Optimal activity and battery scheduling algorithm using load and solar generation forecasts

Yogesh Pipada Sunil Kumar
University of Adelaide
Adelaide, Australia
yogeshpipada.sunilkumar,

Rui Yuan
University of Adelaide
Adelaide, Australia
r.yuan,

Nam Trong Dinh
University of Adelaide
Adelaide, Australia
trongnam.dinh,

S. Ali Pourmousavi
University of Adelaide
Adelaide, Australia
a.pourm@adelaide.edu.au

Abstract—Energy usage optimal scheduling has attracted great attention in the power system community, where various methodologies have been proposed. However, in real-world applications, the optimal scheduling problems require reliable energy forecasting, which is scarcely discussed as a joint solution to the scheduling problem. The 5th IEEE Computational Intelligence Society (IEEE-CIS) competition raised a practical problem of decreasing the electricity bill by scheduling building activities, where forecasting the solar energy generation and building consumption is a necessity. To solve this problem, we propose a technical sequence for tackling the solar PV and demand forecast and optimal scheduling problems, where solar generation prediction methods and an optimal university lectures scheduling algorithm are proposed.

Index Terms—Forecasting, Refined motif, Optimisation, Valley-filling scheduling, Mixed-integer linear programming (MILP)

I. INTRODUCTION

A. Motivation

Even though infrequent, peak load is a cause for increased capital and operating expenses for power networks. The reason for this is twofold; the higher need for grid reinforcements and the use of more expensive fossil fuel-based generators to satisfy the peak loads (for a short duration). Thereby to manage the extra costs due to peak demand, network operators include a peak demand tariff (commensurate to the peak load) in the electricity bills of commercial customers. This motivates large-scale commercial customers such as universities and manufacturers to manage their demand better and invest in assets such as solar photovoltaic (PV) panels and stationary batteries that have the potential to shift their demand and reduce their electricity bills. A secondary effect of this is the reduction in CO_2 emissions because of the lesser usage of fossil fuel generators, which has the potential to combat climate change and aid the process of decarbonization of our power networks.

However, optimal scheduling of (schedulable) loads and batteries (to minimize electricity costs) requires predictions of inflexible load (baseload), solar PV generation and the spot price of electricity. The first two variables majorly depend on weather conditions. Also, increasing renewable generation in the generation mix has led to increasing price volatility in the Australian national electricity market (NEM) which also creates a correlation between electricity prices and the weather [1]. This makes designing such algorithms challenging as most commercial activities are planned over the mid to long term; hence require reliable predictions. Additionally, the scheduling problems are generally mixed-integer linear programs (MILPs) which are NP-hard and may be intractable depending on the formulation and problem size.

Surrounding this premise, the IEEE Computational Intelligence Society (IEEE-CIS) partnered with Monash University (Victoria, Australia) to conduct a competition seeking technical solutions to manage Monash's micro-grid containing rooftop solar panels and stationary batteries [2]. The main challenge was to develop an optimal scheduling algorithm for Monash's lecture theatres and operation of batteries to minimize their electricity bill, considering their baseload, solar generation and NEM electricity spot prices. To this end, the contestants were provided with actual time series data of the building loads without any lecture program (baseload) and solar generation from the micro-grid. So the contestants were expected to predict the baseload and solar generation by taking into account real-world weather data (also provided) for one month in the future. Following this, the contestants had to use actual electricity spot prices for the same duration along with these predictions for the optimal scheduling algorithm.

B. Related work

We developed separate algorithms for the solar and baseload forecasts based on practical insights and the given data. Many approaches are available in the literature for both data types, a relatively mature and crowded research field. Therefore, the methods we have developed for this problem borrow themes from this mature literature.

For solar predictions, the basic theme used is similar to the "clear sky" models for forecasting solar irradiance and thereby estimating PV generation [3]. The general idea here is to create a baseline model for PV generation, assuming the sky is clear, meaning there is no cloud coverage or temperature variance. This baseline is then modified based on actual or expected weather conditions to estimate the actual PV generation. The literature around this idea is based on physical models of irradiance calculations, where equations are used to develop the baseline for a given geographical location [3]. Newer methods use data-driven techniques such as time series forecasting and machine learning models [4]. The data-driven

methods are gaining popularity because of their robustness, speed and geographic adaptability.

Intuitively, this method gives reliable solar forecasts because of solar generation's seasonal and diurnal nature. However, the main drawback is the lack of data specifically related to "clear sky" days; Consequently, we use the most commonly occurring day's generation as baseline, which can be discovered by our previous work [5].

For the baseload forecasting, the theme used is the application of ensemble methods, as the current state of the art identifies these methods to be most accurate when compared to stand-alone methods [6]. We use mainly a combination of random forest (RF), gradient boost (GB), autoregressive integrated moving average (ARIMA) and support vector machine (SVM), which are all well-studied and standard methods for time series forecasting [7]. Specifically, we apply different forecasting methods to different sub-series after disaggregating the load profiles using Seasonal and Trend decomposition using Loess (STL) due to their cyclic patterns [8].

The optimal scheduling of distributed energy resources (DER) and controllable/uncontrollable loads for micro-grids comes under the class of problems commonly referred to as energy management problems. However, these classes of problems tend to be non-convex and non-linear in nature because of the constraints associated with the scheduled devices, e.g., scheduling uninterruptible loads such as lectures that cannot be stopped once started. As reviewed by the authors in [9], the majority of these problems are modeled via classical optimization approaches such as MILP [10] and mixed-integer non-linear programs (MINLPs) [11]. Other approaches include dynamic programming [12], rule-based optimization and meta-heuristic algorithms [13].

Since these problems tend to be non-convex and non-linear, the major challenge associated with solving these problems is tractability and (in the case of meta-heuristic methods) convergence to a global optimum. The mathematical literature surrounding MILPs offers a variety of techniques to simplify these problems before solving them, and the availability of solvers such as Gurobi® for MILPs make them a very attractive option for solving these problems. Also, unlike other models where the algorithms (may) converge to a local optimal solution, modern MILP solvers can obtain globally optimal solutions for this class of problems.

C. Contributions

Based on the related work studied, this paper offers the following contributions to the body of knowledge:

- A solar forecasting algorithm using training data from refined motif (RM) discovery technique and use of an over-parameterized 1D-convoluted neural network (1D-CNN) implemented via residual networks (ResNet). To overcome the lack of "clear sky" data, we incorporated the work done previously by Rui Yuan et al. [5] to identify RMs in the given solar generation dataset. An RM is the most repetitive pattern within a given time series, which can be extracted along with the exogenous variables

(e.g., weather information) associated with this pattern. Using this as a baseline, we estimated solar generation by training an over-parameterized 1D-CNN. Some studies have shown that over-parameterization of CNNs can lead to better performance at the expense of longer training time [14]–[16]. Therefore, a ResNet was implemented to develop a deeper NN but with faster computation time.

- An optimal micro-grid scheduling algorithm is solved based on real-world data for a university-based application. To the best of our knowledge, there has not yet been a study to co-optimize lectures schedule (and associated resources) and battery operation (with PV panels). Therefore, in this paper, we have developed and tested our algorithm using real-world data and practical case instances provided by Monash University. Given the large problem size (one-month schedule) and the presence of a quadratic term in the objective function for the cost of peak demand, we proposed a two sub-problem approach to formulate a tractable problem. The first sub-problem was used to limit the peak demand throughout the month, eliminating the quadratic term from the objective. Then, the peak demand was used to solve the second sub-problem to minimize the total electricity costs.

The rest of the paper is structured as follows. Section II introduces the data, information and problem requirements. Section III proposes the forecasting and optimization methodologies. Section IV presents the numerical results of the scheduling algorithm based on time series forecasting. Finally, we conclude the paper section V.

II. Background Information

This section describes the competition requirements and data provided by the organizers.

A. Problem statement

This section describes the scheduling constraints and objective function provided by the competition organizers and has been adapted from [2]. We were required to develop prediction algorithms for the baseload of six buildings in Monash University and the solar generation of PV panels connected to them. After this, we had to use these predictions to optimally schedule lecture activities and battery operation while minimizing the electricity costs (i.e., electrical energy consumption cost plus peak demand charge). We were allowed to consider the electricity prices as known parameters to simplify the problem and focus on predicting solar generation and baseload.

For each lecture activity, we are provided with several small or large rooms needed, the electrical power consumed per room and the duration of the activities (in steps of 15 minutes). We are also provided with a list of precedence activities, i.e., activities that must be performed at least one day before the activity in question. For the batteries, we are provided with the maximum energy rating, i.e., state of charge (SOC), the peak charge/discharge power and charge/discharge efficiency.

The scheduling must start from the first Monday of the month, and each activity must; 1) take place within office hours (9:00–17:00), 2) happen once every week and recur on the same day and time every week. Also, the number of rooms (associated with each activity) allocated at a given time interval must not exceed the total number of rooms in the six buildings. The batteries can operate in one of three states; charging, discharging or idle, which is determined by the scheduling algorithm. However, the battery power must be at the peak charge/discharge capacity once operated.

B. Dataset information

The organizers provided us with datasets for the electricity consumption of the six buildings and their connected solar panels [2]. About two years' worth of data was provided with a granularity of 15 minutes. However, the datasets contained many missing points, especially the demand profiles with consecutive weeks of unrecorded data, as shown in [17]. Information regarding COVID-19 lockdown and how it influenced the forecast and earlier timelines were also provided to aid in incorporating the effect of COVID-19 pandemic on baseload in our models. ERA5 climate data was also provided as exogenous variables for the time series forecasting via a partnership with OikoLabs [2]. Also, we were allowed to use electricity price information from the Australian NEM [18] as a parameter for objective function calculation. The input parameters used for solar and building forecasting are summarized in Table I.

TABLE I
INPUT DATA USED FOR FORECASTING

Weather inputs (15mins) / Forecast outputs (15mins)	Building (15mins)	Solar (15mins)
Temperature (°C)	✓	✓
Dewpoint temperature (°C)	✓	
Wind speed (m/s)		✓
Relative humidity (0-1)	✓	✓
Surface solar radiation (W/m^2)	✓	✓
Total cloud cover (0-1)		✓
Occupancy (0-1)	✓	
Annual harmonics (0-1)		✓

C. Instance information

To test the algorithms each team developed, the competition organisers provided ten problem instances. Each instance provided information regarding the number of activities to be scheduled and the specifics of each activity; duration of the activity in time steps, power consumed per room for each time interval, number of large/small rooms required and a precedence list. The precedence list contains the activities that must happen one day before the specified task. Depending on the number of activities in each instance, they were divided into two sub-categories; large (200 activities) and small (50 activities) instances. We report simulation results for one large and one small instances in section IV.

III. METHODOLOGY

This section describes the methodology developed to solve the problem described in II-A. Mainly, we outline our data cleaning strategies, forecasting methodology and scheduling algorithms used in the competition.
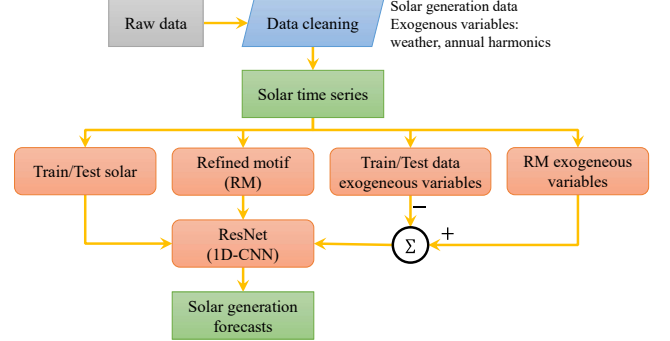
A. Solar generation forecast

Fig. 1. Block diagram of the proposed solar generation prediction model

The solar generation forecast methodology is depicted in Fig. 1. The solar data was mainly clean, with no outliers and a few missing points, which were removed during the cleaning process. Since solar time series data is repetitive in nature, we used the method described in [5] to extract the RM (most repetitive patterns) and its associated weather conditions. The intuitive idea is to train a neural network (NN) to reshape this RM with respect to the difference in weather conditions between the RM weather conditions and the actual data weather conditions.
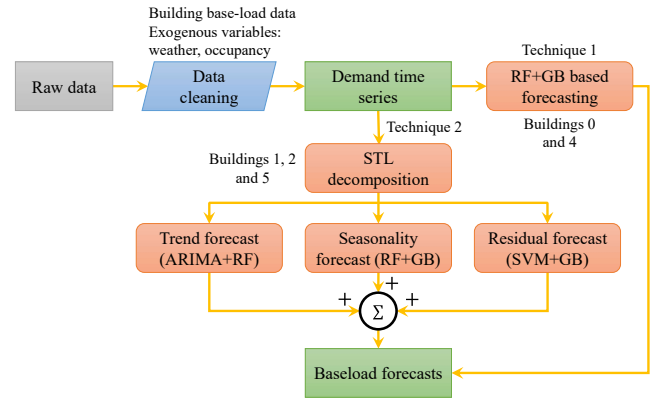
B. Baseload forecast

Fig. 2. Block diagram of the proposed baseload prediction model

The block diagram of the methodology used for baseload forecasting is shown in Fig. 2, which demonstrates the use of different ensemble forecasting methods for different buildings. We applied the same procedure for the data cleaning process as in the PV generation forecast. However, since there were more missing points in the demand profiles, which lasted for many weeks consecutively, we used the average value

from the previous week for the long-period missing data. The choice of ensemble methods was achieved through the use of the Python sklearn package's voting regressor function [19]. For this dataset, this resulted in a combination of RF, SVM, ARIMA and GB methods for best accuracy. In terms of the length of training data, due to the highly dynamic nature of demand profiles, which was exacerbated during the COVID-19 period, we only used 2 months of historical data for training.

Due to the seasonal/cyclic patterns demonstrated in building 1, 2 and 5, as shown in [17], we utilized STL decomposition to capture these properties by disaggregating the load profiles into three components, namely trend, seasonality and residual. We then trained each profile with different methods depending on the performance of the historical data. In the end, the predictions from these three sub-series were combined to obtain the final building demand forecast. For buildings 0 and 4 with no clear cyclic patterns, we observed that either RF or GB (without STL decomposition) were sufficient to capture their repetitive nature. Lastly, for building 3, historical data showed that the load only took on a finite set of discrete values; thus, we fixed the predictions to the median.

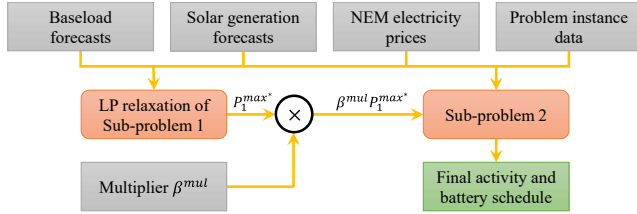### C. Optimal scheduling algorithm

Fig. 3. Block diagram of the proposed optimal scheduling algorithm

We developed a MIP-based optimal scheduling problem to represent all the scheduling constraints, where the objective function was provided by the organizers as follows:

$$O(P_t) = \frac{0.25}{1000} \sum_{t \in \mathcal{T}} P_t \lambda_t^{\text{aemo}} + 0.005 \left(\max_{t \in \mathcal{T}} P_t \right)^2 \quad (1)$$

Here P_t and λ_t^{aemo} are the net demand and AEMO electricity spot price at time t over the time horizon $\mathcal{T}$ respectively. Therefore, it can be seen that the problem is a MIQP problem, which is NP-hard and hence most likely intractable given the size of this problem ($\|\mathcal{T}\| = 2880$). Hence, to improve the tractability, we reformulated the MIQP problem into two MILP sub-problems, as demonstrated in Fig. 3. The first sub-problem is used to estimate an upper bound for the maximum demand ($P_1^{\text{max}} = \max_{t \in \mathcal{T}} P_t$), which is used to remove the quadratic term from the objective function in the second sub-problem.

The scheduling is developed for a time duration Δt with P_t^{base} and P_t^{solar} being the baseload and solar generation forecasts, respectively. For a set of batteries $b \in \mathcal{B}$ over the time horizon $t \in \mathcal{T}$, we define $E_{b,t}^{\text{bat}}$ as battery SOC in kWh, $P_{b,t}^{\text{bat}}$ is the battery power in kW, E_b^{cap} is the battery capacity in kWh, $B_b^{\text{c}}/B_b^{\text{d}}$ is the battery charge and discharge power in kW, $u_{b,t}^{\text{c}}/u_{b,t}^{\text{d}}$ are the battery charge/discharge status, and $\eta_{b,t}^{\text{c}}/\eta_{b,t}^{\text{d}}$

are the charging/discharging efficiencies. For a set of activities $a \in \mathcal{A}$ over the same time horizon, we define A_a^{kW} as the power required per room for the activity in kW, $R_a^{\text{large}}/R_a^{\text{small}}$ are the number of large/small rooms needed for the activity, P_t^{sched} be the total power scheduled in kW, and $u_{a,t}^{\text{start}}/u_{a,t}^{\text{active}}$ are binary variables which track the activity start time and active time, respectively. $\mathcal{D}_a$ is the duration set for the activity, and $\mathcal{P}_a$ is the set of activities that must precede activity a. Additionally, we define $\mathcal{T}_n$ as the set of time intervals where activities must not be active, $\mathcal{T}_f$ is the set of time intervals for the first week and L_t^{day} corresponds to the day of the time interval t. We present the equations used for the two sub-problems in the below sections.

1) Sub-problem 1:

$$\min_{\psi_1} P_1^{\text{max}} \quad (2a)$$

Subject to,

$$0 \le P_t \le P_1^{\text{max}} \quad \forall t \in \mathcal{T} \quad (2b)$$

$$0 \le E_{b,t}^{\text{bat}} \le E_b^{\text{cap}} \quad \forall t \in \mathcal{T}, b \in \mathcal{B} \quad (2c)$$

$$u_{b,t}^{\text{c}} + u_{b,t}^{\text{d}} \le 1 \quad \forall t \in \mathcal{T}, b \in \mathcal{B} \quad (2d)$$

$$P_{b,t}^{\text{bat}} = u_{b,t}^{\text{c}} \frac{B_b^{\text{c}}}{\sqrt{\eta_b^{\text{c}}}} - u_{b,t}^{\text{d}} B_b^{\text{d}} \sqrt{\eta_b^{\text{d}}} \quad \forall t \in \mathcal{T}, b \in \mathcal{B} \quad (2e)$$

$$E_{b,-1}^{\text{bat}} = E_b^{\text{cap}} \quad \forall b \in \mathcal{B} \quad (2f)$$

$$E_{b,t}^{\text{bat}} = E_{b,t-1}^{\text{bat}} + \left(u_{b,t}^{\text{c}} B_b^{\text{c}} - u_{b,t}^{\text{d}} B_b^{\text{d}} \right) \Delta t \quad \forall t \in \mathcal{T}, b \in \mathcal{B} \quad (2g)$$

$$P_t^{\text{sched}} = \sum_{a \in \mathcal{A}} u_{a,t}^{\text{active}} A_a^{kW} \left(R_a^{\text{small}} + R_a^{\text{large}} \right) \quad \forall t \in \mathcal{T} \quad (2h)$$

$$P_t = P_t^{\text{base}} - P_t^{\text{solar}} + P_t^{\text{sched}} + \sum_{b \in \mathcal{B}} P_b^{\text{bat}} \quad \forall t \in \mathcal{T} \quad (2i)$$

$$\sum_{a \in \mathcal{A}} u_{a,t}^{\text{active}} R_a^{\text{large}} \le H^{\text{large}}; \sum_{a \in \mathcal{A}} u_{a,t}^{\text{active}} R_a^{\text{small}} \le H^{\text{small}} \ \forall t \in \mathcal{T} \quad (2j)$$

$$u_{a,t}^{\text{active}} = 0; u_{a,t}^{\text{start}} = 0 \quad \forall t \in \mathcal{T}_n, a \in \mathcal{A} \quad (2k)$$

$$\sum_{t \in \mathcal{T}_f} u_{a,t}^{\text{start}} = 1 \quad \forall a \in \mathcal{A} \quad (2l)$$

$$u_{a,t}^{\text{start}} \le u_{a,t+k}^{\text{active}} \ \forall t \in \{0, 1, \dots, |\mathcal{T}| - |\mathcal{D}_a| - 1\}, k \in \mathcal{D}_a, a \in \mathcal{A} \quad (2m)$$

$$\sum_{t \in \mathcal{T}_f} u_{a,t}^{\text{active}} = |\mathcal{D}_a| \quad \forall a \in \mathcal{A} \quad (2n)$$

$$\sum_{t \in \mathcal{T}_f} \left(u_{a,t}^{\text{active}} L_t^{\text{day}} - u_{k,t}^{\text{active}} L_t^{\text{day}} \right) \ge 1 \ \forall k \in \mathcal{P}_a, a \in \mathcal{A} \quad (2o)$$

$$u_{a,t}^{\text{start}} = u_{a,t-\alpha}^{\text{start}} \quad \forall t \in \mathcal{T}_f^c, a \in \mathcal{A} \quad (2p)$$

where $\psi_1 = \{P, P^{\text{bat}}, P_1^{\text{max}}, E^{\text{bat}}, u^{\text{c}}, u^{\text{d}}, u^{\text{active}}, u^{\text{start}}\}$. Equation (2b) helps find the peak demand of the schedule and prevents energy export to the grid. Equations (2c)–(2g) represent battery related constraints. Equations (2h)–(2i) are used to calculate the net energy imported from the grid, and

TABLE II
BASELOAD AND PV GENERATION PREDICTION NRMSE

Building #	Proposed	BI	Solar #	Proposed	BI
0	**0.43**	0.43	0	**0.72**	0.81
1	**0.31**	0.34	1	**0.77**	0.77
2	**0.23**	0.24	2	**0.74**	0.79
3	**0.32**	0.32	3	**0.74**	0.75
4	**0.40**	0.40	4	**0.78**	0.79
5	**0.13**	0.14	5	**0.72**	0.75

equation (2j) ensures the number of rooms engaged through activities is within the limit. Equation (2k) ensures activities do not start outside office hours and before the first Monday of the month, equation (2l) ensures activities happen once a week, equations (2m)–(2n) ensure activities are active for the necessary duration after starting, equation (2o) satisfies that all precedence (prerequisite) activities happen a day before and equation (2p) ensures activities recur at the same time every week. The objective here is to minimize peak demand while satisfying all the scheduling constraints.

Since this problem is only used to solve for the upper bound maximum demand P_1^{max}, it is not necessary to solve the full problem. We use the property that the solution of the LP relaxation of a minimization-based MILP problem is its lower bound ($y_{\text{MILP}}^* \geq y_{\text{LP}}^*$) [20] to just solve the LP relaxation and use this for sub-problem 2, thereby improving tractability of our method.

2) Sub-problem 2:

$$\min_{\psi_2} \frac{0.25}{1000} \sum_{t \in \mathcal{T}} P_t \, \lambda_t^{\text{aemo}} \qquad (3a)$$

Subject to, Equations (2c)–(2o)

$$0 \leq P_t \leq \beta^{\text{mul}} P_1^{\text{max}^*} \qquad (3b)$$

where $\psi_2 = (\psi_1 \setminus \{P_1^{\text{max}}\})$. Using equation (3b), we ensure that peak demand does not exceed $\beta^{\text{mul}} P_1^{\text{max}^*}$ limiting the demand charge. Intuitively, $\beta^{\text{mul}} \geq 1$ is a multiplier used to change the tightness of the MILP, creating a trade-off between computation time and electricity cost; higher values of β^{mul} decreases computation speed but increases the electricity cost (objective of the problem) and vice versa.

IV. RESULTS AND DISCUSSION

The simulations were run for November 2020 as per the competition stipulation with $\Delta t = 15$ min, $\|\mathcal{T}\| = 2880$ and other instance information as specified in II-C. For the scheduling, two β^{mul} values were chosen; 1.10 for small instances and 1.15 for large instances. These values were obtained via trial and error. The prediction and optimization algorithms were run in Python 3.8.8 using Gurobi® solver.

Since each of the twelve datasets has different scales, we decide to use the normalised root mean square error (NRMSE), which has been used to facilitate the comparison among time series with varying magnitudes [21]. The metric is calculated as follows:

$$\text{NRMSE} = \sqrt{\sum_t \frac{\hat{y}_t - y_t}{\|\mathcal{T}\|}} \times \frac{1}{\bar{y}} \qquad (4)$$

where $\hat{y}_t$ is the model predictions, y_t is the model actual data and $\bar{y}$ represents the mean of the actual data.

In order to demonstrate the superiority of our methods, we compare our proposed techniques with the best individual (BI) prediction models that are mentioned in section I-B. Please note that the BI models are picked from out-of-sample validation on the actual datasets using the same inputs and training length as the proposed methods. Table II displays the NRMSE of the actual data in November 2020 between the proposed and the BI models. It can be seen that the proposed models achieve lower errors in almost all the twelve profiles except for the demand of buildings 0 and 4. As mentioned in section III-B, these two profiles are predicted using either RF or GB due to their non-cyclic behavior. Therefore, they have the highest errors among all the demand profiles. For the PV generation time series, the proposed method also achieves better prediction accuracy with quite consistent NRMSE due to the high dependence of solar PV generation on weather data.

Fig. 4(a) and 4(b) are used to explain the intuition behind the RM method for the PV generation; the former shows the difference between RM and a predicted generation weather condition for November 4th, 2020 while the latter shows the RM and the predicted generation profiles for the same day. We can see that as the difference in temperature is positive, the predicted generation is lower than the RM and vice versa. The ResNet is then trained with the changes in weather conditions to learn how to reshape the RM to predict solar generation.

Fig. 4(c) shows the power scheduled for the small and large instances, the predicted load and the NEM spot price profiles for Novemeber 2nd 2020. It can be seen that during lower price periods (intervals 24–72), most of the load is scheduled and the scheduled load is clipped to the value obtained from sub-problem 1. During high price periods (intervals 72–88), the scheduled load is lower than the predicted baseload due to battery operation. This ensures that electricity cost is minimized.

The organizers provided a base schedule/electricity cost for the problem based on no predictions and an ϵ-greedy algorithm [2]. Table III compares the electricity cost from three different forecast methods, where the base case represents the electricity cost provided by the organizers; whereas BI and the proposed cost represent the cost optimized on their respective forecasts. Our methodology reduced this cost by roughly 23% and 40% compared to the BI and base case, respectively. We were able to solve small instances (158400 binaries) in an average time of 447 seconds per instance and large instances (590400 binaries) in 1550 seconds per instance. It can be seen that even though the problem size grew 3.73 times, the time to solve the problem increased almost linearly and not exponentially with problem size as is the case in MIPs. It further demonstrates

TABLE III
ELECTRICITY COST FROM DIFFERENT FORECASTING RESULTS

Forecast methods	Base case	BI	Proposed
Cost ($AUD)	589356	466652	**357210**

(a) Difference in the RM at and predicted gener- (b) RM vs. predicted solar generation for Nov 4th (c) Scheduled power, predicted net load and NEM spot
ation weather conditions for Nov 4th 2020   2020   price for Nov 2nd 2020
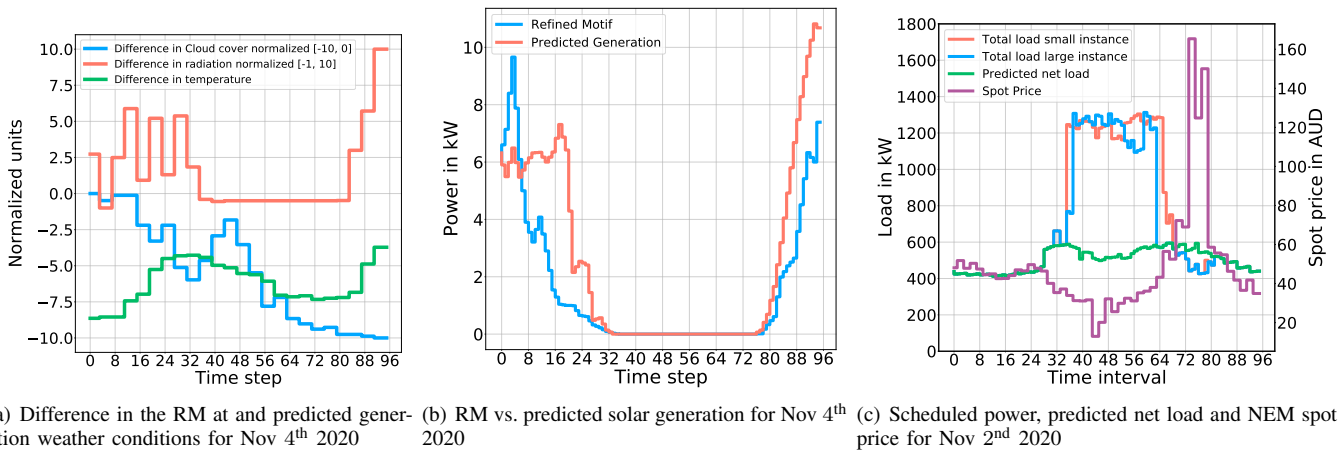
Fig. 4. Prediction and optimization results

the tractability and scalability of our algorithm.

V. CONCLUSION

In this paper, we developed an optimal scheduling solution for solving a real-world electricity cost minimization problem for six buildings at Monash University in Melbourne, Australia. A two-step solution was presented in this paper, which included forecasting and an optimization step. Our solution with energy forecast reduced the total cost of electricity roughly by 40%. We would like to highlight that we placed fifth in the competition among 50 participants and were off from the best solution only by 4% with respect to the objective value due to their better forecast. For future work, the forecasting of baseload and solar generation can be improved by using modern machine learning techniques, such as transformers and info-GANs. For the optimization, a major improvement can be devising an algorithm to select optimal β^{mul} considering the trade-off between the tightness (complexity) of the problem and cost minimization.

REFERENCES

[1] L. Downey, K. Green, J. Grice, D. McConnell, J. Steinert, and N. Twibill, "Untangling the nem: a policymaker's guide to the national electricity market," 2022.

[2] C. Bergmeir. (2022) Ieee cis predict + optimize competition. [Online]. Available: https://ieee-dataport.org/competitions/ieee-cis-technical-challenge-predictoptimize-renewable-energy-scheduling

[3] K. Palani, R. Kota, A. P. Azad, and V. Arya, "Blue skies: A methodology for data-driven clear sky modelling," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 0, pp. 3777–3783, 2017.

[4] H. Long, Z. Zhang, and Y. Su, "Analysis of daily solar power prediction with data-driven approaches," *Applied Energy*, vol. 126, pp. 29–37, 2014.

[5] R. Yuan, S. A. Pourmousavi, W. L. Soong, G. Nguyen, and J. A. Liisberg, "Irmac: Interpretable refined motifs and binary classification for rooftops pv owners," *arXiv preprint arXiv:2109.13732*, 2021. [Online]. Available: https://arxiv.org/pdf/2109.13732.pdf

[6] E. Raviv, K. E. Bouwman, and D. Van Dijk, "Forecasting day-ahead electricity prices: Utilizing hourly prices," *Energy Economics*, vol. 50, pp. 227–239, 2015.

[7] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207015001508

[8] M. Theodosiou, "Forecasting monthly and quarterly time series using stl decomposition," *International Journal of Forecasting*, vol. 27, no. 4, pp. 1178–1195, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207011000070

[9] M. F. Zia, E. Elbouchikhi, and M. Benbouzid, "Microgrids energy management systems: A critical review on methods, solutions, and prospects," *Applied Energy*, vol. 222, pp. 1033–1055, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918306676

[10] N. T. Dinh, S. A. Pourmousavi, S. Karimi-Arpanahi, Y. P. S. Kumar, M. Guo, D. Abbott, and J. A. R. Liisberg, "Optimal sizing and scheduling of community battery storage within a local market," in *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, ser. e-Energy '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 34–46. [Online]. Available: https://doi.org/10.1145/3538637.3538837

[11] S. A. Helal, R. J. Najee, M. O. Hanna, M. F. Shaaban, A. H. Osman, and M. S. Hassan, "An energy management system for hybrid microgrids in remote communities," in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2017, pp. 1–4.

[12] L. N. An and T. Quoc-Tuan, "Optimal energy management for grid connected microgrid by using dynamic programming method," in *2015 IEEE Power & Energy Society General Meeting*, 2015, pp. 1–5.

[13] H. Li, C. Zang, P. Zeng, H. Yu, and Z. Li, "A genetic algorithm-based hybrid optimization approach for microgrid energy management," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2015, pp. 1474–1478.

[14] Y. Balaji, M. Sajedi, N. M. Kalibhat, M. Ding, M. Soltanolkotabi, S. Feizi, and L. Angeles, "UNDERSTANDING OVERPARAMETERIZATION IN GENERATIVE ADVERSARIAL NETWORKS," 2021.

[15] S. Bubeck and M. Sellke, "A Universal Law of Robustness via Isoperimetry," 2021. [Online]. Available: http://arxiv.org/abs/2105.12806

[16] A. Power, Y. Burda, H. Edwards, I. Babuschkin, and V. Misra, "Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets," 2021, pp. 1–9.

[17] FRESNO. (2022) Ieee-cis data challenge - fresno. [Online]. Available: https://gitlab.com/ryuan/ieee-cis-data-challenge-fresno/-/tree/main/data

[18] AEMO. (2022) Nem aggregated demand and price data. [Online]. Available: https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data

[19] Sklearn. (2022) Voting Regressor. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html

[20] J. P. Vielma, "Mixed integer linear programming formulation techniques," *SIAM Review*, vol. 57, no. 1, pp. 3–57, 2015.

[21] B. Das, B. Nair, V. K. Reddy, and P. Venkatesh, "Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of india," *International journal of biometeorology*, vol. 62, no. 10, pp. 1809–1822, 2018.