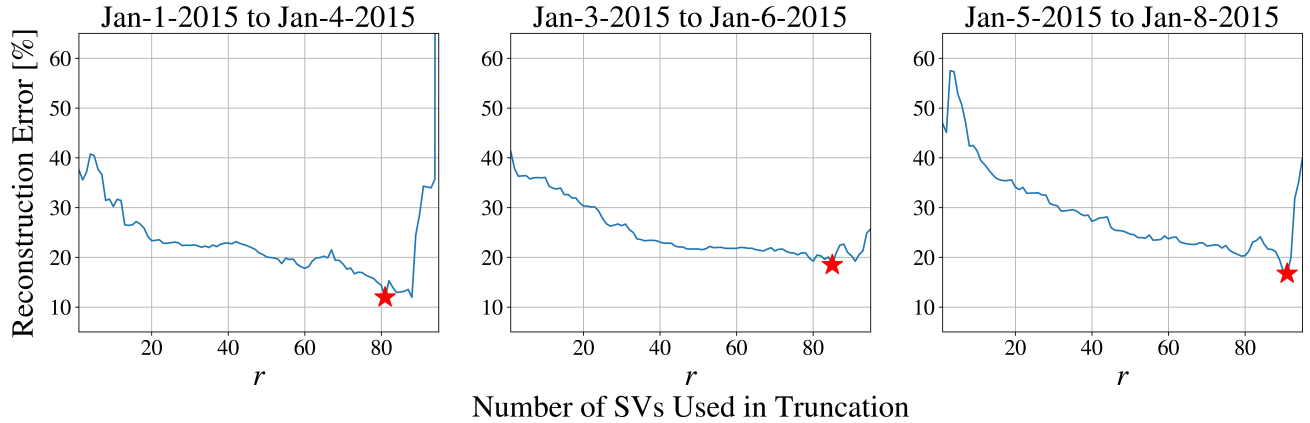


Figure 3: Comparison of reconstruction error aggregated over all 6587 nodes (%) versus the number of retained singular values (r) for three 4-day training windows. The \star symbol marks the minimum reconstruction error: (left) $r = 81$ corresponds with 11.95% error; (middle) $r = 85$ gives 18.46% error; (right) $r = 91$ gives 16.72% error.

error. For these six nodes, we reproduce the sensitivity analysis from Fig. 3. The reconstruction error versus r plots are only 1% different from those shown in Fig. 3. This suggests that the optimal r for the entire dataset (all nodes) is near optimal for individual nodes. Moreover, there are no obvious geographic trends in the distribution of reconstruction errors in Fig. 5 (middle).

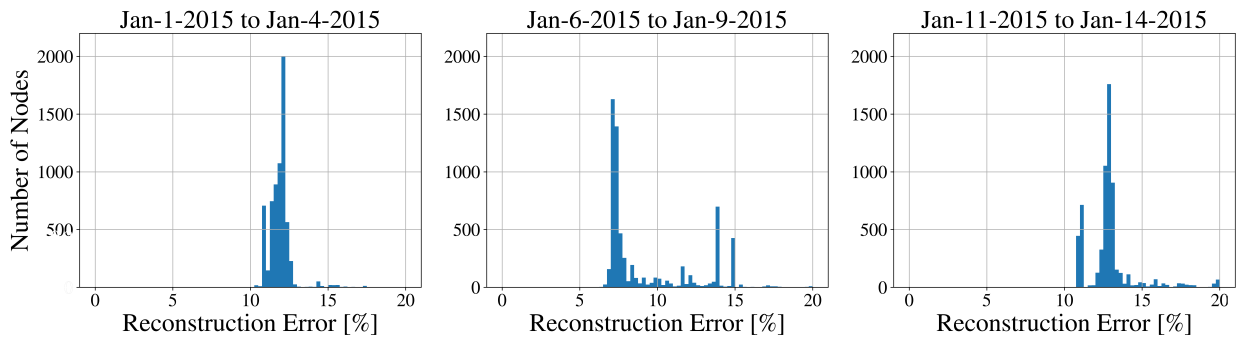


Figure 4: Histogram of reconstruction errors tabulated per node for three training windows using $r = 81$ (left), 91 (middle), and 92 (right), $m = 6587$, and $n = 96$. While each window has a distinct distribution of errors, a majority of errors are between 5% and 15%. Each window shows a few nodes, which are outliers, with reconstruction errors above 15%.

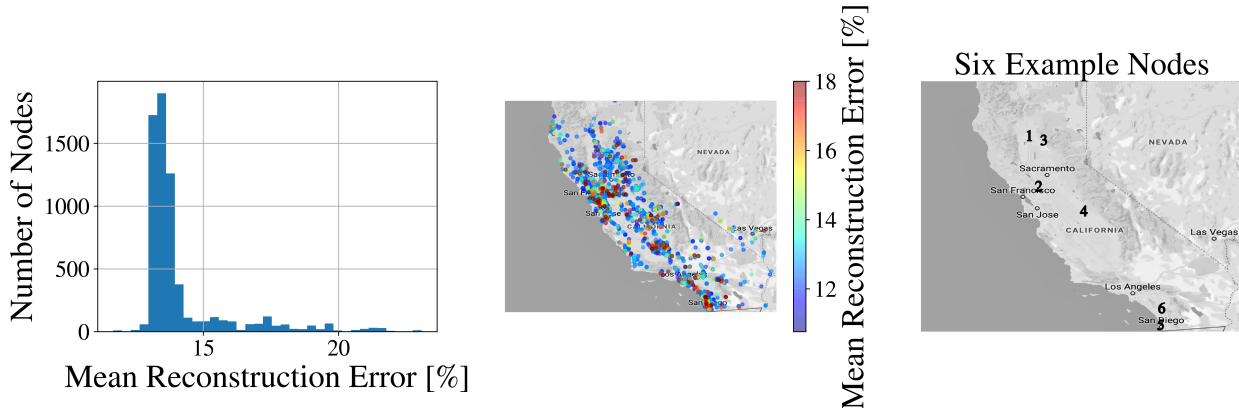


Figure 5: (Left) Distribution of the mean reconstruction error over calendar year 2015 tabulate for each node using $m = 6587$, $n = 96$, and the optimal r for each window. (Center) Spatial distribution of mean reconstruction errors for 2220 nodes in CAISO GPS coordinates publicly available on oasis.caiso.com. (Right) Location for six example nodes detailed in Table 1.

Node Name	Mean % Error	Map Index	Location	General Performance
CORNING_6_N102	10.76	1	Corning, CA	Great
PRAXAIR_1_B1	10.93	2	Antioch, CA	Great
PARADISX_6_N004	12.23	3	Paradise, CA	Average
CERTTEED_1_N002	12.23	4	Madera Acres, CA	Average
URBAN_6_N005	16.87	5	San Diego, CA	Poor
PALA_6_N001	17.01	6	Pala, CA	Poor

Table 1: We select six geographically distinct nodes, shown in Fig. 5 (right), ordered from lowest to highest mean reconstruction error. The node names are the unique identifiers used by CAISO.

3.2. Optimal Training Window Size to Minimize Reconstruction Error

Besides r , selecting the window size, i.e., the dimensions of the input data $\mathbf{X}_t \in \mathbb{R}^{m \times n}$, is another important choice in DMD analysis including financial market forecasting[28]. We repeat sensitivity analysis of reconstruction error versus r for January 1st to 4th ($n = 24 \times 4 = 96$) with $m = 20, 100, 1000, 2000, 3000,$ and 6587 randomly selected nodes. Figure 6 shows that, in general, as m increases, the minimum reconstruction decreases. This suggests that standard DMD is extracting useful multimode dynamics and correlations such that more data (larger m) improves recapitulation of trends (lower error). With only 20 nodes, reconstruction error is between 25% and 40%, which suggests there is insufficient data to extract meaningful trends. We will revisit this finding in Section 4.

Next, we explore how the time horizon n impacts reconstruction error. Training windows for 2 through 15 days ($n \in \{48, 72, 96, \dots, 360\}$) were evaluated for a rolling horizon for a whole year (2015) using all 6587 nodes in Figure 7. Each box and whisker plot in Figure 7 reports the reconstruction error with the optimal r (lowest average reconstruction error for all nodes). For the year 2015 data set, we see the median reconstruction error gradually increases with n . The lowest median error is 16% for $n = 48$ (2 days), which is the shortest time horizon considered. These results are consistent with intuition. DAM prices are strongly periodic with well-documented seasonal, day of the week, and daily trends.[15] A two-day window captures two full cycles. But too long of a training window attempts to learn one monolithic linear operator \mathbf{A} in Eq. (4) to explain dynamics across too many days. We conclude by highlight our findings to maximize m (consider all nodes) and minimize n (short time horizon) are consistent with literature that suggests DMD inputs should be *skinny*, i.e., with $m \gg n$ [24].

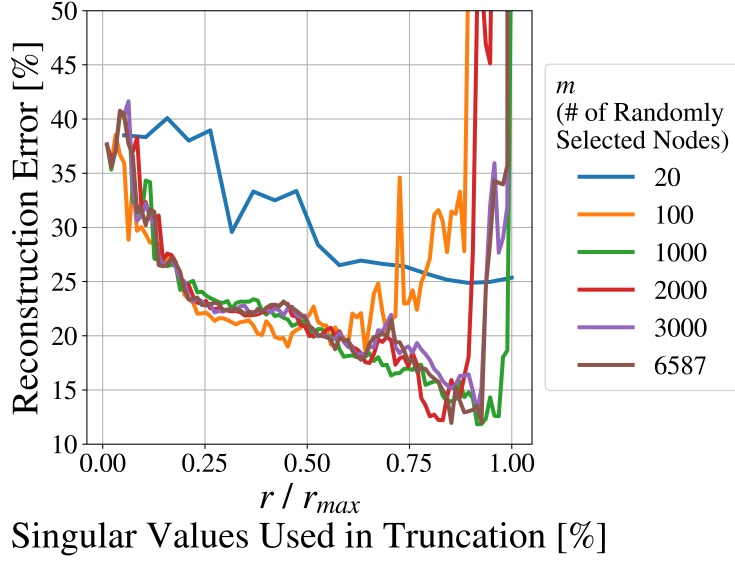


Figure 6: Sensitivity analysis of reconstruction error with respect to truncation level (r/r_{max}) for $m = 20$ to $m = 6587$. The nodes included in the analysis were randomly selected for $m < 6587$. The results show there is a clear benefit for selecting large m .

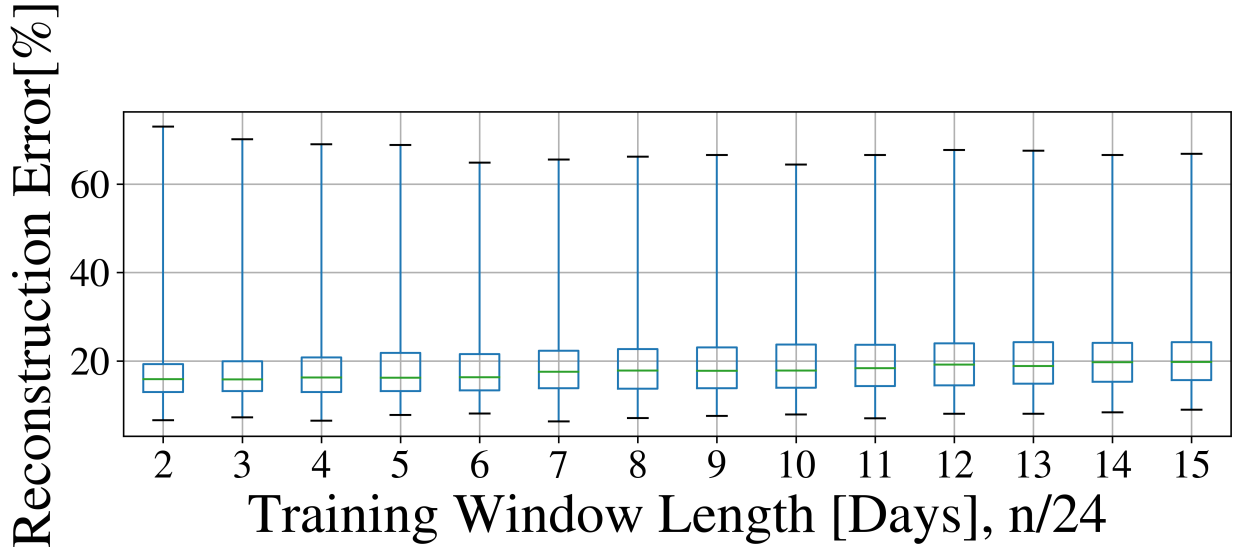


Figure 7: Sensitivity analysis reveals the impact of training window size for $n = 48$ (2 days) to $n = 360$ (15 days) on reconstruction error. The center of each box marks the median reconstruction error calculated across all nodes. The edges of the box mark the quartiles. The whiskers mark the smallest and largest reconstruction errors for an individual node, after excluding outliers. $n = 48$ gives the smallest median reconstruction error of 16%.

3.3. From Optimal Reconstruction to Optimal Forecasts

We now explore how insights for specifying r , m , and n to optimize reconstruction translate to forecasting. Figure 8 shows sensitivity analysis of forecasting and reconstruction errors versus truncation level r . For all three data sets considered, we observe the optimal r for reconstruction and forecasting are similar.

To generalize this trend, we conduct sensitivity analysis over the entire year with 4-day training windows in a rolling horizon framework. On average, there was only a 4% difference between r values that minimized reconstruction error and day 1 forecast errors. Differences were 0.4% and 0.8% for reconstruction error versus day 2 and day 3 forecast errors, respectively. From these results, we propose a simple heuristic: *truncate DMD the with best r for reconstruction to produce for the (near) optimal forecast.*

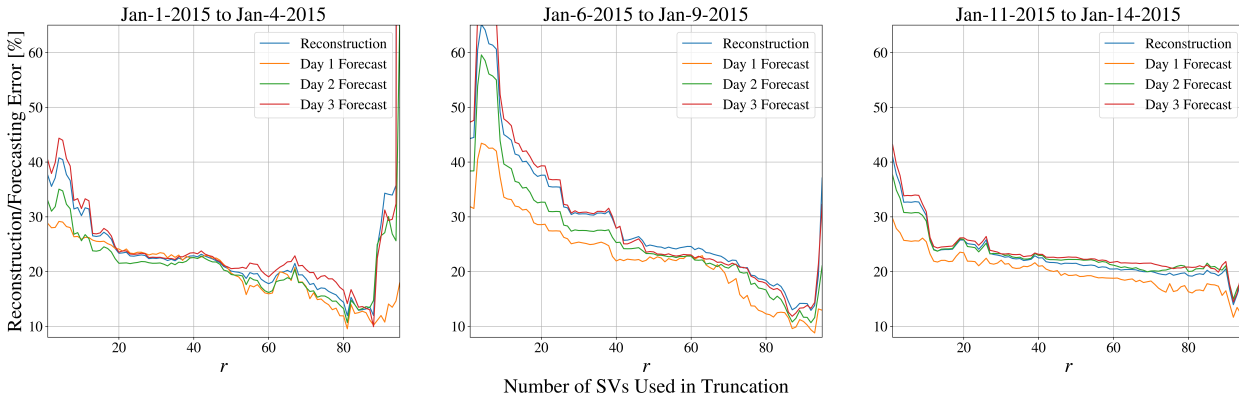


Figure 8: Reconstruction and forecasting errors are compared for three 4-day training windows ($n = 96$, $m = 6587$) for many different number of retained singular values r . For the first window, training is done with data from January 1st through 4th. January 5th, 6th, and 7th correspond to the day 1, 2, and 3 forecasts, respectively.

Next, we develop guidelines for the minimum training horizon length n to create stable multi-day DMD price forecasts. Computing optimal bids for participating in day-ahead energy markets, including CAISO, often require 48-hour or longer forecasts [12]. We perform a sensitivity analysis of all possible 2 to 15-day training windows in the year 2015 and record forecast errors for 1 to 4 days after the end of each training window. Fig. 9 shows forecasting error averaged over all of these windows. When the training window is shorter than the forecasting horizon, the forecasting error is large. For example, forecasting 4 days into the future using only a 1-day training horizon gives 50% average errors. In contrast, forecasting 4 days into the future with a 5-day or longer training horizon gives average errors less than 20%. This trend is consistent for all forecasting horizons considered and leads to another heuristic: *choose the training horizon length n to be longer than the desired forecasting horizon.* Beyond this threshold, the average forecasting error only gradually changes as n increases. For shorter forecasts, e.g., 1 day, longer training windows slightly increase forecasting error. But for longer forecasts, e.g., 4 days, longer training windows slightly decrease forecasting errors. We emphasize that while reconstruction error is minimized with shorter training horizons (see Fig. 7), forecasting requires longer horizons. We speculate that unplanned outages, transmission constraints, uncertain electricity demands, and other factors add noise to the energy prices. This noise causes reconstruction error to increase with training window length. But, for accurate forecasts, more data is required to separate meaningful dynamics from noise, and hence longer training horizons are required.

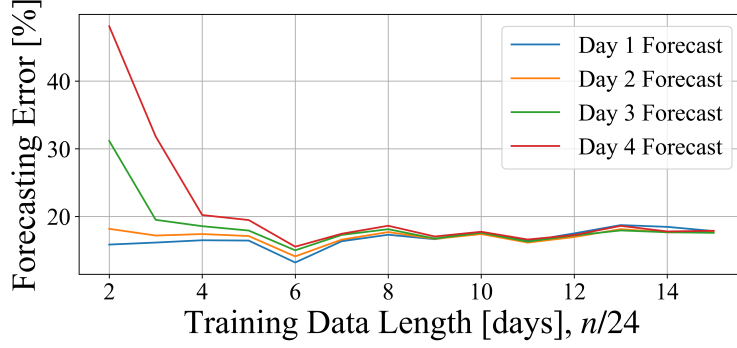


Figure 9: Mean forecasting errors for 6587 nodes ($m = 6587$) calculated with a rolling horizon for 1, 2, 3 and 4 days into the future. The forecasting horizon starts on January 17th and continues through the end of the year. The SV truncation r was performed for each window

3.4. DMD Mode Analysis

We now interpret the DMD modes, similar to previous financial market analyses [24, 28, 20]. Specifically, temporal modes (eigenvalues) capture price periodicity and spatial modes weight different locations (nodes) in the electricity market. Figure 10 shows spatial-temporal DMD modes for a 4-day training window $n = 96$ without truncation. In the top three plots, we see eigenvalues ordered by the prominence of the singular values. The first six singular values explain 99% of the variability in the dataset. Interestingly, as the most important singular values (lowest number) correspond to the lowest frequency oscillations. Beyond singular value 60, this linear trend between singular value number and oscillation frequency ceases. We hypothesize the least prominent singular values are localized fluctuations in energy prices caused by external factors. We find the real eigenvalue components are both positive and negative. We do not see a clear trend with real eigenvalue components and the number of retained singular values r . The middle three plots show how the three most prominent DMD modes ϕ_1 , ϕ_2 , and ϕ_3 are spatially distributed. We see all three modes place the least weight on prices from the San Francisco area. The bottom three plots explore the maximum, minimum, and median contributions to ϕ_1 . We see the minimum contribution has price spikes on the last day. This makes sense, as we expect the most prominent singular value to capture the average trends of the entire CA system and deemphasize localized events.

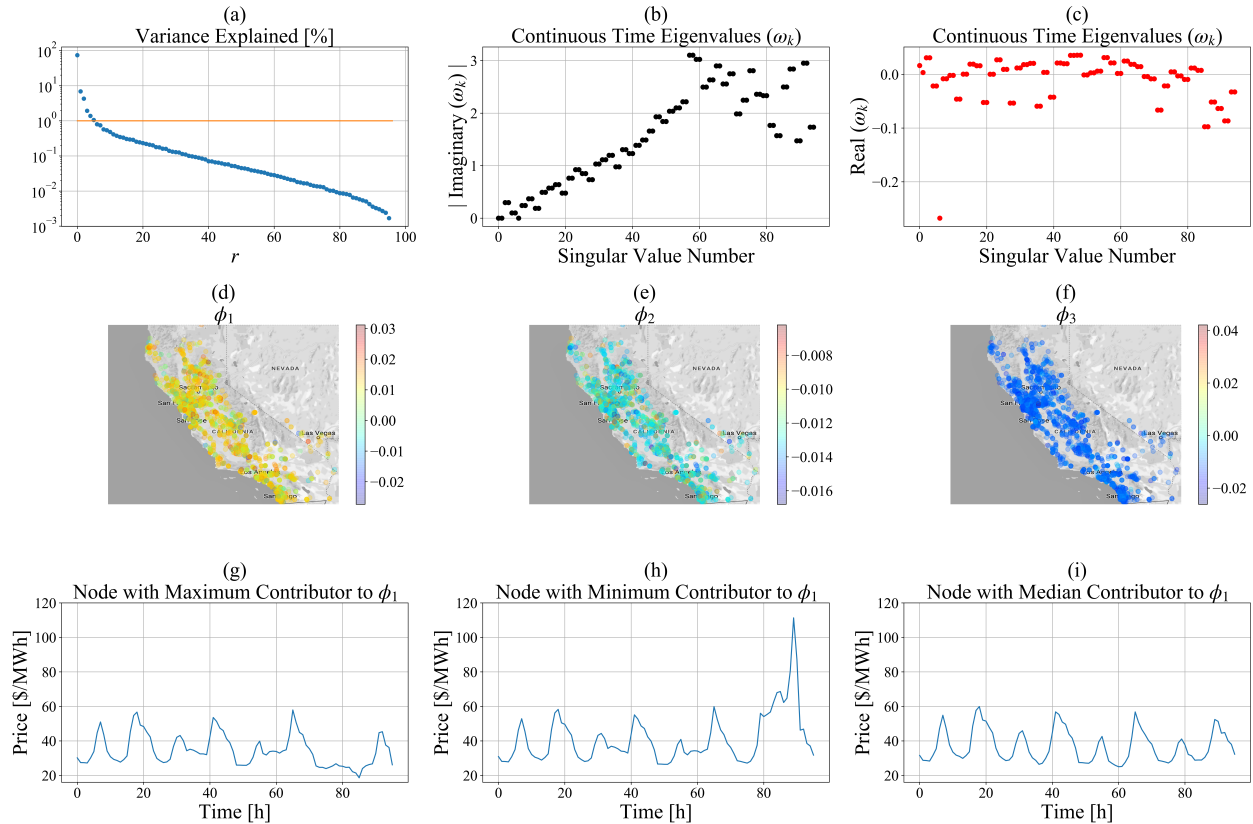


Figure 10: Visualization and interpretation of the DMD modes. **(a)** The fraction of variability that can be attributed to each singular value. The horizontal line marks 1% variability. We see that retaining the $r = 6$ largest singular values in the truncation explains over 99% of the variability in the dataset. **(b)** and **(c)** The imaginary and real components of the eigenvalues associated with each singular value ordered from largest to smallest. **(d)**, **(e)**, **(f)** Three most prominent spatial modes, ϕ_1 , ϕ_2 , and ϕ_3 , respectively. Larger absolute values (red or blue) indicate a larger weight for the specific location in the DMD mode. **(i)** Time-series price data from January 4th to 7th for corresponding to the nodes with the maximum (most positive), minimum (most negative), and median, respectively, weights in ϕ_1 .

4. ADMD Results and Discussion

As discussed in Section 3, standard DMD, even with optimal truncation, is only able to achieve a 16% reconstruction error for the CAISO dataset from 2015. We now show that ADMD overcomes these limitations and consistently achieve less than 5% reconstruction error.

4.1. ADMD is Less Sensitivity to Truncation

First, we explore the sensitivity of ADMD to truncation level r for three $n = 96$ training windows containing 100 randomly selected nodes and augmentation level $l = 48$. Figure 11 clearly shows ADMD is more robust with 1.5 to 3 times smaller reconstruction error than DMD. ADMD reconstruction error reaches a plateau around $r = 15\%$, suggesting ADMD successfully identifies a low-rank structure. To explain this, we verified the condition number of \mathbf{X} is 10,000 times larger than \mathbf{X}_{aug} . These results support our hypothesis that standard DMD struggles to extract low-rank trends because of the standing wave problem.

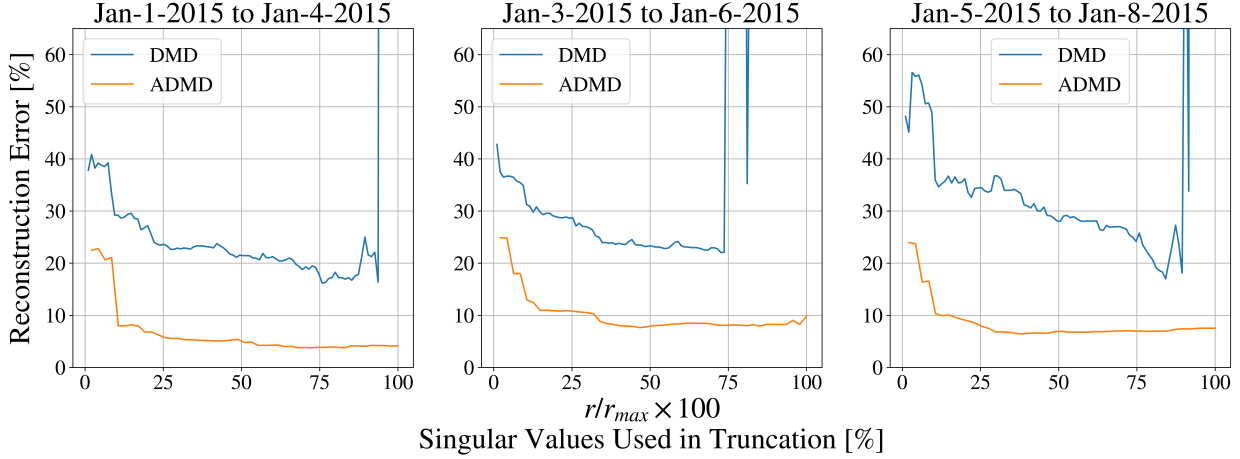


Figure 11: Sensitivity analysis of reconstruction error versus truncation level r shows that ADMD consistently outperforms DMD. For this comparison, $m = 100$, $l = 48$, and $n = 96$ such that $\mathbf{X} \in \mathbb{R}^{96 \times 100}$ and $\mathbf{X}_{\text{aug}} \in \mathbb{R}^{4800 \times 48}$. DMD performs much better in Figure 6 with $m = 6587$ whereas this Figure considers $m = 100$ to benchmark against ADMD.

4.2. ADMD Performs Best Augmenting Over At Least 1 Day

Next, we explore the impact of augmentation level l in Eq. 19 on ADMD performance. In Fig. 12, we see reconstruction error is not robust with $l = 12$ and $l = 18$. In contrast, reconstruction error reaches a plateau around truncation level $r/r_{\max} = 15\%$ with $l = 24$, $l = 36$, and $l = 48$. We hypothesize this is because the dominant price dynamics in energy markets have 24-hour periodicity. Based on these results, we recommend $l = 36$, $l = 48$, or a similar value.

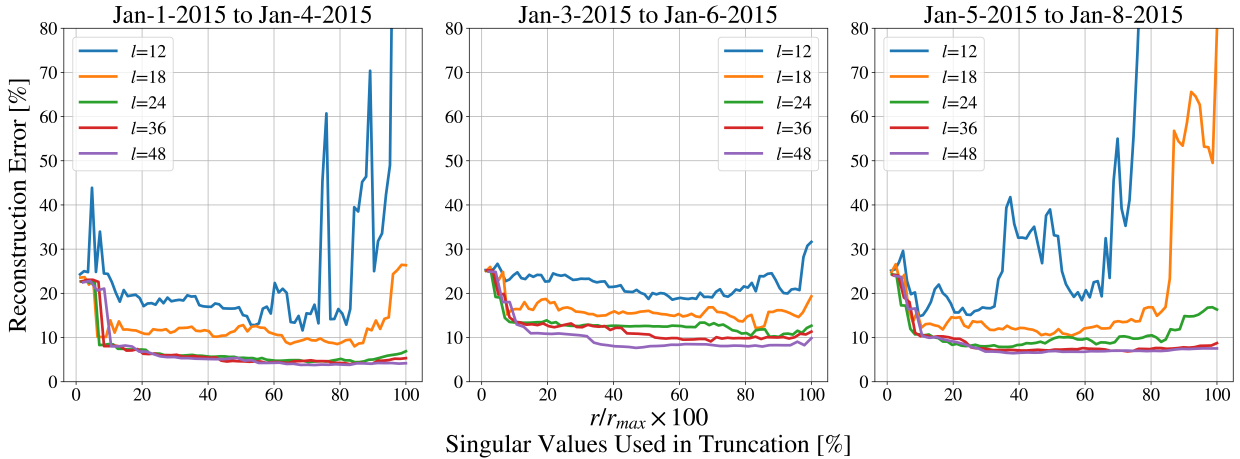


Figure 12: Sensitivity analysis of ADMD reconstruction error with respect to the truncation level r/r_{\max} where $r_{\max} = \min(m \cdot l, n - l)$. Results show $l \geq 24$ is needed for a robust low rank approximation with a small reconstruction error.

4.3. ADMD Does Not Benefit from Training Many Nodes Simultaneously

Next, we determine if ADMD benefits from analyzing multiple nodes simultaneously. Recall, DMD performs best when the data matrix \mathbf{X} is “tall and skinny”, i.e., there are many more nodes than timesteps. As consequence, we show in Section 3.2 DMD works best with $m = 6587$ (all nodes considered). But as

augmentation level l increases, the number of rows in \mathbf{X}_{aug} increase and the number of columns decrease. This means ADMD can be applied to a single node. Figure 13 compares reconstruction error for a single node and 100 randomly selected nodes over with $r = 8$ and $l = 48$ for 96 hours of training data. Surprisingly, we find reconstruction error only improves 0.65% when forecasting many nodes simultaneously. This suggests there is less valuable information available in the spatial dynamics than previously postulated.

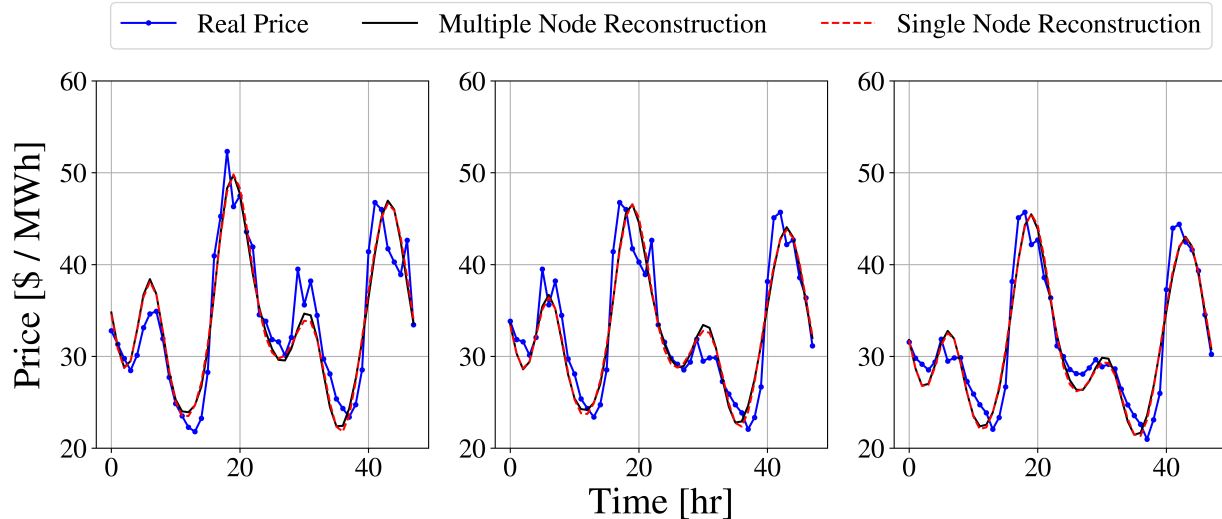


Figure 13: Comparison of ADMD with $m = 1$ (red) and $m = 100$. Each panel is a different 48-hour training window. There is little benefit to reconstructing and forecasting multiple nodes simultaneously.

4.4. ADMD Outperforms DMD for Forecasting

Next, we compare forecasting error from DMD and ADMD for a single node. Fig. 14 shows forecasts for January 16th and 17th, 2015 (48 hours) using the previous 96 hours for training. DMD forecasts were constructed with all 6587 nodes, whereas ADMD was only trained using the node of interest (same as Fig. 14) with $r = 8$ and $l = 48$. The results show the ADMD forecast closely tracks the realized price with around 10% forecasting error. In contrast, the DMD forecast fails to predict major temporal features and is out of phase.

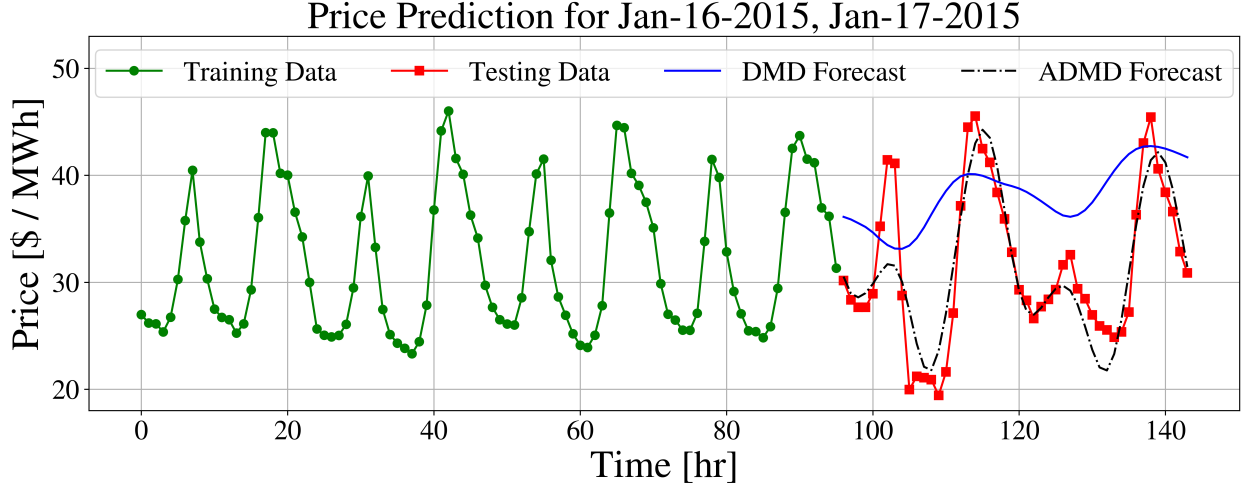


Figure 14: Comparison of DMD and ADMD forecasts using $n = 96$ training data (January 12th through 15th, 2015) and two-days of testing data (January 16th and 17th, 2015). The ADMD forecast is superior and most importantly captures the phase of the testing data.

5. Optimal Market Arbitrage for Energy Storage Systems

Accurate price forecasts are important for many aspects of grid reliability, including optimal control of individual energy systems such as conventional generators, grid-connected storage, microgrids, and virtual power plants. In this section, we benchmark the performance of deterministic model predictive control of a price-taker battery energy storage system self-scheduling into a wholesale energy market using different price forecasting techniques including DMD and ADMD.

5.1. Optimization Formulation

We consider a storage system with time-varying energy hold-up E_i at time i . The system can purchase (charge) or sell (discharge) energy to and from the market at rates c_i and d_i , respectively, during time interval $[i, i + 1)$. The energy storage operator seeks to maximize net revenue as follows:

$$\max_{\mathbf{c}, \mathbf{d}, \mathbf{E}} \sum_{i=0}^n x_i (d_i - c_i) \quad (21a)$$

$$\text{s.t. } E_{i+1} = E_i + c_i \sqrt{\eta} - \frac{d_i}{\sqrt{\eta}} \quad \forall i \in \{0, \dots, n-1\} \quad (21b)$$

where x_i is the market price forecast during time interval $[i, i + 1)$. The hold-up is constrained by an energy balance and subject to round trip efficiency η . Variable bounds, parameter values, and units are specified in Table 2. Figure ?? illustrates the interaction between the energy storage system and the energy market.

Fig. 15 compares the solution to Eq. (21) for the market node near Chino, CA. Price 48-hour forecasts were generated with ADMD using the previous 96 hours of prices for the node of interest as training data with $r = 8$ and $l = 48$. As expected, the optimal solution is to sell energy when the forecasted price is lowest and buy energy when the forecasted price is the highest. Although price forecasts are consistently biased (e.g., higher than the actual price) the periodicity (e.g., when prices increase and decrease) is correct, leading to an effective control strategy.

Variables	Description	Units	Dimensions	Range
\mathbf{c}	Charging Rate	MW	\mathbb{R}^n	$[0,1]$
\mathbf{d}	Discharging Rate	MW	\mathbb{R}^n	$[0,1]$
\mathbf{E}	Battery Charge	MWh	\mathbb{R}^n	$[0,4]$

Parameters	Description	Units	Dimensions	Range
$\sqrt{\eta}$	Round Trip Efficiency	-	\mathbb{R}^1	0.88
n	Length of Forecast	h	\mathcal{Z}	
E_0	Initial Battery Charge	MWh	\mathbb{R}^1	$[0,4]$
\mathbf{x}	Price Forecast	\$/MWh	\mathbb{R}^n	

Indices	Description	Units	Dimensions	Range
i	Time index	h	\mathcal{Z}	$[0,n]$

Table 2: Optimization variables and parameters used in the battery optimal control problem.

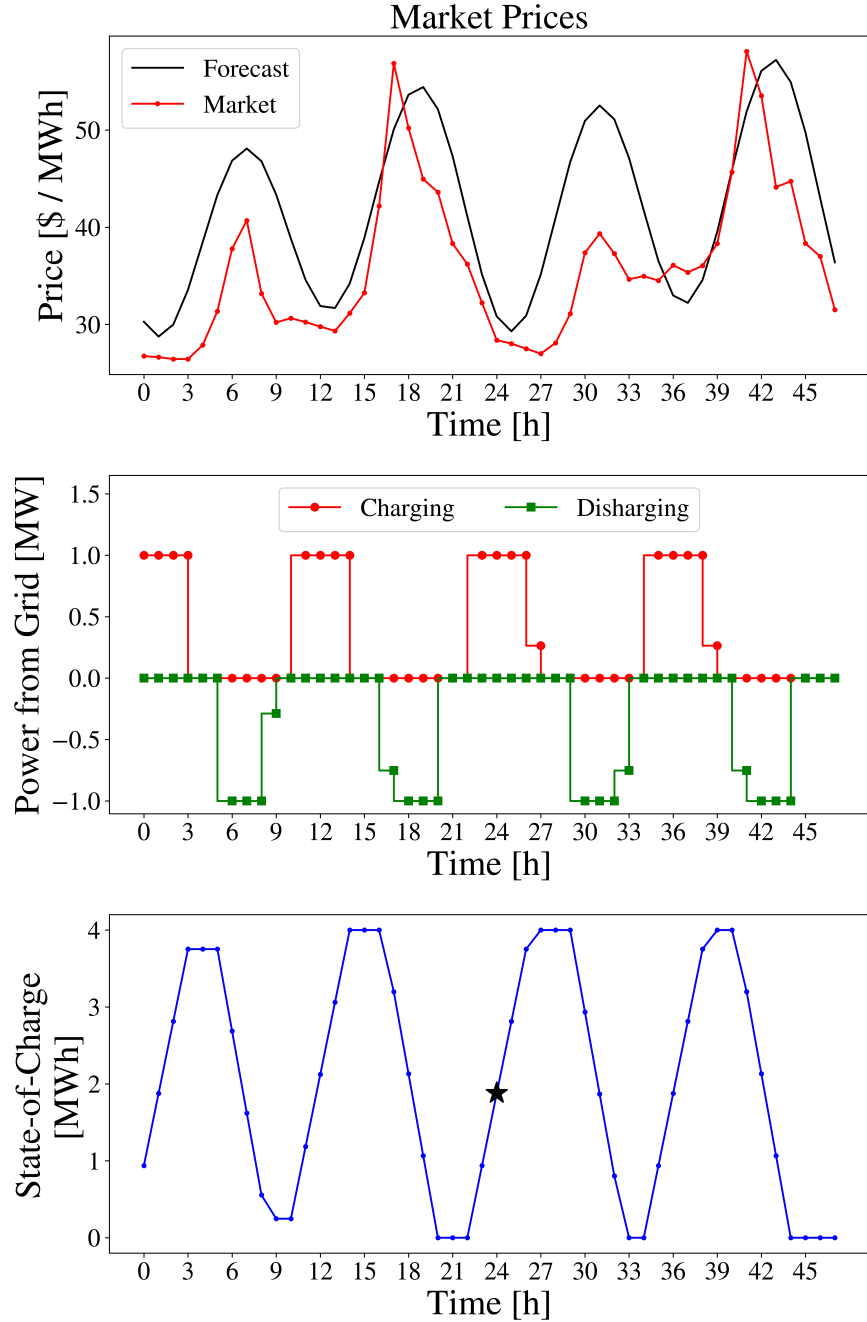


Figure 15: Example battery optimization solution for Chino, CA for January 7th and 8th, 2015. **(Top)** The historical market price and ADMD forecast both show 4 distinct price maximums which correspond with early morning (6 am) and late afternoon (3 or 4 pm). **(Middle)** As expected, the optimal arbitrage policy is to charge the battery (buy from the market) during low price periods and discharge (sell) at higher prices. **(Bottom)** The state-of-charge (SOC) in the battery is the integral of the charge and discharge signals. The black \star marks the SOC at the end of the first day (E_{24}). In the rolling horizon framework, this value is the fixed initial condition E_0 for the subsequent day.

5.2. Backcasting Benchmark

We use simple backcasting as a benchmark forecasting strategy[43]. In backcasting, prices for the n most recent observed days as used as forecasts for the next n days. We compare 48-hour rolling horizon forecasting errors for backcasting, DMD (48 hours of training data, optimal truncation), and ADMD (96 hours of data using, $r = 8, l = 48$) averaged over all 6587 nodes in CAISO for January 5th, 2015 to December 7th, 2015. Surprisingly, we found median forecasting errors of 17%, 26%, and 31% for backcasting, ADMD, and DMD respectively. Below we show the lowest median forecasting error does not always correspond with the maximum realized revenue. Sioshansi *et al* observe backcasting is extremely effective because it recapitulates the periodicity of day-ahead energy market prices. [42]

5.3. Rolling Horizon Computational Results

We now compare the performance of DMD, ADMD, and backcasting for energy arbitrage by repeatedly solving Eq. (21) in a rolling horizon framework for January th, 2015 to December 9th, 2015. Each instance of Eq. (21) is first solved with a 48 hour prediction horizon. The computed policy is then simulated over the first 24 hours and profits are recorded with the real (historical) price. The simulation window is then advanced 24 hours. The value for E_{24} for the previous day is copied to E_0 for the next day.

We start by considering all 6587 nodes in the CAISO market. For DMD, we use a 48-hour training window ($\mathbf{X}_t \in \mathbb{R}^{48 \times 6587}$) with optimal truncation. For ADMD, we generate forecasts individually for each node using $r = 8$ and $l = 48$ with 96 hours of training data (thus 6587 unique instances of $\mathbf{X}_{\text{aug}} \in \mathbb{R}^{48 \times 48}$). Backcasting was performed using the previous 48 hours of data. Figure 16 compares the market revenues for optimal arbitrage with perfect information, backcast, DMD, and ADMD price forecasts. The perfect information scenario, which uses historical prices in Eq. (21), represents the maximum available revenue. We see hot spots in the perfect information map, which show the greatest potential for arbitrage is in central California as well as the coasts. We find that DMD performs poorly, yielding only 24% of the perfect information revenue. This is expected as DMD suffers from the so-called standing wave problem. In contrast, ADMD and backcasting capture 81% and 84% of the perfect information revenue, respectively. Backcasting and ADMD capture the hot spot in central California but miss the opportunities along the coast. These results are summarized in Table 3.

Next we identified 2209 nodes ($2209 / 6587 = 33.5\%$ of the original data set) for which the prices were always between 0 \$ / MWh and 200 \$ / MWh during calendar year 2015. We refer to these as *Select Nodes*. We find that ADMD and backcasting recover 92% and 88% of potential revenues, respectively, for only select nodes. This is a modest increase compared to all nodes. DMD, however, recovers less relative profit when only considering select nodes. We hypothesize this is because DMD benefits from forecasting as many nodes as possible.

	Backcasting	DMD	ADMD
All Nodes	83.7%	24.3%	80.8%
Select Nodes	88.3%	21.7%	92.2%

Table 3: Comparison of net revenues, relative to the perfect information case, for each forecasting method for January 5th to December 9th in 2015. Select nodes are locations where the market price was always between 0 \$ / MWh and 200 \$ / MWh.

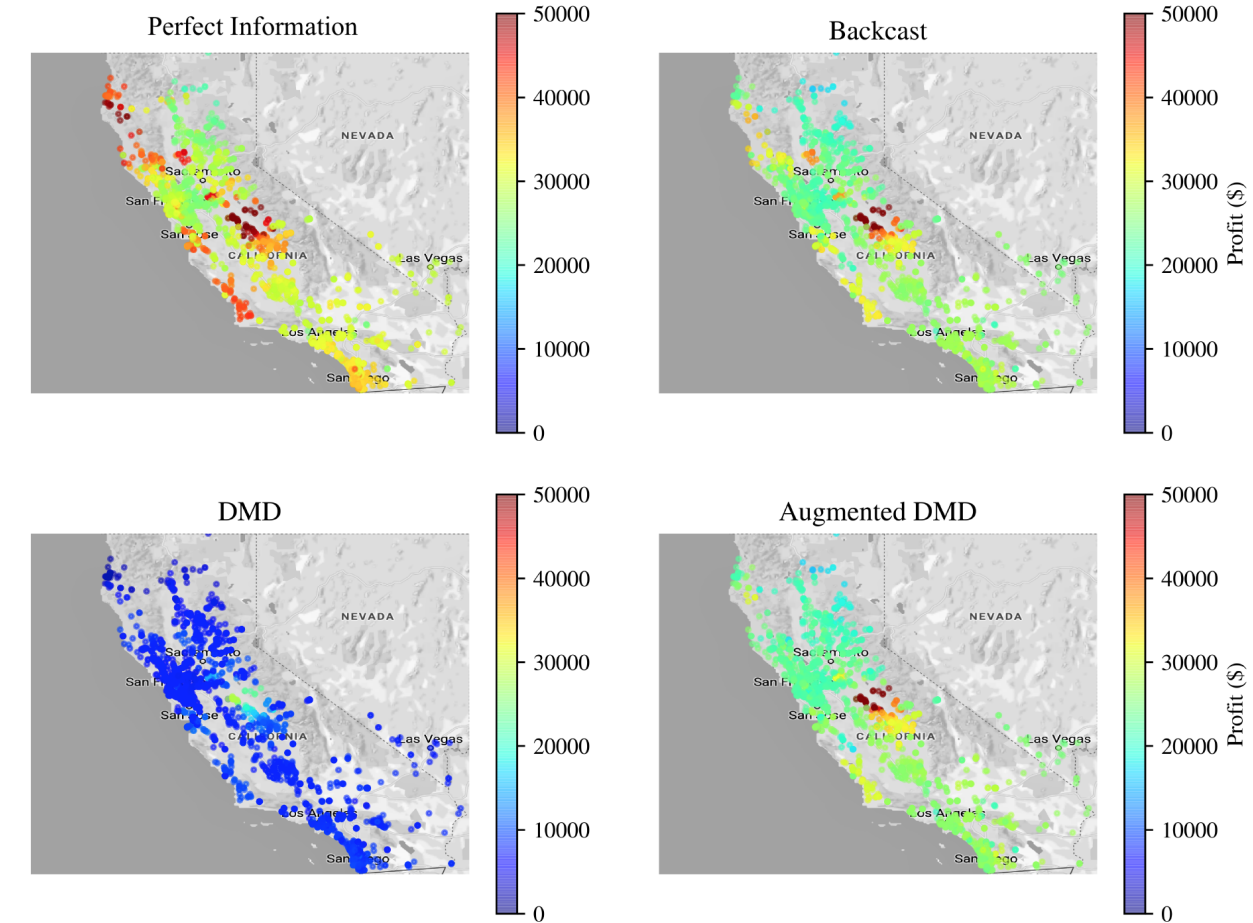


Figure 16: Simulated net profits for energy arbitrage for all nodes in the CAISO day-ahead market in 2015.

DMD forecasting greatly underperforms relative to ADMD and backcasting. ADMD and backcasting perform well through most of CAISO, but fail to capture the largest revenue opportunities across the coast. The perfect information reference shows the maximum available revenue.

6. Computational Timing Results

A key advantage of DMD and ADMD is computational speed. Training is quick, often limited by the SVD calculation, and forecasting is almost instantaneous using the explicit formula in Eq. (18). To demonstrate these benefits, we performed simple computational benchmarks using an Auto-Regressive Integrated Moving Average (ARIMA) model implemented in the StatsModel package. We considered 48 hours of training data and ARIMA model form $(p, d, q) = (7, 1, 1)$. Training and forecasting with an ARIMA model took 33 seconds per node on average. In contrast, it took 11.2 seconds in total for ADMD ($r = 8, l = 48, n = 96$) to train and forecast for all 6587 nodes, which is 0.002 seconds per node. On a per-node basis, ADMD is 20,000 times faster than ARIMA. Our custom implementation of DMD using SciPy is available on GitHub (<https://github.com/dowlinglab/dmd-energy-markets>). These benchmark tests were performed on a 2014 MacBook Pro (3.0 GHz Intel Core i7 processor, 16 GB of RAM) using an Anaconda Python environment.

7. Conclusions and Future Work

In this paper, we extend the success of DMD in forecasting financial markets[28, 10] to wholesale electricity markets. We find that standard DMD is unable to extract coherent low-rank structures in California Independent System Operator (CAISO) day-ahead market prices for the calendar year 2015. Instead, we

find Augmented DMD (ADMD), which is equivalent to differencing in a time-series model, overcomes the so-called standing wave problem and can consistently reconstruct and forecast with errors less than 5% with a low-rank approximation. We benchmark DMD, ADMD, and backcasting in a rolling horizon model predictive control case study for energy arbitrage with a market connected battery. We show that ADMD outperforms backcasting by capturing 92% versus 88% of available net market revenue at nodes without extreme prices. We also find that ADMD is up to 20,000 times faster than ARIMA for training and forecasting.

There are several opportunities to further enhance DMD for energy market forecasting. Denoising procedures can greatly improve the performance of DMD with experimental data and should be exploring for energy price forecasting.[11, 19] Multiresolution DMD may better identify and recapitulate dynamics at disparate time-scales in electricity markets.[23] Both of these methods may help explain why ADMD as currently implemented does not benefit from forecasting multiple nodes simultaneously. We also plan to more thoroughly benchmark DMD against other forecasting techniques, including emerging machine learning methods, to understand trade-offs between data requirements, computational time, need for model tuning, and forecasting error. Real-time markets (RTMs) set prices in 5 to 15-minute increments and are much more volatile than day-ahead markets.[13] Studying the performance of DMD for RTMs and other geographic regions is another future opportunity.

References

- [1] Jamshid Aghaei and Mohammad-Iman Alizadeh. “Demand response in smart electricity grids equipped with renewable energy sources: A review”. In: *Renewable and Sustainable Energy Reviews* 18 (2013), pp. 64–72.
- [2] N. Amjady. “Day-Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network”. In: *IEEE Transactions on Power Systems* 21.2 (2006), pp. 887–896. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2006.873409. URL: <http://ieeexplore.ieee.org/document/1626395/>.
- [3] Nima Amjady and Farshid Keynia. “Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method”. In: *International Journal of Electrical Power & Energy Systems* 30.9 (2008), pp. 533–546. ISSN: 0142-0615. DOI: 10.1016/J.IJEPES.2008.06.001. URL: <https://www.sciencedirect.com/science/article/pii/S0142061508000628>.
- [4] S Ashok and Rangan Banerjee. “An optimization mode for industrial load management”. In: *IEEE Transactions on Power Systems* 16.4 (2001), pp. 879–884.
- [5] Pedro M Castro, Iiro Harjunkoski, and Ignacio E Grossmann. “Optimal scheduling of continuous plants with energy constraints”. In: *Computers & chemical engineering* 35.2 (2011), pp. 372–387.
- [6] J.P.S. Catalão et al. “Short-term electricity prices forecasting in a competitive market: A neural network approach”. In: *Electric Power Systems Research* 77.10 (2007), pp. 1297–1304. ISSN: 0378-7796. DOI: 10.1016/J.EPSR.2006.09.022. URL: <https://www.sciencedirect.com/science/article/pii/S0378779606002422>.
- [7] Chris Chatfield. *The analysis of time series: an introduction*. Chapman and Hall/CRC, 2003.
- [8] A.J. Conejo et al. “Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models”. In: *IEEE Transactions on Power Systems* 20.2 (2005), pp. 1035–1042. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2005.846054. URL: <http://ieeexplore.ieee.org/document/1425601/>.
- [9] Antonio J. Conejo et al. “Forecasting electricity prices for a day-ahead pool-based electric energy market”. In: *International Journal of Forecasting* 21.3 (2005), pp. 435–462. ISSN: 0169-2070. DOI: 10.1016/J.IJFORECAST.2004.12.005. URL: <https://www.sciencedirect.com/science/article/pii/S0169207004001311>.
- [10] Ling-xiao Cui and Wen Long. “Trading strategy based on dynamic mode decomposition: Tested in Chinese stock market”. In: *Physica A: Statistical Mechanics and its Applications* 461 (2016), pp. 498–508. ISSN: 0378-4371. DOI: 10.1016/J.PHYSA.2016.06.046. URL: <https://www.sciencedirect.com/science/article/pii/S0378437116303144>.
- [11] Scott TM Dawson et al. “Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition”. In: *Experiments in Fluids* 57.3 (2016), p. 42.

- [12] Alexander W. Dowling, Ranjeet Kumar, and Victor M. Zavala. “A multi-scale optimization framework for electricity market participation”. In: *Applied Energy* 190 (2017), pp. 147–164. ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2016.12.081. URL: <https://www.sciencedirect.com/science/article/pii/S0306261916318487>.
- [13] Alexander W. Dowling and Victor M. Zavala. “Economic opportunities for industrial systems from frequency regulation markets”. In: *Computers & Chemical Engineering* 114 (2018), pp. 254–264. ISSN: 0098-1354. DOI: 10.1016/J.COMPCHEMENG.2017.09.018. URL: <https://www.sciencedirect.com/science/article/pii/S009813541730323X>.
- [14] Alexander W Dowling, Tian Zheng, and Victor M Zavala. “A decomposition algorithm for simultaneous scheduling and control of CSP systems”. In: *AIChE Journal* 64.7 (2018), pp. 2408–2417.
- [15] Alexander W Dowling, Tian Zheng, and Victor M Zavala. “Economic assessment of concentrated solar power technologies: a review”. In: *Renewable and Sustainable Energy Reviews* 72 (2017), pp. 1019–1032.
- [16] Robert L Fares, Jeremy P Meyers, and Michael E Webber. “A dynamic model-based estimate of the value of a vanadium redox flow battery for frequency regulation in Texas”. In: *Applied Energy* 113 (2014), pp. 189–198.
- [17] Feng Gao et al. “Forecasting power market clearing price and quantity using a neural network method”. In: *2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134)*. Vol. 4. IEEE, 2000, pp. 2183–2188. ISBN: 0-7803-6420-1. DOI: 10.1109/PESS.2000.866984. URL: <http://ieeexplore.ieee.org/document/866984/>.
- [18] R.C. Garcia et al. “A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices”. In: *IEEE Transactions on Power Systems* 20.2 (2005), pp. 867–874. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2005.846044. URL: <http://ieeexplore.ieee.org/document/1425583/>.
- [19] Maziar S Hemati et al. “De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets”. In: *Theoretical and Computational Fluid Dynamics* 31.4 (2017), pp. 349–368.
- [20] Jia-Chen Hua et al. “Using dynamic mode decomposition to extract cyclic behavior in the stock market”. In: *Physica A: Statistical Mechanics and its Applications* 448 (2016), pp. 172–180. ISSN: 0378-4371. DOI: 10.1016/J.PHYSA.2015.12.059. URL: <https://www.sciencedirect.com/science/article/pii/S0378437115010870>.
- [21] MG Ierapetritou et al. “Cost minimization in an energy-intensive plant using mathematical programming approaches”. In: *Industrial & engineering chemistry research* 41.21 (2002), pp. 5262–5277.
- [22] Christoph Kost, Christoph M Flath, and Dominik Möst. “Concentrating solar power plant investment and operation decisions under different price and support mechanisms”. In: *Energy Policy* 61 (2013), pp. 238–248.
- [23] J Nathan Kutz, Xing Fu, and Steven L Brunton. “Multiresolution dynamic mode decomposition”. In: *SIAM Journal on Applied Dynamical Systems* 15.2 (2016), pp. 713–735.
- [24] J. Nathan Kutz et al. *Dynamic Mode Decomposition*. 2017. DOI: 10.1137/1.9781611974508. arXiv: 1506.00564.
- [25] Enrique Lizarraga-Garcia et al. “Optimal operation of a solar-thermal power plant with energy storage and electricity buy-back from grid”. In: *Energy* 51 (2013), pp. 61–70.
- [26] Alicia Troncoso Lora et al. “Electricity Market Price Forecasting Based on Weighted Nearest Neighbors Techniques”. In: *IEEE Transactions on Power Systems* 22.3 (2007), pp. 1294–1301. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2007.901670. URL: <http://ieeexplore.ieee.org/document/4282040/>.
- [27] Seyed Hossein Madaeni, Ramteen Sioshansi, and Paul Denholm. “How thermal energy storage enhances the economic viability of concentrating solar power”. In: *Proceedings of the IEEE* 100.2 (2011), pp. 335–347.
- [28] Jordan Mann and J. Nathan Kutz. “Dynamic mode decomposition for financial trading strategies”. In: *Quantitative Finance* 16.11 (2016), pp. 1643–1655. ISSN: 1469-7688. DOI: 10.1080/14697688.2016.1170194. URL: <https://www.tandfonline.com/doi/full/10.1080/14697688.2016.1170194>.

- [29] Sumit Mitra et al. “Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes”. In: *Computers & Chemical Engineering* 38 (2012), pp. 171–184.
- [30] Iman Naziri Moghaddam, Badrul Chowdhury, and Milad Doostan. “Optimal sizing and operation of battery energy storage systems connected to wind farms participating in electricity markets”. In: *IEEE Transactions on Sustainable Energy* 10.3 (2018), pp. 1184–1193.
- [31] Tomas W. Muld, Gunilla Efraimsson, and Dan S. Henningson. “Flow structures around a high-speed train extracted using Proper Orthogonal Decomposition and Dynamic Mode Decomposition”. In: *Computers & Fluids* 57 (2012), pp. 87–97. ISSN: 0045-7930. DOI: 10.1016/J.COMPFLUID.2011.12.012. URL: <https://www.sciencedirect.com/science/article/pii/S0045793011003847>.
- [32] Joannah I Otashu and Michael Baldea. “Scheduling chemical processes for frequency regulation”. In: *Applied Energy* 260 (2020), p. 114125.
- [33] Raúl Pino et al. “Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks”. In: *Engineering Applications of Artificial Intelligence* 21.1 (2008), pp. 53–62. ISSN: 0952-1976. DOI: 10.1016/J.ENGAPPAI.2007.02.001. URL: <https://www.sciencedirect.com/science/article/pii/S0952197607000097>.
- [34] Pierre Pinson, Henrik Madsen, et al. “Benefits and challenges of electrical demand response: A critical review”. In: *Renewable and Sustainable Energy Reviews* 39 (2014), pp. 686–699.
- [35] C.P. Rodriguez and G.J. Anders. “Energy Price Forecasting in the Ontario Competitive Power System Market”. In: *IEEE Transactions on Power Systems* 19.1 (2004), pp. 366–374. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2003.821470. URL: <http://ieeexplore.ieee.org/document/1266590/>.
- [36] P. J. Schmid et al. “Applications of the dynamic mode decomposition”. In: *Theoretical and Computational Fluid Dynamics* 25.1-4 (2011), pp. 249–259. ISSN: 0935-4964. DOI: 10.1007/s00162-010-0203-9. URL: <http://link.springer.com/10.1007/s00162-010-0203-9>.
- [37] Peter J. Schmid. “Application of the dynamic mode decomposition to experimental data”. In: *Experiments in Fluids* 50.4 (2011), pp. 1123–1130. ISSN: 0723-4864. DOI: 10.1007/s00348-010-0911-3. URL: <http://link.springer.com/10.1007/s00348-010-0911-3>.
- [38] PETER J. SCHMID. “Dynamic mode decomposition of numerical and experimental data”. In: *Journal of Fluid Mechanics* 656 (2010), pp. 5–28. ISSN: 0022-1120. DOI: 10.1017/S0022112010001217. URL: http://www.journals.cambridge.org/abstract{_}S0022112010001217.
- [39] M. Shafie-khah, M. Parsa Moghaddam, and M.K. Sheikh-El-Eslami. “Price forecasting of day-ahead electricity markets using a hybrid forecast method”. In: *Energy Conversion and Management* 52.5 (2011), pp. 2165–2169. ISSN: 0196-8904. DOI: 10.1016/J.ENCONMAN.2010.10.047. URL: <https://www.sciencedirect.com/science/article/pii/S0196890410005212>.
- [40] Yue Shao and Victor M Zavala. “Space-time dynamics of electricity markets incentivize technology decentralization”. In: *Computers & Chemical Engineering* 127 (2019), pp. 31–40.
- [41] Fereidoon P Sioshansi, ed. *Evolution of Global Electricity Markets: New paradigms, new challenges, new approaches*. Academic Press, 2013.
- [42] Ramteen Sioshansi and Paul Denholm. “The value of concentrating solar power and thermal energy storage”. In: *IEEE Transactions on Sustainable Energy* 1.3 (2010), pp. 173–183.
- [43] Ramteen Sioshansi et al. “Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects”. In: *Energy economics* 31.2 (2009), pp. 269–277.
- [44] Farshud Sorourifar, Victor M Zavala, and Alexander W Dowling. “Integrated Multiscale Design, Market Participation, and Replacement Strategies for Battery Energy Storage Systems”. In: *IEEE Transactions on Sustainable Energy* 11.1 (2018), pp. 84–92.
- [45] B.R. Szkuta, L.A. Sanabria, and T.S. Dillon. “Electricity price short-term forecasting using artificial neural networks”. In: *IEEE Transactions on Power Systems* 14.3 (1999), pp. 851–857. ISSN: 08858950. DOI: 10.1109/59.780895. URL: <http://ieeexplore.ieee.org/document/780895/>.

- [46] Zhongfu Tan et al. “Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models”. In: *Applied Energy* 87.11 (2010), pp. 3606–3610. ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2010.05.012. URL: <https://www.sciencedirect.com/science/article/pii/S0306261910001807>.
- [47] Jonathan H. Tu et al. “On Dynamic Mode Decomposition: Theory and Applications”. In: (2013). DOI: 10.3934/jcd.2014.1.391. arXiv: 1312.0041. URL: <http://arxiv.org/abs/1312.0041><http://dx.doi.org/10.3934/jcd.2014.1.391>.
- [48] Rahul Walawalkar, Jay Apt, and Rick Mancini. “Economics of electric energy storage for energy arbitrage and regulation in New York”. In: *Energy Policy* 35.4 (2007), pp. 2558–2568.
- [49] Qi Zhang et al. “Air separation with cryogenic energy storage: optimal scheduling considering electric energy and reserve markets”. In: *AIChE Journal* 61.5 (2015), pp. 1547–1558.
- [50] Xiao Zhang et al. “Demand response of ancillary service from industrial loads coordinated with energy storage”. In: *IEEE Transactions on Power Systems* 33.1 (2017), pp. 951–961.

Acknowledgments

C.T.E. is grateful for support and guidance from Dr. Scott Hampton at the Notre Dame Center for Research Computing. C.T.E. also thanks Prof. Jeffery Kantor and Prof. Patrick Wensing for serving on his senior thesis committee.