Integrating Data-Driven Forecasting and Optimization to Improve the Operation of Distributed Energy Storage

Khalid Abdulla, Kent Steer and Andrew Wirth
Melbourne School of Engineering
The University of Melbourne
Melbourne VIC 3010 Australia
(kabdulla@student.unimelb.edu.au)

Julian de Hoog
IBM Research Australia
19/60 City Road
Melbourne VIC 3006 Australia

Saman Halgamuge
Research School of Engineering
College of Engineering & Computer Science
The Australian National University
Canberra, ACT 2601 Australia

Abstract—**Distributed energy technologies, such as residential energy storage, embedded generation, and microgrids, are likely to play an increasing role in future energy systems. Getting the most value from these distributed assets is often dependent on the ability to optimize their operation in a distributed manner. This distributed optimization, in turn, calls for effective short-term forecasts of the output of small-scale generating assets, and the demand of small-scale aggregations of users. This paper introduces the integration of data-driven forecasting and operational optimization methods into a single model, avoiding the need to explicitly produce forecasts. The method is tested against two empirical energy storage operational optimization problems; the minimization of peak energy drawn by a small aggregation of customers, and the minimization of the energy costs of a collection of households which have rooftop PV systems. The integrated forecasting and operational optimization approach performs well at the peak demand minimization problem for intermediate-sized aggregations (50 residential customers or more), while an approach with separate forecasting and optimization performed better on the energy cost minimization problem. These results suggest that the integrated approach can be effective in applications where (i) forecasting difficulty is intermediate, and (ii) the exact operational optimization formulation can be well approximated by a data-driven model trained on a small fraction of the available forecast training data.**

I. Introduction

The control of many energy assets can be improved if accurate forecasts are available for the output of non-dispatchable generators and the load of non-deferrable demands. This is perhaps most obvious when operating energy storage assets, which produce an inherent coupling (dynamics) between time intervals: the operational constraints at any given interval are directly affected by preceding operational decisions.

Fortunately, there is a large body of research and practice dedicated to forecasting, as is briefly surveyed in Section II. Forecasting methods can be broadly characterized into (i) *Causal* methods, where an understanding of the physical processes driving the value being forecast is developed, allowing a physical model (initialized to known conditions) to be simulated into the future; and (ii) *Data-driven* methods, where historic data is used to identify patterns in the variable being forecast, and correlations between the variable being forecast and other variables (for which a forecast is available, or can be made). There are also methods combining both approaches.

In data-driven forecasting a model is produced which is characterized by a number of (often physically meaningless) parameters, whose values are chosen to minimize some forecast error metric on an available historic (training) dataset. The error metric chosen for this parameter-optimization process is typically not given much attention. However, when forecasting variables which are 'difficult' to forecast, the chosen error metric can have a significant impact on the forecasts which are produced. A simple illustration of this is provided by Kolassa and Martin [1], in the example of rolling a fair six-sided die. The prediction of any future roll which minimizes MAPE (Mean Absolute Percentage Error - a common error metric for electricity demand) is 2.0, despite the expected value being 3.5. This is due to MAPE's asymmetry; it penalizes over-forecasts more heavily than under-forecasts. Small-scale aggregations of demand are known to be 'difficult' to forecast [2] [3], and so is the output from a single roof-top PV system (see Table 4 of [4], for example), hence the operation of distributed energy storage is an interesting area in which to consider the role of forecast errors on controller performance.

In this paper a framework is presented which combines a data-driven forecast model, with a data-driven optimization model in an attempt to improve the operation of a distributed energy storage asset. By combining these two data-driven models we remove the need to explicitly produce a forecast, which avoids the sensitivity of the output of such forecasts to the selected forecast error metric (as no selection needs to be made). The basic

idea is to use a historical dataset, which would typically be used to train a forecast model, in combination with an exact operational optimization model to directly learn a mapping from state measurements and predictor variables onto good control actions (see Fig. 1).

This framework can be applied wherever (i) a controller requires a point forecast over some horizon, (ii) a data-driven controller would be an acceptable replacement, and (iii) the controller and plant-under control can be accurately simulated ahead of time (to allow the generation of a training dataset). The following contributions are made:

i A framework is introduced which allows a single data-driven optimization model to replace separate data-driven forecast models and a data-driven or exact optimization model;
ii The performance of this combined model is assessed in two distinct applications for operating a distributed energy storage asset, and empirical results are presented;
iii The reasons for the relative effectiveness of the approach in the two applications is discussed, and guidelines for when this approach may be attractive are given.

The remainder of this paper is structured as follows: Section II gives a brief review of some of the relevant forecasting and model-free optimization literature, Section III details the framework for producing integrated forecasting and optimization models and provides details of the empirical applications considered; results are discussed and conclusions drawn in Sections IV and V respectively.

II. Literature Review

A. Short-Term Univariate Forecasting

This paper focuses on univariate forecasting methods, where the only input available for making a forecast are a time-history of previous values of the variable being forecast. These methods are of interest for two reasons, firstly by not requiring external inputs the resulting forecast models can be applied to a very wide range of problems, and secondly they are perhaps the most pure form of data-driven forecasting; in that univariate models can only hope to spot patterns and trends in the data, and not exploit a physical understanding of what causes them.

In addition we restrict our review to short-term forecasting of the output of roof-top PV systems and the electrical demand of households; as these are the forecasts required in our applications, however the framework presented could be applied wherever data-driven forecasts are (or could be) used.

1) Load Forecasting: Short-Term Load Forecasting (STLF) has traditionally been focused on larger aggregations of demand (Hong's dissertation provides a helpful review, [5]), but with the greater adoption of smart-meters interest in forecasting at smaller aggregations is increasing. In [2] Sevlian *et al*. compare the performance of three univariate forecasting models; seasonal autoregressive moving average, support vector regression, and feed-forward neural networks (NN), in the application of forecasting demand at various levels of aggregation. They find that as aggregation level is increased the predictability of aggregated demand also increases, as might be expected as the random component of each household's demand profile is uncorrelated. Mirowski *et al*. , [3], perform a similar study but look at five different models.

2) PV Forecasting: [4] offers a recent review of state-of-the art methods for solar and PV forecasting. Figs. 7 and 8 of [4] indicate that similar to demand forecasting increasing the area and number of PV systems which are being forecast reduces the forecast errors achieved by state-of-the-art methods.

B. Model-Free Optimization

The distinction between *causal* and *data-driven* forecasting methods, discussed in the introduction, can also be considered to broadly categorize operational optimization models. An *exact* optimization model will maximize the performance of some plant subject to constraints, in some mathematically exact sense, whereas a *model-free* optimization will attempt to find a solution with a high performance, but in general offers no guarantee of optimality. The benefits of exact optimization models are obvious, but they also have a number of drawbacks. For example, exact approaches tend to put conditions on the form which objective functions and constraints can take. Also the exactness of any optimization which receives a forecast is conditional on that forecast being accurate (having zero residual in the case of point forecasts provided to a deterministic optimization, or having an accurate distribution in the case of a probabilistic forecast provided to a stochastic optimization).

Model-free operational optimization and similar methods have different names depending on the field of application, and the details of methods used: *Simulation Optimization*, *Reinforcement Learning* and *Approximate Dynamic Programming* are some of the most common. [6] offers a review of different simulation optimization approaches. Model-free optimization is often resorted to when exact approaches are not available due to analytical and computational intractability. However, this is not the case in the applications considered here; the problems considered have tractable exact solutions provided a perfect forecast is available. Model-free operational optimization is applied because ideal forecasts are not available. To the best of the authors' knowledge, no previous studies have specifically looked at using a model-free optimization, as a means of integrating

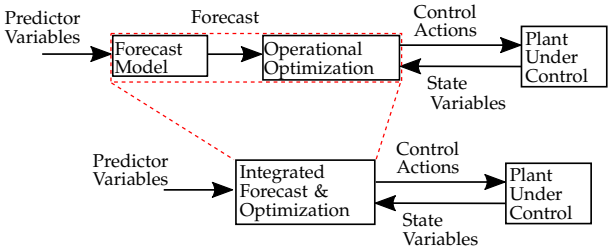

Fig. 1. Illustration of separate (top) and combined (bottom) forecast and operational optimization models

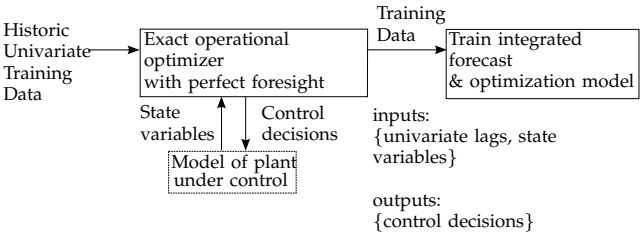

Fig. 2. Framework for Training Integrated Forecast & Optimization (IFO) model

an operational optimization model and a data-driven forecast model into a single data-driven model.

C. Energy Storage Operational Optimization

A significant body of literature is concerned with the possible roles, and operational optimization approaches for energy storage systems; a comprehensive review is beyond the scope of this paper. A specific example relevant to the first case-study considered, are two recent papers by Rowe *et al*. [7], [8] which seek to operate energy storage systems for peak-demand minimization. In [7] an off-line scheduling approach to peak-reduction is considered for aggregations of 30 homes, with a battery sized at 25% of the maximum demand. The approach is able to achieve a peak-reduction of approximately 5%, however this is only after aggregations for which the method performed poorly have been removed from consideration. [8] considers a deterministic model-predictive-control (MPC) and a stochastic receding horizon controller (SRHC) approach to the peak demand minimization problem. With a battery sized at 25% of the maximum demand, and an aggregation of 35 customers, these methods achieve average peak reductions of 7.5% and 8% respectively.

III. Method

A. Framework

Fig. 1 illustrates the combination of a forecasting model and an operational optimization model which this framework seeks to achieve. The integrated forecast & optimization (IFO) model has inputs composed of all of the inputs which could separately be presented to a data-driven forecast model, and an exact operational optimization model, and its outputs are control actions which are passed to the plant under control. This approach is schematically similar to reinforcement learning, where an agent interacts with an (initially unknown) environment, and updates its policy in response to rewards received when it takes particular actions from particular states. The main difference to conventional reinforcement learning is the means by which we train the IFO model.

To train the IFO model, it is necessary to generate a training dataset. The 'predictor variables' are available from historic data of the variable being forecast (we only consider univariate forecasts), and the state-variables of the plant under control can be generated by simulating the behavior of the plant ahead of time; but it is not immediately obvious where the training set of control actions should come from. We propose using an exact operational optimization model provided with an artificial perfect foresight forecast. The overall training process for the IFO model is illustrated in Fig. 2. The steps of this process are:

1) Simulate an exact operational optimization model and the plant on the training dataset. Provide the model with a perfect-foresight forecast;
2) Use this simulation to generate a set of inputs {predictor variables, plant state variables}, and outputs {control actions};
3) Use these inputs and outputs to train a data-driven IFO model;
4) This model can then be used on-line in place of both a data-driven forecast model, and an exact operational optimization model.

We consider two applications of distributed energy storage, and in both cases we consider operation at different levels of aggregation (*i.e*. different numbers of houses aggregated and operated as a single entity). In this process we make the simplifying assumption that there is sufficient communication and control infrastructure to enable this centralized operation, whilst noting that in practice such facilities would be costly to install, and the impact of communication network failures would need to be considered in a more detailed study.

B. Applications

Two empirical applications of the framework are considered in this paper; the operation of a distributed energy storage asset to minimize the peak power drawn by an aggregation of customers over a billing period; and the operation of a residential battery to minimize the cost of meeting local demand given a roof-top PV system, and a residential two-part time-of-use tariff. In both applications a receding horizon controller is used as the *exact* (subject to the myopia caused by a finite

horizon) operational optimization method; but in the first linear programming is used to find the optimal charging decisions for each horizon, and in the second dynamic programming is used.

1) Application One: Demand Peak-Minimization: The first application considers the operation of a distributed energy storage asset to reduce the peak power drawn by an aggregation of customers. This problem is illustrated in Fig. 3, and the formulation described in [9] is used.

The problem is solved using receding horizon control; in each interval the following linear program is solved to determine the optimal charging operations over a horizon T intervals long:

Minimize:

$$E \tag{1}$$

Subject to:

$$\sum_{t=1}^{\tau} b_t \geq \text{SoC} - B \quad \forall \tau \in \{1, \ldots, T\} \tag{2}$$

$$\sum_{t=1}^{\tau} b_t \leq \text{SoC} \quad \forall \tau \in \{1, \ldots, T\} \tag{3}$$

$$E \geq \hat{d}_t - b_t - E_{\text{bill}} \quad \forall t \in \{1, \ldots, T\} \tag{4}$$

$$-b_{\text{lim}} \leq b_t \leq b_{\text{lim}} \quad \forall t \in \{1, \ldots, T\} \tag{5}$$

Where:

- E is the amount by which the peak power (measured as energy over an interval in [kWh]) drawn over the horizon is expected to exceed the running peak;
- b_t is the energy to discharge from the battery during interval t - the decision variables [kWh];
- T is the number of steps in the forecast and planning horizon, for this instance $T = 48$;
- SoC (State of Charge) is the amount of energy in the battery at the start of the horizon for which the optimization is being run [kWh];
- $\hat{d}_t$ is the forecast demand during interval t (a power, but measured as energy over the interval in [kWh]);
- E_{bill} is the highest per interval demand within the billing period so far [kWh];
- b_{lim} is the maximum energy which can be transferred to/from the battery within an interval [kWh];
- B is the usable energy capacity of the battery [kWh].

(1) ensures that the exceedance of the previous demand peak is minimized over the horizon; (2) and (3) ensure that the state-of-charge constraints of the battery are not violated in any interval; (4) ensures that E is appropriately set as the amount the peak demand during the horizon is expected to exceed the running peak so far in the billing period (assuming all control actions are implemented and forecasts are correct); (5) ensures the energy exchanged with the battery in an an interval is within the battery's rate-of-charge limits. This optimization outputs the optimal discharging energies for the battery, b_t, over the next T intervals assuming that the forecast, $\hat{d}_t$, has zero error. Robustness to forecast errors is achieved by using the solution within a receding-horizon-controller: only the first of these decisions, b_1, is implemented and the optimization is re-run at the next interval, with a new forecast.

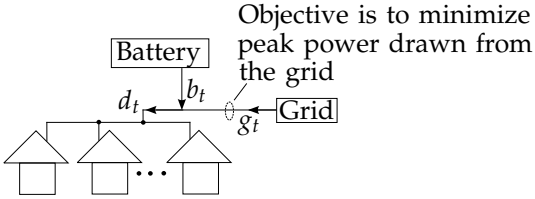

Fig. 3. Application One: Minimizing the peak power drawn by an aggregation of customers.

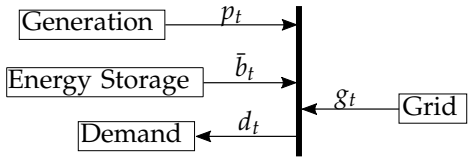

Fig. 4. Application Two: Minimizing the cost of meeting local demand. Positive sign conventions for power flows are shown. b_t is controlled. b_t and g_t take negative values when energy is transmitted to the battery or grid.

TABLE I
TARIFF STRUCTURE FOR APPLICATION TWO

Export Tariff, r_t	0.05 \$[1]/kWh	flat rate
Import Tariff, c_t	0.40 \$/kWh	7:00AM – 10:00PM,
	0.20 \$/kWh	at other times

2) Application Two: Residential tariff optimization: The second application considers the optimal operation of a residential energy storage asset in order to minimize the cost of meeting the local demand of a set of households which have roof-top PV systems. The sign convention for power flows is depicted in Fig. 4. The tariff structure considered is given in Table I, and is intended to be representative of tariffs which Australian residential electricity customers with roof-top PV systems are likely to have over the coming decade. The tariff structure means there are two ways in which a battery may be able to reduce the costs of meeting local demand: (i) By minimizing energy exports which receive only a small price, compared to if the energy from a PV system were used to offset local demand at another time; (ii) By optimizing use of the time-of-use tariff (minimizing imports during peak times).

The *exact* (subject to finite-horizon myopia) operational optimization is once again implemented using receding horizon control, considering a 24-hour horizon. The optimization of an individual horizon is achieved using dynamic programming, within which the battery state of charge, q_t, provides a complete description of the *state*, and the interval within the horizon is the *stage*. We use a simplified version of the formulation presented in [10].

If $CTG_t(q_t)$, is the optimal 'cost to go' from stage t given that the battery capacity at the start of interval t is q_t; and $STC_t(b_t)$ is the 'state transition cost' for stage t given that control action b_t has been chosen.

The recursive relationship determining the cost-to-go is then:

$$CTG_t(q_t) = \min_{b_t}\{STC_t(b_t) + CTG_{t+1}(q_t - b_t)\} \quad (6)$$

Included in (6) is the state-update rule which ensures that the correct cost-to-go at interval $t+1$ is considered given that state-transition b_t is chosen.

If $\bar{b}_t$ is defined as the amount of energy supplied to the grid when battery's state-of-charge is decreased by b_t kWh, it can be calculated as:

$$\bar{b}_t = \begin{cases} b_t/\eta_c & b_t < 0 \\ b_t\eta_d & b_t \geq 0 \end{cases} \quad (7)$$

where η_c and η_d are the charging and discharging efficiencies of the battery which lie in the range $(0, 1]$.

The state transition cost is then:

$$STC_t(b_t) = c_t[g_t]^+ - r_t[g_t]^- \quad (8)$$

where $[.]^+$ represents taking the positive component of [.], i.e. $[x]^+ = \max(0, x)$, similarly $[x]^- = \max(0, -x)$. The terms of the state transition cost are; the cost of any energy imported from the grid (g_t), and the reward received for energy exported to the grid.

Constraints are implemented as follows:

- Energy balance is ensured by defining energy from the grid during an interval, t, as $g_t = d_t - p_t - \bar{b}_t$, and asserting that the grid can supply or accept power as required.
- Rates of charging and discharging are kept feasible by limiting the minimum (most negative) and maximum values of b_t, i.e. $b_t \in [-b_{\text{lim}}, b_{\text{lim}}]$
- Battery capacity constraints are respected by further limiting the minimum and maximum values of b_t for particular states, q_t, i.e.:

$$\begin{aligned} b_t &\geq \max(q_t - B,\ -b_{\text{lim}}) \\ b_t &\leq \min(q_t,\ b_{\text{lim}}) \end{aligned} \quad (9)$$

This dynamic program is solved by backwards induction. The cost-to-go from the final stage of the horizon is zero for all possible states (as there is no demand left to be satisfied in the horizon). For each preceding stage $t = \{T-1, T-2, \ldots, 0\}$, and possible state to be in q_t, the feasible charge decision, b_t, which minimizes (6), is found by exhaustive search, and the optimal costs-to-go for that stage are computed. This process gives the minimum-cost charging profile from all possible starting states-of-charge, and the decision corresponding to the current state of charge (q_0, which is known) is chosen. The optimal first discharge decision, b_0^*, is implemented and a new solution is computed one interval later.

C. Application Data

For *Application One*, demand peak-minimization of an aggregation of customers, we use data from the Irish Social Science Data Archive (ISSDA), [11]. It includes the energy consumed in each half-hourly interval from the smart-meters of 4225 residential users for the period July 2009 to December 2010. To prepare the data for analysis, customers have been combined into random aggregates of demand (by randomly selecting the desired number of customers, and summing their demand in each metering interval). The first year of data is used for training, and the remaining 6 months for testing. For each aggregation we consider a battery capacity, B which is 5% of the aggregate daily demand, a 24-hour (48-interval) forecast and planning horizon, and a 7-day billing period over which a peak demand is calculated.

For *Application Two*, minimizing the cost of meeting demand subject to a residential tariff, the demand and generation data is derived from the publicly available Ausgrid 'Solar Home Electricity Data' [12]. It contains two years of half-hourly meter readings (kWh over the half-hour interval) for 300 households in New South Wales, Australia, all of whom have a solar PV generation system. Each customer's gross demand and solar generation are separately metered. The first year of data is used for training, and the following 6 months for testing. We consider a 24-hour (48-interval) forecasting and planning horizon, and consider a battery of 2kWh capacity per household included in the aggregation.

For both applications we consider a battery charging and discharging rate of 2C (*i.e.* (dis)charging rate in kW is twice the battery capacity in kWh).

D. Data-driven Model

There are several data-driven methods to choose from, and none of them out-perform all others in all cases. Here single hidden layer feed-forward neural networks (NN) are used, as they are a widely understood and flexible class of data-driven model, and with sufficient nodes in the hidden layer are capable of approximating any function arbitrarily closely. Details of the implementation are not given, as the MATLAB neural-network toolbox is used; but a repository of the source code used in this paper is provided [13]. For all forecast and IFO models 50 hidden nodes are considered, 10 random

[1]$ symbol refers to Australian Dollars throughout this paper

initializations are used and training of each network is halted when the error on a 20% held-out validation dataset increases for 6 training epochs in a row, to avoid over-fitting.

For each application in addition to the IFO model, we consider an exact operational optimization model provided with three different forecasts. The first is a daily naive-periodic (NP) forecast which assumes the demand (and PV) output is equal to that from 24-hours ago. The second is a univariate neural-network (NN) forecast whose inputs are the demand (PV) in the 24-hours (here, 48 intervals) leading up to the present interval, and whose outputs are the demand (PV) for the current interval and the subsequent 47 intervals. These models are trained on the first year of training data, and tested on the following 6 months. The final method is an artificial perfect foresight (PF) forecast which is not implementable, but is included to provide a prescient upper-bound on performance.

E. Set-Point Control

Given that the reason for considering an IFO model is that forecasts may be poor, it is helpful to consider simple set-point controllers which act on state information alone, so have no need of forecasts.

For *Application One* there is an intuitively obvious set-point controller which charges the battery by $E_{\text{bill}} - d_t$, *i.e.* the difference between the running demand peak, and the present demand (subject to rate-of-charge and state-of-charge constraints). This method has the advantage that it can act on the actual demand value for the present interval, rather than a forecast, but is myopic, in that it does not consider the impact of current decisions on future intervals.

For *Application Two* we consider a set-point controller which seeks to avoid export to the grid (for minimal return) at all times. It chooses to charge the battery by $p_t - d_t$ subject to rate-of-charge and state-of-charge constraints.

IV. Results

The results of applying the set-point, forecast-driven, and IFO models to *Application One* are shown in Fig. 5; which shows the peak reduction ratio $(1 - E_{\text{bill}}/\max(d_t))$, where E_{bill} is the peak energy imported from the grid over a billing period whilst operating the battery, and $\max(d_t)$ is the peak of local demand.

It can be seen that set-point control performs well for small aggregations of demand (fewer than 50 households), at these aggregation levels forecasts are poor, as indicated by the forecast noise:signal ratio. The noise:signal ratio is estimated as:

$$\text{N:S} = \frac{\sqrt{\sum (d_t - \hat{d}_t)^2}}{\sqrt{\sum \hat{d}_t^2}} \tag{10}$$

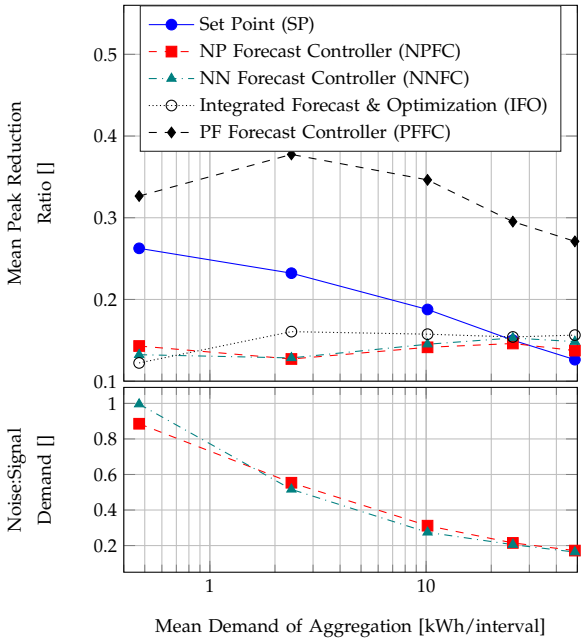

Fig. 5. *Application One*: Peak reduction ratio of a battery operated using various control Methods, at different levels of aggregation: {1, 5, 20, 50, 100} households. Bottom plot shows the noise:signal ratio of the demand forecasts for the forecast-driven methods.

Where d_t is the actual demand (PV) in interval t, $\hat{d}_t$ is the forecast demand (PV) during interval t, and the sum is taken over intervals in a horizon. This calculation assumes that the forecast for the horizon represents the discernible signal, and that any error is the result of noise (which is unforeseeable given the predictor variables). This will not generally be the case, however, as it is likely that some improvement in forecasting performance is always possible, even with a fixed set of predictor variables.

Fig. 5 shows that the set-point method performs most effectively below noise:signal ratios of about 0.25. It also shows that the IFO model outperforms forecast-driven optimization approaches below a signal:noise ratio of about 0.7. Above this noise ratio the poor predictability of demand over the horizon means it is difficult to learn the control function mapping. For the largest aggregations considered (of 50 and 100 households) the IFO method out-performed others considered, with the exception of an exact controller receiving a perfect foresight forecast, which of course is not implementable in practice. The finding that a MPC-based method out-performs set-point control only at or above an aggregation of about 50 customers is consistent with findings in [8]. Note that for a single customer a controller provided with a NP forecast out-performs that given a NN forecast, despite the NN achieving lower mean-squared-errors (see Fig. 6). This suggests mean-squared-error is not the best metric to minimize for this particular control problem; this concept is expanded on in [9].

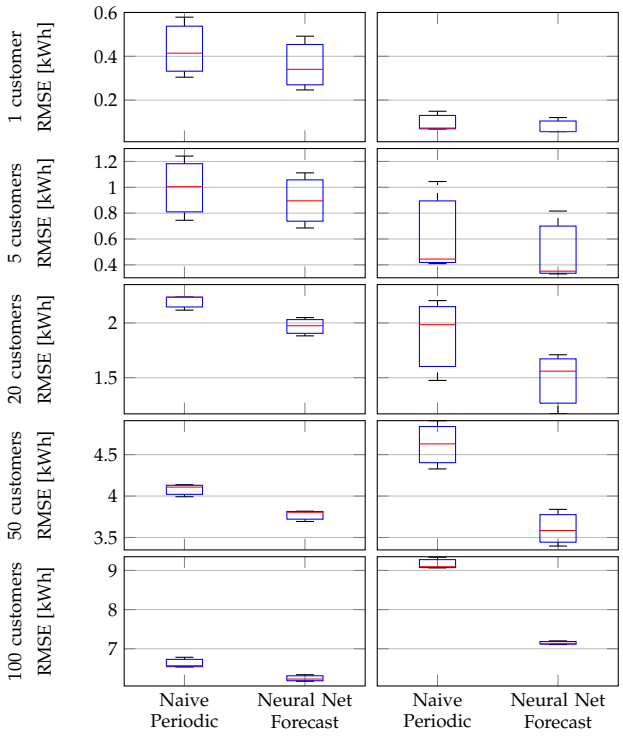

Fig. 6. *Application Two*: Root-mean-squared-error (RMSE) for Demand (left) and PV (right) forecasts for the forecast-driven methods

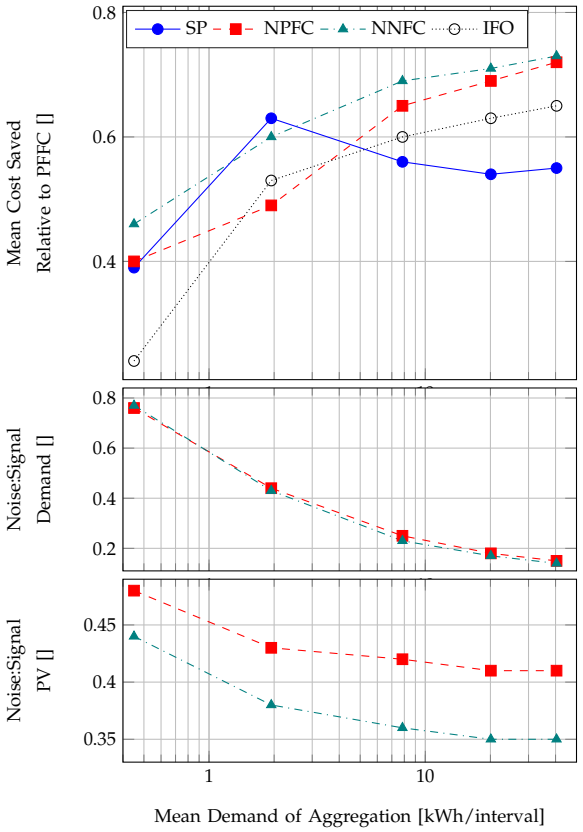

Fig. 7. *Application Two*: Cost Saving of a Battery Operated Using Various Control Methods, at different levels of aggregation: {1, 5, 20, 50, 100} households, shown relative to the cost-saving of an exact controller provided with a perfect-foresight forecast. The second and third plots show the noise:signal ratio of the demand and PV forecasts respectively, for the forecast-driven methods.

Fig. 6 shows the performance of the NP and NN forecasts used for the PV and demand in *Application Two*. It shows that the NN forecast provides reduced forecast errors compared to naive periodic forecasting, especially at higher levels of customer aggregation. Fig. 7 shows the cost saved in *Application Two* by operating the battery to minimize the cost of meeting local demand subject to a residential tariff, at varying levels of aggregation. An exact controller with a separate data-driven forecast model outperforms the IFO model at all levels of aggregation considered.

A. Learn-ability of exact operational optimization model

The IFO modeling approach (as illustrated in Figs. 1 and 2) relies on being able to learn good operational behavior as a mapping {state variables, univariate lags} $\rightarrow$ {control decisions}, based on a finite set of historical data. For this to be effective, it is necessary that the controller behavior (*i.e.* its mapping of {forecasts, state variables} $\rightarrow$ {control actions}) can be learned using the data-driven model being used, with a number of samples which is smaller than the available historic training data.

This requirement was tested by generating random inputs (of both state and horizon-forecasts) for the forecast-driven model, generating the corresponding model output (from the exact model), and training a NN to learn this mapping (*i.e.* how well can the data-driven model represent the forecast-driven controller, without the additional task of making forecasts?). The results are presented in Figs 8 and 9 for *Application One* and *Application Two* respectively. These plots demonstrate that whilst the forecast-driven models used in both applications can be represented to some degree of accuracy using the data-driven models considered, the dynamic programming optimization of *Application Two* requires substantially more data, and has larger errors.

Seeing this difference in how easily learned the controller mappings are for the two applications, it is not surprising that the IFO approach is significantly more effective in *Application One*. This would suggest that for the IFO approach to be effective, it is necessary that the forecast-driven controller being replaced can be well represented by a data-driven model trained on a small fraction of the available historic training data. This seems intuitively reasonable, as within that same historic dataset it is necessary to also learn the mapping from univariate lags to forecast outcomes.

V. Conclusions

A method has been presented which allows an exact forecast-driven operational optimization model to be replaced by an integrated forecast and optimization model.

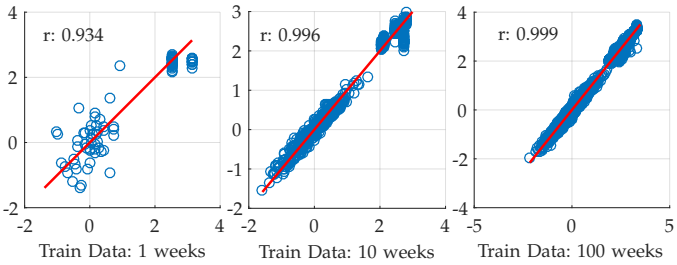

Fig. 8. Out-of-sample Neural Network Regression Fit to *Application One* controller mapping. In each plot the x-axis gives the controller response to a set of randomly generated {state, forecast} inputs, and the y-axis gives the NN response having been trained on a separate training set of the length indicated below the plot.

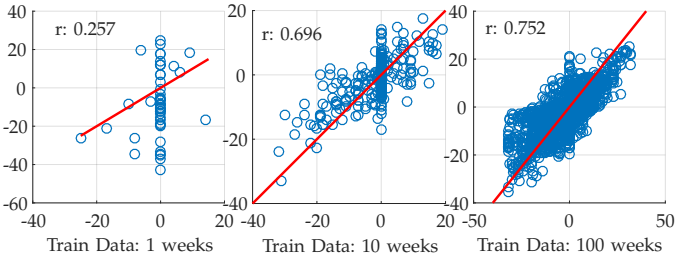

Fig. 9. Out-of-sample Neural Network Regression Fit to *Application Two* controller mapping. In each plot the x-axis gives the controller response to a set of randomly generated {state, forecast} inputs, and the y-axis gives the NN response having been trained on a separate training set of the length indicated below the plot.

The potential benefit of doing so is that it is no longer necessary to produce forecasts, so the difficult question of what forecast error metric should be minimized in a particular application is avoided.

The method has been tested in two empirical applications in which we seek to optimize the operation of an energy storage asset. In the first application a battery is operated so as to minimize the peak power drawn by an aggregation of customers, and for large aggregations (more than 50 households) the proposed integrated approach out-performed forecast-driven approaches. In the second application we seek to operate a battery so as to minimize the cost of meeting residential demand, with a roof-top PV system available, and subject to a residential tariff. In this second application the integrated approach performed poorly compared to forecast-driven approaches provided with a separate data-driven forecast.

The difference in performance of the proposed approach between the two applications is thought to be a result of two main factors, which may provide some guidance on what other problems might be effectively tackled using a similar integrated approach. The integrated approach is only likely to be worth considering when:

i The mapping of the forecast-driven controller it replaces can be well-represented by the data-driven model being used when trained on a number of samples corresponding to a small fraction of the available historic training data;

ii The forecasting sub-problem is of intermediate difficulty: If forecasting is very difficult performance of the integrated approach is likely to be poor (see left of Figs. 5 and 7), if it is very easy then forecasts are likely to be good, and similar regardless of the forecast error metric minimized, so forecast-driven approaches are likely to be more effective.

VI. Acknowledgments

We would like to thank four anonymous reviewers for their feedback which improved the final version of this paper. This work was supported by Melbourne International Research and Fee Remission Scholarships, and an Australia-Indonesia Center grant.

References

[1] S. Kolassa and R. Martin, "Percentage errors can ruin your day (and rolling the dice shows how)," *Foresight*, vol. Fall, pp. 21–27, 2011.

[2] R. Sevlian and R. Rajagopal, "Short term electricity load forecasting on varying levels of aggregation," *arXiv preprint arXiv:1404.0058*, 2014.

[3] P. Mirowski, S. Chen, T. K. Ho, and C.-N. Yu, "Demand forecasting in smart grids," *Bell Labs Technical Journal*, vol. 18, no. 4, pp. 135–158, 2014.

[4] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. D. Brabandere, "Photovoltaic and Solar Forecasting: State of the Art," *International Energy Agency: Photovoltaic Power Systems Programme, Report IEA PVPS T14*, pp. 1–40, 2013. [Online]. Available: http://www.iea-pvps.org/index.php?id=278&eID=dam_frontend_push&docID=1690

[5] T. Hong, "Short term electric load forecasting," September 2010.

[6] M. C. Fu, F. W. Glover, and J. April, "Simulation optimization: a review, new developments, and applications," *Simulation Conference, 2005 Proceedings of the Winter*, p. 13 pp., 2005.

[7] M. Rowe, T. Yunusov, S. Haben, C. Singleton, W. Holderbaum, and B. Potter, "A Peak Reduction Scheduling Algorithm for Storage Devices on the Low Voltage Network," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 2115–2124, 2014.

[8] M. Rowe, T. Yunusov, S. Haben, W. Holderbaum, and B. Potter, "The Real-Time Optimisation of DNO Owned Storage Devices on the LV Network for Peak Reduction," *Energies*, vol. 7, no. 6, pp. 3537–3560, May 2014. [Online]. Available: http://www.mdpi.com/1996-1073/7/6/3537/

[9] K. Abdulla, K. Steer, A. Wirth, and S. Halgamuge, "Improving the on-line control of energy storage via forecast error metric customization," *Journal of Energy Storage*, vol. 8, pp. 51 – 59, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352152X16301554

[10] K. Abdulla, J. D. Hoog, V. Muenzel, F. Suits, K. Steer, A. Wirth, and S. Halgamuge, "Optimal operation of energy storage systems considering forecasts and battery degradation," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[11] ISSDA. CER smart meter customer behaviour trials data. Accessed via the Irish Social Science Data Archive, ver. CER Electricity Revised March 2012, on 15th Sept. 2015: www.ucd.ie/issda. [Online]. Available: www.ucd.ie/issda

[12] Ausgrid solar home electricity data, v.2. (Accessed: 15th Sept. 2015: http://www.ausgrid.com.au/Common/About-us/Corporate-information/Data-to-share/Data-to-share/Solar-household-data.aspx). [Online]. Available: http://www.ausgrid.com.au/Common/About-us/Corporate-information/Data-to-share/Data-to-share/Solar-household-data.aspx

[13] K. Abdulla, "Integrating data-driven forecasting and optimization," https://github.com/khalida/uceso, 2016.