Properties of Energy-Price Forecasts for Scheduling

Georgiana Ifrim, Barry O'Sullivan, and Helmut Simonis

Cork Constraint Computation Centre
Department of Computer Science, University College Cork, Ireland
{g.ifrim|b.osullivan|h.simonis}@4c.ucc.ie

Abstract. Wholesale electricity markets are becoming ubiquitous, offering consumers access to competitively-priced energy. The cost of energy is often correlated with its environmental impact; for example, environmentally sustainable forms of energy might benefit from subsidies, while the use of high-carbon sources might be discouraged through taxes or levies. Reacting to real-time electricity price fluctuations can lead to high cost savings, in particular for large energy consumers such as data centres or manufacturing plants. In this paper we focus on the challenge of day-ahead energy price prediction, using the Irish Single Electricity Market Operator (SEMO) as a case-study. We present techniques that significantly out-perform SEMO's own prediction. We evaluate the energy savings that are possible in a production scheduling context, but show that better prediction does not necessarily yield energy-cost savings. We explore this issue further and characterize, and evaluate, important properties that an energy price predictor must have in order to give rise to significant scheduling-cost savings in practice.

1 Introduction

Short-term forecasting of electricity prices is relevant for a range of applications, from operations scheduling to designing bidding strategies. For example, an increasing amount of work has recently focused on the impact of energy-aware schedules for operating data centres and production lines [7, 8, 18, 22, 27], where tasks are scheduled based on technical requirements and energy-saving objectives. Furthermore, [22] shows that simple scheduling strategies that can react to the spot-electricity price can save existing systems millions of dollars every year in electricity costs. Since energy costs constitute the largest proportion of a data centre's or production line's running costs [14], designing energy-price-saving schedules becomes critical. A number of recent papers [5, 17, 19, 26] focus on reducing both power usage and power cost by taking real-time energy price into consideration. The scheduling problems addressed differ, e.g., [5] focuses on scheduling in data centres, where tasks can be executed at various time slots.

Prior work has shown that electricity spot prices are one of the most challenging types of time series in terms of simulation and forecasting [4, 16, 21, 29]. A large number of techniques have been employed to predict energy prices; see [4, 29] for a review. Time series models, Neural Networks (NN) and Support Vector Machines (SVM) were proven to have some degree of success depending on the actual design and volatility of the markets. For example, [6] has studied NN models for forecasting prices in the Pennsylvania, New Jersey, Maryland as well as the Spanish market. NN and SVM were employed in [3, 12, 23] for forecasting prices in the New England, Ontario and the Australian markets. SVM models were generally shown to perform similarly to the NN models, while often being more scalable and more accurate than competitors. An interesting aspect of price forecasting is that the user may not neccesarily require a good numeric estimate of the actual price, but rather need a good estimation of the price class, e.g., is the price higher or lower than a pre-defined user-threshold. This problem was recently addressed in [29] which trained SVM-based price classification models for the Ontario and Alberta markets.

In this paper, we analyze the Irish electricity market, from market design and publicly available data, to several machine learning modelling strategies and their impact on day-ahead energy price forecasting, as judged by classical error measures (e.g., Mean Squared Error). Furthermore, we investigate the effect of using the proposed forecasting models for designing energy-price-saving schedules. To the best of our knowledge, this is the first analysis of the impact of various properties of day-ahead electricity price forecasts on price-aware scheduling.

Under EU initiatives Ireland has an obligation to supply at least 20% of its primary energy consumption from renewable sources by 2020 [10]. The Irish government has set the following targets in 2007 for its energy usage: no oil in electricity generation by 2020, 15% of electricity from renewable resources by 2010, and 33-40% by 2020. Wind is the most abundant renewable energy source available in Ireland [10, 13]. However, introducing such renewable energy sources introduces volatility into the market, making energy price prediction and cost-efficient planning considerably more challenging.

We build on prior literature for short-term (numeric) price forecasting and price classification, and investigate SVM models for these tasks. We propose two SVM models that reduce the numeric price forecasting error of the market operator by 24-28% (Mean Squared Error). Furthermore, we investigate the usage of these models for price classification and for designing price-aware operation schedules. We evaluate the savings that are possible in a production scheduling context, but show that better numeric prediction does not necessarily yield energy-cost savings. We explore this issue further and characterize, and evaluate, important properties that an energy price predictor must have in order to give rise to significant scheduling-cost savings in practice. Therefore, this paper brings together the disciplines of machine learning and combinatorial optimisation to study the most appropriate types of energy price prediction models to use in a scheduling context. In addition, this paper considers this problem in a real-life setting: the Irish energy market.

2 The Irish Electricity Market

The Irish electricity market is an auction-based market, with spot prices being computed every half-hour of a trading day. The methodology for calculating the price in the Irish all-island market is as follows: every half-hour of the trading day, the Single Electricity Market Operator (SEMO)[1] calculates the System Marginal Price (SMP). The SMP has two components: the *Shadow Price* representing the marginal cost per 1MW of power necessary to meet demand in a particular half-hour trading period, within an unconstrained schedule, i.e., no power transmission congestions; and the *Uplift* component, added to the Shadow Price in order to ensure the generators recover their total costs, i.e., start-up and no-load costs [16].

One day ahead of the trade-day the generators have to submit technical and commercial offer data [24]: incremental price-quantity bids summarizing how much supply for what price does a generator offer to provide every half-hour, and the technical specifications of the generator such as start-up costs, maximum capacity, minimum on/off times. Only price-making generator units, that are not under test, are represented individually within the Market Scheduling and Pricing (MSP) software. Non-price making units are scheduled either based on submitted nominations or forecast data in the case of wind units. Once the load met by generation from non-price making units has been removed, then price making generator units are scheduled in merit order according to their bids to meet the remaining load. The SMP is bounded by a Market Price Cap (€1000/MWh) and a Market Price Floor (€-100/MWh), which are set by the Regulatory Authorities.

Two runs of the Market Scheduling and Pricing Software are particularly relevant each trading day. The Ex-Ante (EA) run is carried out one day prior to the trade date which is being scheduled and as such uses entirely forecast wind and load data. A schedule of half-hourly forecasted SMP, shadow price, load and wind generation is produced by the market operator (SEMO) for the coming trade-day. The Ex-Post Initial (EP2) run is carried out four days after the trade date which is being scheduled, and as such is able to utilize full sets of actual wind and load data with no forecast values. The system marginal prices produced in the EP2 run are used for weekly invoicing and the SMP determined in the EP2 run for a given half hour trading period is the price applicable to both generators and suppliers active in such a trading period.

3 Price Forecasting Models

We present an approach to building price forecasting models for the Irish electricity market: the factors influencing the price, data collection and appropriate forecasting models. We show that we significantly improve the forecast of the market operator, thus providing a more reliable price prediction for the next trade-day.

[1] `http://www.sem-o.com`

3.1 Data Collection and Analysis

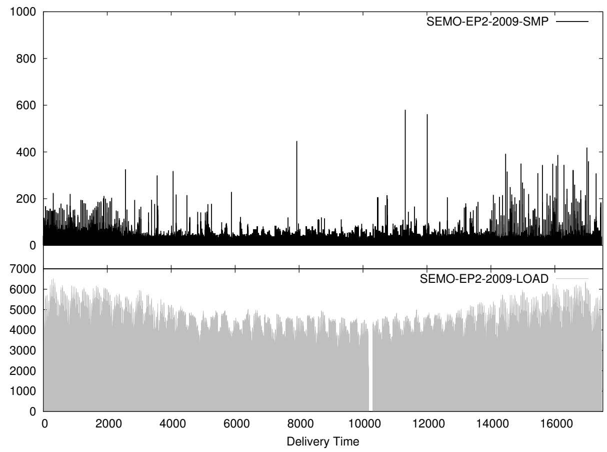
We begin our study of the Irish electricity market by analyzing the price and load (i.e., demand) profile from January 2009 to June 2011. Figure 1 shows the actual half-hourly price (top frame) and demand (bottom frame). We notice that the load profile is fairly similar over time, showing clear periodicity, with higher load in the cold months. The price is much more volatile, with high variations during both cold and warm months. We also notice that the price volatility and magnitude increased considerably from 2009 to 2011. Table 1 shows some statistics about the price in this period. We notice the increasing median and average price, as well as the increased price volatility (captured by the standard deviation) over time. We believe this could be explained by increasing fuel prices and the ramp-up of wind-generated power, as well as other factors, such as a higher percentage of unscheduled generator outages in 2011.

Table 1: Statistics of the Irish SMP for 2009 to mid-2011.

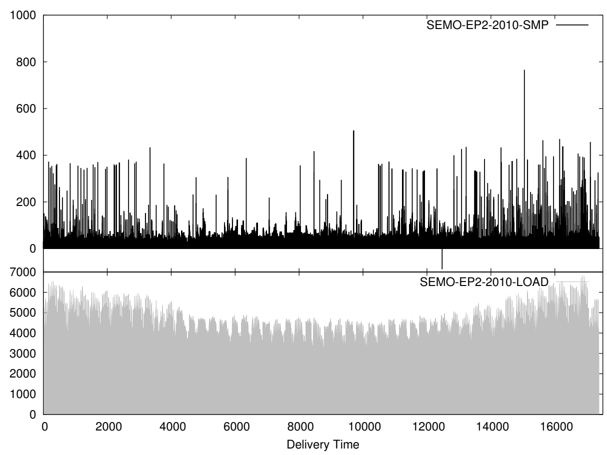
Year	Min	Median	Mean	Stdev	Max
2009	4.12	38.47	43.53	24.48	580.53
2010	-88.12	46.40	53.85	35.49	766.35
2011	0	54.45	63.18	35.79	649.48

Although there seems to be some correlation between demand and price, the price spikes seem to be highly influenced by a combination of additional factors. For example, the half-hourly demand is typically covered by the available wind-generated power, with the remaining load being covered using the generator bids sorted by price merit, with generators using more expensive fuel having higher price bids. If the forecasted load, wind-power and expected supply quantities are unreliable, due to the poor quality of the forecast or unexpected outages, this will affect the market operator scheduling of generators, leading to price volatility. Figure 2 offers a closer look at the price versus load pattern in 2011, for the first week of the year and the week with the maximum price up to mid-2011.

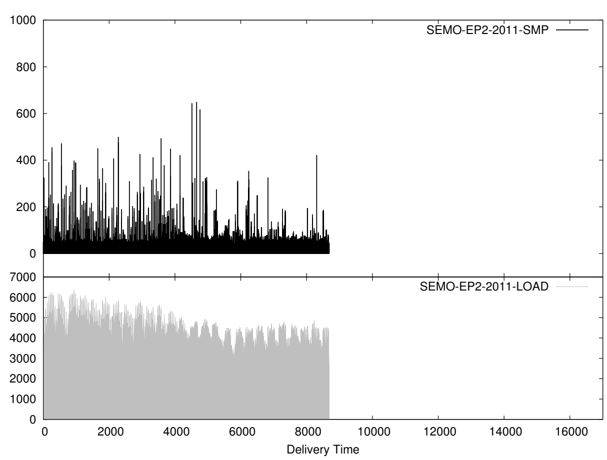
SEMO provides a web interface for public access to the historical SMP, Shadow Price and load details back to January 2008. In November 2009, SEMO started providing day-ahead half-hourly forecasts for load, SMP, Shadow Price and available wind-supply. Considering the changing price profile starting from 2009 to mid-2011 and the later availability of the forecasts (beginning 2010), we decided to use data starting January-2010 to June-2011 for training and evaluating price forecasting models. More concretely, we use year 2010 for training, three months of 2011 (January, March and May) for validation (i.e., calibrating model parameters), and another three months of 2011 (February, April and June) for testing. The choice of training, validation and test is made in order to respect the time dependency in which we train on historical data of the past and forecast prices into the future, as well as testing on months from different



(a) 2009

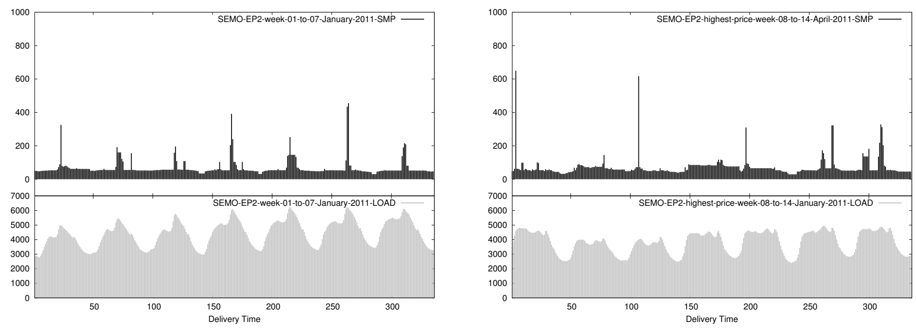


(b) 2010



(c) 2011

Fig. 1: Half-hourly price (top-black) and demand (bottom-gray) from January-2009 to June-2011. The X axis represents the delivery time (every half-hour of a trading day). The Y axis for the top-black plot represents the SMP in €/MWh and for the bottom-gray plot, the Load in MWh.



(a) First week of 2011. (b) Highest price week of 2011.

Fig. 2: Half-hourly price (top-black) and demand (bottom-gray) for two weeks in 2011.

seasons to avoid the bias of forecasting prices for summer or winter months exclusively (since prices in the winter tend to be more volatile than prices in the summer). We also paid attention to the detail that in the Irish electricity market, the actual values for SMP, load, etc., are made available only four days after the tradeday, which means that for prediction we can only use historical data with a gap of four days back into the past from the current day. All evaluations of our models and comparisons to the SEMO price forecasts are done on the three test months that are not otherwise used in any way during training or validation.

We began our analysis with simple models, using only the historical SMP for predicting the SMP of the next tradeday. We then gradually introduced information about the Shadow Price, load and wind-generation and studied the effect of each of these new variables on the prediction quality. Additionally, we investigated the impact of weather forecasts and calendar information (weekend, bank or school holidays) on the quality of the models. In order to estimate the expected supply, we have extracted information on the daily generator bids and planned generator outages available from SEMO and Eirgrid [24, 11]. Information about demand and supply is important since price peaks are typically an effect of the mismatch between high load and low supply. Our data integrity checks revealed missing days/hours in the original SEMO data. We have filled in the missing half-hours by taking the data of the closest half-hour. The data collected was available in different granularity (e.g., wind-supply obtained from Eirgrid was sampled every 15 mins) and units (Eirgrid wind-supply was in MW vs SEMO in MWh); we aggregated it to half-hourly granularity and converted the data to same units (MWh). Since we rely on SEMO forecasts for building our models, we estimated the SEMO forecast quality for each of the variables involved: load, wind, shadow price, and SMP. Our evaluation on the training set showed that in terms of forecast quality, the load forecast is most reliable, followed by shadow price, SMP and wind. In our models we use local forecast-quality-estimates as additional features. All the data collected is available online in csv files.[2].

3.2 Machine Learning Models

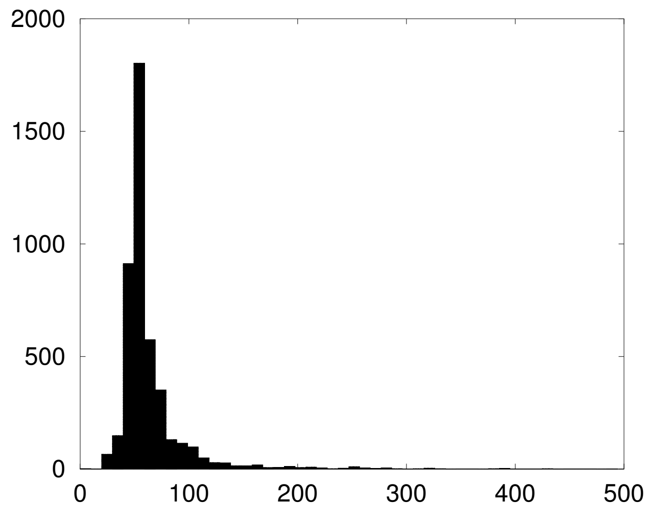
We have investigated a range of regression models (e.g., linear regression, linear SVM, various kernel-SVM), as well as analysed a variety of features and feature combinations, and present here the two best approaches.

Model 1: Predicting the SMP using historical and forecasted SMP, Shadow Price, Load and Supply. This approach follows the classical line of price prediction in international electricity markets where the main idea is to use historical data, e.g., past prices, load, and supply, for training a price model for the next trade-day. Since machine-learning models were shown to outperform other techniques such as time series models [3, 23, 29], we focus on non-linear regression techniques for building price forecasting models (e.g., Support Vector Machines).

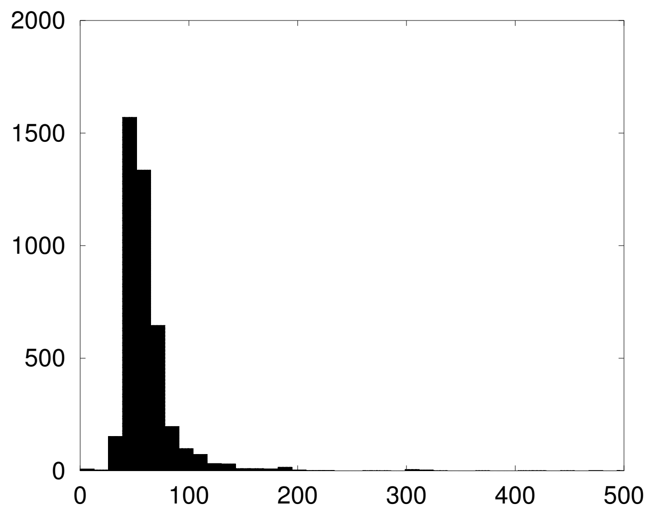
From the time series data, we extract regression vectors as follows. For each half-hour of a tradeday, we take the actual SMP as a prediction target and use historical data for the same half-hour in the past as features. For example, if the SMP on 1st of January 2010, 7 AM, is €31.04/MWh, we take this as a learning target and the SMP at 7AM of D past days as features (in this case the most recent historical data is from 27 December 2009, due to the 4 days gap). The number of historical days D is a parameter of the model and is calibrated using the validation dataset.

Since we also have access to day-ahead forecasts of SMP, shadow price, load, wind and other-supply, we study those as additional features. We have additionally investigated weather and calendar information as features, but these have not increased the quality of the model. This may happen since calendar and weather data is already factored into the load and wind-supply forecast, thus it does not add new information to the model. We compute estimates of the weekly and daily available supply from the information on outages and generator bids publicly available from [24, 11]. From Eirgrid, we use information on planned outages to estimate the weekly maximum available supply based on the maximum capacity of the available generating plants. From the day-ahead generator bids, we extract features on daily available supply. For example, we set thresholds on the maximum bid price (e.g., €40) in order to obtain estimates of expected *cheap* supply. The maximum price thresholds of bids are set at €40, €50 and €60, based on the bids and empirical SMP distribution on the validation set (Figure 3). Once the data required for preparing features is processed, we scale all features and use an SVM with an `rbf` kernel for learning. We use the LIBSVM package [9] widely accepted as state-of-the-art for SVM implementations.

[2] `http://4c.ucc.ie/~gifrim/Irish-electricity-market`



(a) Validation set.



(b) Test set.

Fig. 3: Empirical SMP distribution in our datasets. The X axis represents the SMP truncated at €500. The Y axis represents the frequency bins for the SMP values.

Model 2: Predicting the SMP using the local average-SMP and a learned difference-from-average-model. This approach moves away from the more traditional Model 1, presented above, and builds on the following observation: the actual historical SMP is a good indicator for the *average* electricity price at a given half-hour, but does not capture the particular behaviour of a given day in terms of the magnitude of the SMP peaks and valleys. It may be that due to the particular features of the next tradeday (e.g., strong wind, lower load, enough cheap supply), the SMP stays more or less flat, without exceptional peaks or valleys, thus using the local average SMP itself as an estimate is not sufficient for good prediction quality. Nevertheless, we can estimate the characteristics of the next tradeday using the publicly available forecasts. Hereby, we propose computing the final SMP as a sum of a locally computed average-SMP (e.g., over the last $D = 7$ days) and a learned SMP-difference from the average, estimated from the training set, capturing whether the SMP is going up or down with respect to the average. For example, for forecasting the SMP on 1st of January 2010, 7AM (equal to €31.04), we use the local average-SMP (equal to €29.57) over the most recent seven days, respecting the four day gap explained above, as a first component. The second component is the learned-difference between the actual SMP and the average-SMP. As learning features we use the difference between the forecasted tradeday characteristics (load, wind-supply, shadow price) from their local averages. Intuitively, lower than average load and higher than average wind, should trigger a decrease in price, thus a negative difference of SMP from the average. A regression model can estimate the SMP-difference from the differences of its features. We finally compute the SMP as the sum of the local average and the predicted difference for the SMP.

Since this model relies heavily on the quality of the forecasts, for each forecasted variable used in the model, e.g., load, wind, smp, shadow price, we have an additional feature capturing the local quality of that forecast (i.e., we measure over the past week the MSE between true and forecasted value for load, wind, etc.). There is a large body of research literature on using uncertain data for learning [1, 2, 28]. At the moment we use this simple approach for dealing with forecast uncertainty, and we plan to investigate more advanced approaches in our future work. Our experiments show that even this simple approach of integrating variable uncertainty leads to considerable model improvement.

Evaluation. We use the Mean Squared Error (MSE) as a primary means for evaluating the quality of price forecasts. This is a classical measure of both bias and variance of the models [20]. It typically penalizes gross over or underestimates of the actual values. Additionally we show the Mean Absolute Error (MAE) and the Skill-Score. The Skill-Score indicates the fractional improvement in the MSE over a reference model [20]. We use SEMO (the market operator's forecast) as a reference model, and show the Skill-Score for our forecasts. The parameters of our models are optimized for minimizing the MSE, and can be calibrated for any quality measure the user finds fit.

Table 2: Half-hourly price forecasts: error rates for SEMO and our forecasts FM1 and FM2 with respect to the true price.

Method	MAE	**MSE**	Skill-Score
SEMO	12.64	**1086.25**	NA
FM1	11.14	**821.01**	0.24
FM2	11.21	**781.72**	0.28

Table 2 shows the evaluation of our price forecasts (FM1 and FM2), using the market operator's forecast (SEMO) as a strong[3] baseline. The two forecasting models we proposed show between 24-28% improvement over the MSE of the SEMO price forecast.

Paired t-tests on the MSE and MAE (at 0.95 confidence level) show our price forecasts are statistically significantly better than those of SEMO. Table 3 gives details of the confidence interval of the MSE for all three forecasts.

Table 3: t-tests details for the three price forecasts. Each row is a baseline. Each column (SEMO, FM1, FM2) is compared against it. The upper (U) and lower (L) limits of the 95% confidence intervals are shown.

Baseline Price		SEMO	FM1	FM2
Actual	L	761.8	513.5	486.9
	U	1410.7	1128.4	1076.4
SEMO	L	-	172.4	209.7
	U	-	358.0	399.3
FM1	L	-	-	11.5
	U	-	-	66.9

So far we have focused on developing and analyzing electricity price forecasts with respect to classical error measures (e.g., MSE, MAE). Next, we focus on the effect of these forecasts on price-aware scheduling.

4 Price-Aware Scheduling Model

To test the quality of the price forecasts on a realistic scheduling problem, we adapted a variant of the feedmill scheduling problem from [25]. The schedule is generated from orders on the current day for delivery in the next morning. Tasks i are scheduled on four disjunctive press lines with their allocated machine m_i, duration d_i, power requirement p_i and due date e_i, satisfying an overall power limit l_t at each time point t. We express the problem as a mixed integer programming (MIP) minimization problem following [15], where the main decision variables are 0/1 integers x_{it} indicating whether task i starts at time t, and

[3] For building price forecasts, SEMO uses more data than what is publicly available.

non-negative, continuous variables pr_t denoting the power use at time t. For this evaluation we choose the MIP formulation over a more conventional constraint programming model, as it allows us to find the optimal solutions for the core problem. The objective function is based on the predicted price v_t, while the evaluation of the quality uses the actual price a_t. This corresponds to a scenario whereby a forecast price is available 24 hours in advance, but the price paid will be the actual market price. As executing the schedule requires significant preparation work, we cannot continuously reschedule based on the current, actual price. Therefore, the energy cost of a schedule is computed as:

$$\text{cost} = \sum_t pr_t^* a_t \tag{1}$$

where pr_t^* is the profile value at time t of the optimal solution to the following MIP problem:

$$\min \sum_t pr_t v_t \tag{2}$$

subject to:

$$\forall_i : \quad \textstyle\sum_t x_{it} = 1 \tag{3}$$

$$\forall_t : \textstyle\sum_i \sum_{t-d_i+1 \le t' \le t} p_i x_{it'} = pr_t \le l_t \tag{4}$$

$$\forall_m \forall_t : \quad \textstyle\sum_{i|m_i=m} \sum_{t-d_i+1 \le t' \le t} x_{it'} \le 1 \tag{5}$$

$$\forall_i \forall_{t|t+d_i>e_i} : \quad x_{it} = 0 \tag{6}$$

We generated problem instances randomly, filling each production line to capacity for 24 hours, choosing random durations uniformly between 25 and 100 minutes, the last task generated will be truncated to fit into 24 hours, and power requirements uniformly chosen between 0 and 200 kW. The time resolution was set to 5 minutes, so that optimal solutions could be found within a 10 minute timelimit. For each prediction day, from the same test period used for evaluating the price forecasts, we generated 10 samples, in total 880 runs. For each instance we computed the actual cost based on an optimal solution for the actual price, for the SEMO forecast and for our two forecasts (FM1, FM2). The schedule based on the actual price provides a lower bound, but since the actual price is not known in advance, it is not realizable.

We also experimented with another scenario, where each production line will be busy for only 12 hours. This allows us to avoid the peak price periods completely, which presents a much less challenging problem. The schedule overhead decreases accordingly, and the results are similar to the ones presented here.

Table 4 shows summary results over all sample runs. It provides statistics of the scheduling cost for the different price forecasts. We ran paired t-tests to assess which price forecasts lead to significantly cheaper scheduling. We found that the costs using all the three forecasts are very close to the optimal cost (within 5-10%), and that neither FM1 or FM2 forecasts were significantly better than SEMO. In fact, the price forecast with best MSE (FM2, as shown in Table 2) was significantly worse than the other two with respect to the energy cost of

Table 4: Summary Results of Price-Aware Schedule Costs

Price	Min	Median	Mean	Max
Actual	4,383,718	5,934,654	6,093,365	9,805,821
SEMO	4,507,136	6,054,220	6,272,768	10,218,804
FM1	4,499,811	6,058,093	6,266,800	10,070,541
FM2	4,570,552	6,094,818	6,283,261	10,059,264

the optimal schedule. Table 5 shows the confidence intervals for the average difference between the optimal schedule cost and the costs obtained using the forecasts. This was a somewhat surprising result: the best numeric forecasting model (as judged with respect to classical learning measures, e.g., MSE), was the worst model with respect to scheduling cost. We analyse in the following section what is the correct approach to significantly reduce the energy-costs in a scheduling context.

Table 5: Confidence intervals (95%) for price-aware scheduling costs comparing optimal solutions priced using actual price, and each of the three forecasts (SEMO, FM1, FM2). Baseline for comparison is the method identified on each row.

Price		SEMO	FM1	FM2
Actual	L	$-200,564.9$	$-193,646.7$	$-211,094.4$
	U	$-158,241.3$	$-153,222.5$	$-168,697.4$
SEMO	L	-	$-1,506.1$	$-17,262.6$
	U	-	$13,443.1$	$-3,722.9$
FM1	L	-	-	$-23,968.3$
	U	-	-	$-8,954.2$

5 Properties of Energy-Price Forecasts for Scheduling

We analyze here the key properties of price forecasts that positively affect cost-aware scheduling. What seems to matter most for scheduling is that the forecasting model captures the price-trend rather than the exact real value (as measured by MSE). Therefore, forecasting models that capture well the peaks and valleys of the energy price have better behaviour when used for scheduling. To study this hypothesis, we analysed the three previous forecasts in a classification framework, where prices belong to one of two classes: *p*eak or *l*ow price. The class is decided using a threshold inferred from the empirical distribution of the price on the validation set. We set the threshold at the 66th percentile (about €60, see Figure 3). Thus, if a price is above the threshold, it is in the peak class, otherwise it is in the low class.

Table 6: Confidence intervals for scheduling-costs of price forecasts with increasingly better peak-price classification accuracy. FP-82% stands for forecast obtained by correcting the false positive error, to obtain a forecast with 82% classification accuracy. Statistically significant improvements are highlighted in bold.

Price	SEMO-78%	FP-82%	FP-86%
SEMO-78% L	-	**19,610.6**	**29,274.0**
SEMO-78% U	-	**28,795.2**	**39,735.0**
FP-82% L	-	-	**7,388.3**
FP-82% U	-	-	**13,214.8**

Table 7: Confidence intervals for scheduling-costs of price forecasts with increasingly better peak-price classification accuracy. FN-82% stands for forecast obtained by correcting the false negative error, to obtain a forecast with 82% classification accuracy. Statistically significant improvements are highlighted in bold.

Price	SEMO-78%	FN-82%	FN-86%
SEMO-78% L	-	**30,223.1**	**46,582.4**
SEMO-78% U	-	**66,446.1**	**86,151.5**
FN-82% L	-	-	**9,604.2**
FN-82% U	-	-	**26,460.4**

Analyzed in this context, all three forecasts have similar classification accuracy, about 78%. This could explain the lack of difference with respect to effect on the energy cost of the schedule. To further study this hypothesis, we performed the following experiment. Starting from the SEMO price forecast (with 78% classification accuracy), we artificially obtained better peak-price classifiers by correcting the classification error. There are two types of classification error: false positives (missing lows) and false negatives (missing peaks), and we believe that for scheduling it is more important to reduce the false negatives, than the false positives. To test this, we have first corrected 50% of the false positives (if the classifier predicts *peak*, but the truth is *low*, we replace the SEMO price with the true price), and then 100% of the false positives, to obtain two classifiers with 82% and 86% classification accuracy and their associated price forecasts. Similarly, starting from SEMO's forecast, we have corrected the same number of errors as before, but this time from the false negatives, to obtain two more classifiers with 82% and 86% accuracy. Finally, we have used SEMO versus the four improved forecasts for cost-aware scheduling. The results show that improved classification accuracy leads to reduced scheduling cost, and that the type of classification error matters, with false negatives having more impact on scheduling. Paired t-tests on the scheduling costs obtained with the different forecasts showed that the improvement is statistically significant with very high confidence (higher than 0.99). The schedule-cost improvement over SEMO for the first two artificially improved forecasts is between 0.4-0.5%, and for the second two between 0.8-1.0%. Given that SEMO was already within 5% of the optimum schedule cost, this is a significant improvement. Tables 6 and 7 give detailed results.

In conclusion, we show that developing good regression techniques where quality is evaluated using traditional learning measures is not effective for cost-aware scheduling. We have also tested peak-price classification models trained similarly to [29], but the thresholded regression models had slightly better classification accuracy. We believe one needs to rather focus on various types of *cost-sensitive* peak-price classifiers and their impact on scheduling cost, where ideally learning should inform scheduling decisions and scheduling should inform learning decisions (i.e., the classifications costs have to be motivated by the scheduling application).

6 Conclusion

We have shown that using classical price-prediction features and machine learning techniques, one can obtain better price forecasts with respect to classical error measures (e.g., MSE, MAE). When plugging the improved price forecasts into cost-aware scheduling, we nevertheless do not observe the same benefit on the schedule-cost. This suggests that scheduling requires specific features from the price forecasts. This paper focuses on pruning the large space of learning strategies to identify the most promising models for designing energy-efficient schedules. We have shown that good peak-price classification behaviour is an important model property and that the type of classification error directly affects cost-aware scheduling. This opens new research directions towards designing cost-sensitive price classification forecasts for scheduling.

Our scheduling experiments also give some insights into the usefulness of alternative tariff models. Besides a time-variable tariff based on the actual market price, tariffs based on long- or short-term price prediction have been proposed under the term *time-of-use* tariffs. In this case the customer knows in advance the price to be paid, and the provider carries the risk of a wrong prediction. In our scheduling problem, the value of the optimal solution based on the forecast will then give the final cost of the schedule. This optimal value can be above or below the cost based on the actual prices, but shows much higher variability as compared to the cost using the actual market price. From our preliminary experiments, it therefore seems preferable to use tarrifs based on the actual price. We plan to investigate this further in our future work.

Acknowledgements

This work was funded by Science Foundation Ireland under Grant 10/IN.1/I3032.

References

1. C. Aggarwal and P. Yu. A survey of uncertain data algorithms and applications. *IEEE Transactions on Knowledge and Data Engineering*, 21(5):609 –623, may 2009.
2. C. C. Aggarwal. On multidimensional sharpening of uncertain data. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pages 373–384, 2010.
3. S. K. Aggarwal, L. M. Saini, and A. Kumar. Day-ahead price forecasting in ontario electricity market using variable-segmented support vector machine-based model. *Electric Power Components and Systems*, 37(5):495–516, 2009.
4. S. K. Aggarwal, L. M. Saini, and A. Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power and Energy Systems*, 31(1):13 – 22, 2009.
5. D. Aikema, C. Kiddle, and R. Simmonds. Energy-cost-aware scheduling of hpc workloads. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011*, pages 1 –7, june 2011.
6. N. Amjady and F. Keynia. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Transactions on Power Systems*, 24(1):306 –318, feb. 2009.
7. P. Bodík, R. Griffith, C. Sutton, A. Fox, M. Jordan, and D. Patterson. Statistical machine learning makes automatic control practical for internet datacenters. In *Proceedings of the 2009 conference on Hot topics in cloud computing*, HotCloud'09, Berkeley, CA, USA, 2009. USENIX Association.
8. N. Buchbinder, N. Jain, and I. Menache. Online job-migration for reducing the electricity bill in the cloud. In *Proceedings of the International Conference on Networking (NETWORKING)*, volume 6640 of *Lecture Notes in Computer Science*, pages 172–185. Springer Berlin / Heidelberg, 2011.
9. C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, May 2011.
10. D. Connolly, H. Lund, B. Mathiesen, and M. Leahy. Modelling the existing irish energy-system to identify future energy costs and the maximum wind penetration feasible. *Energy*, 35(5):2164 – 2173, 2010.
11. EirGrid `http://www.eirgrid.com/`, 2012.
12. S. Fan, C. Mao, and L. Chen. Next-day electricity-price forecasting using a hybrid network. *Generation, Transmission Distribution, IET*, 1(1):176 –182, january 2007.
13. P. Finn, C. Fitzpatrick, D. Connolly, M. Leahy, and L. Relihan. Facilitation of renewable electricity using price based appliance control in irelands electricity market. *Energy*, 36(5):2952 – 2960, 2011.
14. B. Guenter, N. Jain, and C. Williams. Managing cost, performance, and reliability tradeoffs for energy-aware server provisioning. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pages 1332 –1340, april 2011.
15. J. Hooker. *Integrated Methods for Optimization*. Springer, New York, 2007.
16. M. Jablonska, A. Mayrhofer, and J. P. Gleeson. Stochastic simulation of the uplift process for the Irish electricity market. *Mathematics-in-Industry Case Studies Journal*, 2(pp 86-110), 2010.
17. T. Kim and H. Poor. Scheduling power consumption with price uncertainty. *IEEE Transactions on Smart Grid*, 2(3):519 –527, sept. 2011.

18. K. Le, R. Bianchini, M. Martonosi, and T. Nguyen. Cost-and energy-aware load distribution across data centers. In *Proceedings of HotPower*. Citeseer, 2009.
19. S. Mani and S. Rao. Operating cost aware scheduling model for distributed servers based on global power pricing policies. In *Proceedings of the Fourth Annual ACM Bangalore Conference*, COMPUTE '11, pages 12:1–12:8, New York, NY, USA, 2011. ACM.
20. S. Pelland, G. Galanis, and G. Kallos. Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. *Progress in Photovoltaics: Research and Applications*, 2011.
21. P. Ptak, M. Jablonska, D. Habimana, and T. Kauranne. Reliability of arma and garch models of electricity spot market prices. In *Proceedings of European Symposium on Time Series Prediction*, september 2008.
22. A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the electric bill for internet-scale systems. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, SIGCOMM '09, pages 123–134, New York, NY, USA, 2009. ACM.
23. L. Saini, S. Aggarwal, and A. Kumar. Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in national electricity market. *Generation, Transmission Distribution, IET*, 4(1):36 –49, january 2010.
24. SEMO `http://www.sem-o.com/`, 2012.
25. H. Simonis. Models for global constraint applications. *Constraints*, 12(1):63–92, 2007.
26. H. Simonis and T. Hadzic. A family of resource constraints for energy cost aware scheduling. In *Third International Workshop on Constraint Reasoning and Optimization for Computational Sustainability*, St. Andrews, Scotland,UK, September 2010.
27. S. Srikantaiah, A. Kansal, and F. Zhao. Energy aware consolidation for cloud computing. In *Proceedings of the 2008 conference on Power aware computing and systems*, HotPower'08, pages 10–10, Berkeley, CA, USA, 2008. USENIX Association.
28. J.-L. Yang and H.-X. Li. A probabilistic support vector machine for uncertain data. In *Computational Intelligence for Measurement Systems and Applications, 2009. CIMSA '09. IEEE International Conference on*, pages 163 –168, may 2009.
29. H. Zareipour, A. Janjani, H. Leung, A. Motamedi, and A. Schellenberg. Classification of Future Electricity Market Prices. *IEEE Transactions on Power Systems*, 26(1):165–173, Feb. 2011.