Improving the On-line Control of Energy Storage via Forecast Error Metric Customization

Khalid Abdulla[a,*], Kent Steer[a], Andrew Wirth[a], Saman Halgamuge[a]

[a]*Department of Mechanical Engineering, University of Melbourne, Melbourne, Australia, VIC, 3010.*

Abstract

The economical operation of many distributed energy assets relies on effective on-line control, which in turn often requires forecasts to be made. To produce and evaluate forecasts, the error metric by which one measures forecast accuracy must be selected. A new method is presented which customizes a forecast error metric to a given on-line control problem instance, in order to improve the controller's performance. This method is applied to the real-time operation of a battery with the objective of minimizing the peak power drawn by an aggregation of customers over a billing period. In the empirical example considered, customizing the forecast error metric to each problem instance, improved performance by 45% on average, compared to a controller provided with a forecast of the same type, but trained to minimize mean-squared-error. Error metric customization is made possible by two newly proposed parametrized error metrics. The proposed method can be applied to any on-line optimization problem which requires a point forecast as an input, and which can be accurately simulated ahead of time. The method is likely to be most effective in applications where forecasting errors are quite high, as in these applications the choice of forecast error metric significantly affects the forecasts which are produced.

Keywords: Demand forecasting, forecast uncertainty, predictive control, optimization, energy storage

1. Introduction

Increasing the proportion of electrical energy derived from Renewable Energy Sources (RES) is widely accepted as being a laudable goal, be this for energy security, sustainability, or to reduce environmental impacts. RES tend to be non-dispatchable and can only be forecast with limited accuracy, which makes it challenging to rely on them for a large proportion of our electricity supply. To address these challenges researchers are developing methods to find effective means of integrating RES.

The effective integration of non-dispatchable RES will often require on-line control. Due to the uncertainty of electrical demand and RES generation, a first step in achieving this control is typically to produce forecasts. This paper presents a method for generating forecasts which are most useful to a given on-line control problem, in particular forecasts of low-level demand aggregations (equivalent loads of up to 1000 households).

The remainder of the introduction presents and defines key concepts and reviews relevant literature, Section 2 details the methods used and the proposed approach for customizing a forecast error metric to a specific problem. Section 3 presents and discusses the results of applying this approach to an empirical example, and Section 4 draws some conclusions.

This paper makes the following contributions:

1. A method is proposed for customizing a forecast error metric to a particular instance of a control problem (Section 2);
2. Two new parametrized error metrics: Parametrized Forecast Error Metric (PFEM), and Parametrized Earth Mover's Distance (PEMD) are proposed, which allow error metric customization (Section 2.3);
3. The performance of this error metric customization method is demonstrated in the application of producing forecasts for a controller which seeks to operate a local energy storage asset to minimize the peak power drawn by an aggregation of customers over a billing period (Section 3).

1.1. Forecasting and Control

The effective control of energy assets often requires an ability to predict future conditions with reasonable accuracy. This is particularly evident in the case of controlling energy storage, where control decisions made now (to charge or discharge) affect what decisions are available to be made in the future.

Forecasting and control are often treated as separate problems. Papers focusing on forecasting typically concentrate on minimizing an arbitrary forecast error metric on some unseen test dataset; and papers focusing on control

*Corresponding author

Email address: kabdulla@student.unimelb.edu.au (Khalid Abdulla)

URL: http://people.eng.unimelb.edu.au/kabdulla/ (Khalid Abdulla)

often assume some forecast of future conditions is available (sometimes with a randomly distributed error), or use a standard forecasting method to produce forecasts for their controller. In the present work the forecasting and control problems are connected, by allowing a forecast error metric to be customized for a particular on-line control problem (see Fig. 1 for a schematic illustration of the approach).

1.2. Forecasting Small Aggregations of Demand

The distributed availability of many RES has motivated attempts to exploit them using distributed solutions such as micro-grids, distributed generation, and distributed energy storage. Optimally controlling these distributed assets requires forecasts for small demand aggregations, as presented in this paper, and the generation from small RES. Historically, Short-Term Load Forecasting (STLF) has focused on large aggregations of demand, at the level of entire cities or regions, because electricity supply systems have been based on centralized generation and dispatch. Recent studies, for example those by Sevlian *et al.* [1] and Mirowski *et al.* [2], show that forecasting small demand aggregations is more difficult. The increased difficulty of forecasting small demand aggregations has motivated some researchers to develop new forecasting methods (*e.g.* [3]). However, the present work considers the issues associated with traditional forecast error metrics when forecasting small aggregations of demand, and an approach to addressing these.

1.3. Forecast Error Metrics

Error metrics measure the accuracy of a forecast, compared to the actual values, and are used as the cost-function (or loss-function) of forecast models. What makes a good error metric depends on the application but generally, in addition to providing some distance measure between forecast and actual values, it should:

- Be readily interpretable;
- Be quick to compute;
- Reflect the net increase in costs resulting from decisions made based on incorrect forecasts.

Many existing error metrics fulfill the first two requirements, but make little attempt at directly addressing the third, and arguably most important, requirement. In many applications it is difficult or impossible to assess how useful a forecast is. For example, a weather forecast has many users who derive different benefits, hence computing a single appropriate error metric based on users' decisions is not practical.

In [4], Murphy discusses three distinct types of 'goodness' of forecasts. The two which are of relevance here are defined:

- Quality: Correspondence between forecasts and observations,
- Value: Incremental benefits of forecasts to users.

High 'Quality' forecasts are close (according to some error metric) to observations, whereas high 'Value' forecasts are useful to their intended users. In this paper it is argued that by appropriately selecting an error metric to be minimized (by which one measures Quality), forecasts of high Value can be produced. Forecasting for an on-line controller makes this possible; because there is only a single user, and the decisions that user will make based on forecasts is known in advance.

Commonly used error metrics in STLF include Mean Absolute Percentage Error, and Mean Squared Error, their acronyms and definitions are given below:

$$\mathrm{MAPE}(\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{N} \sum_{t=1}^{N} \frac{|y_t - x_t|}{x_t} \tag{1}$$

$$\mathrm{MSE}(\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{N} \sum_{t=1}^{N} (y_t - x_t)^2 \tag{2}$$

These metrics evaluate the performance of a forecast based on each interval, t, within a horizon $\{1, \ldots, N\}$ individually, based on the error between the forecast, y_t, and actual value, x_t. As a result they all suffer from the 'double penalty' effect described by Haben *et al.* [5]; where an error metric penalizes a forecast twice for an event which is correctly predicted but displaced along the time-axis.

Another issue affecting the selection of error metric is symmetry; the relative penalty applied for over-forecasts and under-forecasts. MAPE penalizes over-forecasts more severely than under-forecasts. This is illustrated by Kolassa and Martin [6] in the example of rolling a fair six-sided die, where the MAPE-minimizing forecast of the value shown on the face is 2, despite the expected value being 3.5. The asymmetry of MAPE is particularly relevant when forecasting electrical demand, as studies have suggested the economic impact on the Economic Dispatch problem is larger when demand is under-forecast, compared to when it is over-forecast [7]. A further problem with MAPE, relevant when forecasting low-level demand aggregations, is that it is infinite when the actual value is zero (and large when actual values are near zero). To overcome the issues with MAPE and similar measures, Hyndman and Koehler [8] proposed the use of an alternative error metric, Mean Absolute Scaled Error (MASE), which normalizes errors to those of the persistence forecast.

[9] provides a discussion on the importance of appropriate error metric selection, highlighting that "once the metric has been selected the decision as to which forecasting method to select becomes a less difficult problem." A study by Elliott *et al.* [10] considers the problem of how to optimally combine a number of forecasts under different loss functions, arguing that when a decision maker's loss to forecast errors is asymmetric, forecasts should be combined to produce a corresponding bias in the errors of the combined forecasts. Similarly, Christoffersen and

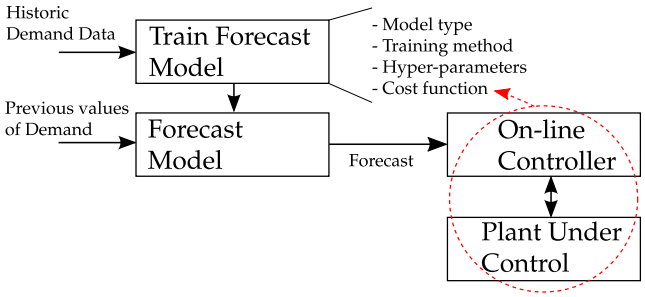

Figure 1: Block diagram showing process of producing a univariate forecast model from historic data, which produces forecasts for an on-line controller. Shown in dashed is the feedback of information from controller performance to forecast error metric presented in this paper.

Diebold [11] analytically derive optimal predictors under a restricted class of asymmetric loss functions. These methods are attractive because they provide exact results for optimal forecast combination and predictor derivation for a subclass of error metrics, but they are more limited than the frame-work presented here because they assume a loss function that can be computed one interval at a time, so do not address the double penalty effect.

1.4. Forecast Methods

A forecast method refers to both a model type, and a means of training it on (fitting its parameters to) a dataset. There are a large number of forecast methods available. This paper focuses on STLF, and in particular on univariate methods where the only available input is a time-series of previous values. Fig. 1 illustrates how a forecast model is produced from historic data, and then used in an on-line setting to produce forecasts for a controller. Historic data is first converted into a set of training instances. For a forecast origin at interval t, an instance consists of a set of predictor variables (previous values of demand, *i.e.* $\{y_{t-1}, \ldots, y_{t-n}\}$) and the realized value over the horizon (*i.e.* $\{y_t, \ldots, y_{t+N}\}$). A data-driven regression model is then fitted to this input-to-output mapping, further details of which are given in Section 2.2.

Much research is dedicated to STLF methods, Hong's dissertation provides a thorough review [12]. Some recent method comparison papers are reviewed here: in [13] Taylor gives a review of six univariate methods for forecasts up to a day ahead and concludes that double-seasonal exponential smoothing performs best, [14] offers a study of various machine learning techniques applied to forecasting hourly residential electrical consumption, and [15] gives a thorough review of the use of Feed-forward Neural Networks (FFNN) for STLF.

A large research effort, including a number of forecasting competitions (*e.g.* [16]), has been dedicated to producing the most accurate forecasts, as measured by some error metric. This paper does not propose a new forecasting method, but rather a means for customizing an error metric (which any forecasting method seeks to minimize) to a given problem instance.

1.5. On-line Controller Application

There are many types of controller which offer some robustness to forecast errors, this paper considers one of the most basic - Model Predictive Control (MPC). In MPC at each interval an optimization is performed, assuming that the forecast provided has zero error, and actions are selected for the control horizon. Only the first action is implemented, and the optimization is solved again, with updated forecasts, at the next interval. This paper provides a method for customizing a forecast error metric to a given on-line controller, and associated plant, such that when forecasts minimize that error metric they will be of most use to the controller. This feedback is shown schematically in Fig. 1.

This error metric customization is applied to optimizing the charging and discharging of a local battery in order to minimize the peak power drawn from the grid by an aggregation of customers. An application in which this objective would be useful is minimizing the peak power drawn for a set of customers (or single large customer) who have a tariff component linked to their peak demand over a billing period. Another application is the operation of a Distribution System Operator-owned energy storage asset to prevent (or delay) the need to upgrade network infrastructure which is at risk of exceeding its capacity during peak loading events. The specifics of the control problem and its formulation are given in Section 2. This problem has been addressed by Rowe *et al.* in two recent papers, [17, 18]; wherein they consider the performance of various control methods for operating a storage device for peak reduction.

2. Method

The method implemented is outlined in Fig. 1. A forecast model is trained on historic demand data for a set of customers. This forecast model is then used in an on-line setting (provided with previous values of demand) to produce forecasts which are presented to the MPC. The MPC then attempts to optimally control the plant; the control variables are the amount of energy to be sent to (withdrawn from) the battery at each interval in the horizon. For the newly proposed parametrized error metrics an additional step is necessary (see Table 1), a forecast model is produced for each set of parameter values considered, and the performance of the controller provided with forecasts from that model on the 'parameter selection' dataset is used to decide which forecast model to use on the unseen 'test' dataset.

2.1. Data

The demand data for this study are from the Irish Social Science Data Archive, [19]. It includes the energy consumed in each half-hour interval from the smart-meters of

Table 1: Use of Historic Demand data for Conventional and Parametrized forecasts

	Available Historic Data		
Conventional Forecast	Training & Validation		Testing
Parametrized Forecast	Training & Validation	Parameter Selection	Testing

4225 residential users for the period July 2009 to Dec 2010. 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).

2.2. Forecast Model Selection and Training

Two types of forecast model are considered: A Seasonal Auto-Regressive Moving Average (SARMA) model and a FFNN model. FFNN and SARMA are common STLF methods, included in seven and four (respectively) of the eight papers referenced which discuss forecasting methods ([1, 2, 7, 8, 13, 14, 15, 16]). These models allow flexibility in the selection of an error metric to be minimized. The models are outlined only briefly as they are entirely standard.

For FFNN and SARMA models parameter selection is achieved by minimizing the error metric under consideration on a training dataset, and model selection (the number of hidden nodes in the FFNN and the order of the SARMA model) is achieved by assessing the performance of the trained models on unseen validation data. Model selection is carried out only once, so that the same order of model is used for all aggregations of demand, but parameter selection is carried out for each aggregation.

A restricted class of SARMA models with zero Moving Average (MA) component is considered (as per Sevlian *et al.* [1]); giving a model for forecast, y_t as:

$$y_t = \sum_{k=1}^{p} \theta_k x_{t-k} + \sum_{k=1}^{P} \phi_k x_{t-sk} \tag{3}$$

Where x_t are previous demand values[1], p is the Auto-Regressive (AR) order of the model, $\theta_k, \forall k \in \{1, \ldots, p\}$ are the AR coefficients, s is the number of intervals within a season (here 48 to represent a 24-hour day with half-hourly data), P is the seasonal AR order of the model, and $\phi_k, \forall k \in \{1, \ldots, P\}$ are the seasonal AR coefficients. The accuracy of trained SARMA models of different orders was evaluated on an unseen validation dataset; and based on this 3$^{\text{rd}}$ order AR, and 1$^{\text{st}}$ order seasonal AR ($\text{SARMA}(3,0) \times (1,0)_{48}$) was chosen. Parameters $\{\theta_k, \phi_k\}$ are selected to minimize the error metric under consideration on the training dataset, for each demand aggregation.

The FFNN model is a single-hidden-layer network with a hyperbolic tangent activation function from input to hidden layer, and a linear (identity) activation function from the hidden layer to the output, as is commonly used in applying FFNN to regression problems. This was implemented using the MATLAB Neural-Net toolbox. However, to train (update the weights and biases of) the FFNN using custom loss functions it was necessary to modify the parts of the toolbox used in the loss and gradient calculations. The neural network inputs are taken as the actual demand from the 48 intervals leading up to the forecast origin $< x_{t-48}, x_{t-47}, \ldots, x_{t-1} >$. The output is forecast demand for the current interval and for the following 47 intervals $< y_t, y_{t+1}, \ldots, y_{t+47} >$.

A grid-search was performed to determine the number of nodes in the hidden layer, and the length of historic data used to train the FFNN: this determined that 50 hidden nodes, and 200-days of training data produced the best validation results, over a range of aggregation levels. Significantly more data is required to train the FFNN compared to the SARMA model because of the larger number of parameters. Even with 200-days of training data the model is highly parametrized. With 200-days$\times$48-intervals, there are 9600 training examples, and with 48-inputs$\times$50-hidden$\times$48-output nodes there are 4898 parameters in the weight matrices and bias vectors, resulting in a significant risk of over-fitting. To prevent this, the training data was divided into training and validation sets, and training was halted once the validation error increased more than six times in a row without a reduction.

Back-propagation training of a FFNN normally requires the loss function to be differentiable with respect to the network outputs, but the new forecast error metrics considered here (described in Section 2.3) are not differentiable. To overcome this difficulty the gradients of the error metric are computed numerically[2]; and errors are then fed back through the network using a conventional back-propagation process. This offers complete flexibility in the selection of a forecast error metric, but results in slower FFNN training.

A Naive daily Periodic (NP) forecast is also considered which assumes the demand in any interval is equal to the demand 24-hours (here 48 intervals) ago.

2.3. Error Metrics

Two potential issues with conventional forecast error metrics have already been discussed: symmetry (the relative penalty applied to over-forecasts and under-forecasts); and the 'double penalty effect' which is inherent to all single-interval error metrics. Two new parametrized error metrics are presented, which are evaluated over the whole forecast horizon; the Parametrized Earth Mover's

[1] When x_t needs to be computed for some t in the future, it is replaced with y_t; *i.e.* the model is applied recursively.

[2] Numerical gradients are found using the symmetric difference quotient; $f'(x) = (f(x+h) - f(x-h))/2h$, h is chosen as $\sqrt{\epsilon}$, where ϵ is the machine epsilon (unit roundoff).

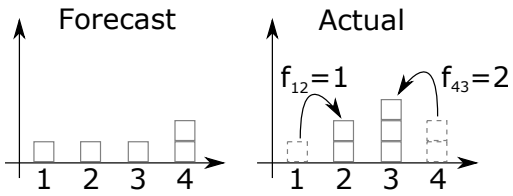

Figure 2: Illustration of the least cost EMD flows for a four-interval example

Distance (PEMD), and the Parametrized Forecast Error Metric (PFEM). The parametrization of these metrics allows them to be customized to a given on-line control problem, by appropriate selection of the parameters.

2.3.1. Parametrized Earth Mover's Distance

Earth Mover's Distance (EMD) has been used to measure the distance between probability distributions, and for image comparison [20]. EMD's name comes from the analogy of the least cost way of converting one mound of earth (here energy) into another, assuming that there is a cost for each unit of earth which is flowed (moved) from one location on the ground (interval in the horizon) to another.

In the context of forecasting, EMD computes the total cost of a minimum-cost set of flows which convert the forecast into the actual values. An illustrative example with a four-interval horizon is shown in Fig. 2, where all non-zero flows have been labeled. The cost of flows, from one interval to another, is given in the ground distance matrix; an $N \times N$ matrix (where N is the number of intervals), whose ij^{th} element gives the cost of moving a unit of energy from interval i (in the forecast), to interval j (in the actual).

When comparing forecast demand to actual values, there is no guarantee that the integrals over the horizon match; so it is necessary to extend the EMD to allow energy to be created or destroyed at some cost. A Parametrized EMD (PEMD) is proposed; if f_{ij} is the flow of energy from interval i to interval j, then the cost associated with it is:

$$\text{Cost}(f_{ij}) = \begin{cases} f_{ij} \min(c(j-i), d) & \text{if } i < j < N+1 \\ f_{ij} ab & \text{if } i < j = N+1 \\ f_{ij} \min((i-j), d) & \text{if } j < i < N+1 \\ f_{ij} a & \text{if } j < i = N+1 \end{cases} \quad (4)$$

An artificial interval, $N+1$, is used for cases when energy needs to be created or destroyed. The total cost of flows is minimized subject to constraints which ensure the flows convert the forecast into the actual values, the implementation of this optimization is as per Pele and Werman [21].

In (4) the cost of moving a unit of energy to an interval, from the following interval, is 1 (an arbitrary factor setting the overall scale of the metric). The error metric parameters can then be described as:

a the cost of adding a unit of energy to any interval;

b the cost of removing a unit of energy from any interval, relative to a;

c the cost of moving a unit of energy to an interval, from the previous interval;

d a threshold limiting the maximum value of elements of the ground distance matrix.

The following observations can be made on the interpretation of these parameters: if $a > 1$, then the cost of adding energy at any interval is greater than that of moving it from the following interval; if $b < 1$ over-forecasts are penalized less severely than under-forecasts; if $c < 1$ then moving forecast energy to later intervals is less costly than moving it to earlier intervals (this seems appropriate as MPC can re-optimize for future events, but not for past ones).

2.3.2. Parametrized Forecast Error Metric

A second error metric has been developed by the authors and attempts to capture some of the characteristics of discrepancy between forecast and actual demands which are likely to be of importance to forecasts produced for a MPC. This is termed the Parametrized Forecast Error Metric (PFEM), and it has the following parameters:

α The cost of under-forecasts relative to over-forecasts;

β Error exponent, *i.e.* the relative importance of large errors compared to small ones;

γ The ratio of the cost of errors in the first interval compared to the final interval (with linear scaling in between);

δ The allowable permutation radius in finding a minimum cost value.

The forecast error metric is then computed for an N-step horizon as:

$$\text{PFEM}(\boldsymbol{e}) = \min_{\text{radius } \delta} \left[\frac{1}{N} \sum_{t=1}^{N} \left(\frac{[e_t]^- + [e_t]^+ \alpha}{(1+\alpha)/2} w_t \right)^{\beta} \right]^{1/\beta} \quad (5)$$

$$w_t = \frac{2}{(\gamma+1)(1-N)} \left[t(\gamma-1) + (1-\gamma N) \right] \quad (6)$$

Where $e_t = x_t - y_t$ is the error in the t-th interval, $\boldsymbol{e}$ is a vector of errors over the horizon, the $[\,]^+$ operator takes the positive component (*i.e.* $[x]^+ = \max(0, x)$, similarly $[x]^- = \max(0, -x)$). w_t are the weights applied to each interval in the horizon as defined in (6), where the first factor ensures the weights sum to N. The minimization over the permutation radius, δ, computes the error metric at all

permutations of the forecast, in which each y_t is moved at most δ intervals earlier or later, and selects the minimum value. δ is equivalent to the 'adjustment limit' described in [5]. With the parameters $(\alpha = 1, \beta = 2, \gamma = 1, \delta = 0)$ PFEM becomes the root-mean-squared-error. Note that the proposed error metrics are not symmetric, for example PFEM($\boldsymbol{e}$) $\neq$ PFEM($-\boldsymbol{e}$), therefore, they are not mathematically true distance metrics.

These error metrics were formulated to have parameters which should allow them to be customized to a wide range of control problems. For example, problems which suffer disproportionately as a result of under-forecasts should have $\alpha > 1$, and problems in which large errors dominate over smaller ones should have $\beta > 1$, and so on. However, referring to the desirable features of error metrics discussed in Section 1.3 they take longer to compute, and are less interpretable than conventional error metrics. There is a trade-off between simplicity and performance, but in applications where forecasts are being produced for use by a controller (rather than a person) interpretability is not as critical.

2.4. Model Predictive Controller

The control problem considered is that of operating (charging and discharging) a battery with the objective of minimizing the peak power drawn over a billing period by a set of customers. This is shown schematically in Fig. 3. The problem is solved using MPC, in each interval the following linear program is solved:

Minimize:

$$E \tag{7}$$

Subject to:

$$\sum_{t=1}^{\tau} c_t \leq B - C_{\text{init}} \quad \forall \tau \in \{1, \ldots, N\} \tag{8}$$

$$\sum_{t=1}^{\tau} c_t \geq -C_{\text{init}} \quad \forall \tau \in \{1, \ldots, N\} \tag{9}$$

$$E \geq c_t + y_t - E_{\text{bill}} \quad \forall t \in \{1, \ldots, N\} \tag{10}$$

$$-c_{\text{lim}} \leq c_t \leq c_{\text{lim}} \quad \forall t \in \{1, \ldots, N\} \tag{11}$$

Where:

E is the amount by which the peak power over the horizon exceeds the running peak of the current billing period. Power is measured in units of energy consumed over a single interval [kWh];

c_t is the energy to send to the battery during interval t - the decision variables [kWh];

N is the number of intervals in the forecast and planning horizon, for this instance $N = 48$;

C_{init} is the amount of energy in the battery at the start of the horizon for which the optimization is being run, above the minimum allowable charge-state [kWh];

y_t is the forecast demand during interval t [kWh];

E_{bill} is the highest per interval demand within the billing period so far [kWh];

c_{lim} is the maximum energy which can be transferred to/from the battery within an interval [kWh];

B is the use-able energy capacity of the battery [kWh].

(7) ensures that the controller's objective (minimizing the peak power drawn from the grid) is optimized; (8) and (9) ensure that the upper and lower state-of-charge constraints of the battery are not violated in any interval, respectively; (10) 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); (11) ensures the energy exchanged with the battery in an interval is within the battery's rate-of-charge constraints. Note that as formulated E is allowed to be negative, *i.e.* solutions which offer some margin from setting a new peak demand are favored.

This optimization outputs the optimal charging energies for the battery, c_t, over the next $t \in \{1, \ldots, N\}$ periods assuming that the forecasts, y_t have zero error. The MPC only implements the first of these decisions c_1, and the optimization is re-run at the next interval, with a new forecast. The optimization is solved for a 24-hour horizon at a time, but the billing period (over which the peak power, E_{bill}, is determined) is 7-days. Before the next horizon is solved, the running peak demand, E_{bill}, is updated as required (*i.e.* increased if grid import during the implemented interval exceeds E_{bill}, and reset if a billing cycle is over).

Set-point-based recourse of the optimal charging decision, c_1, is also considered. If the actual demand during the current interval is such that implementing c_1 would create a new grid-demand peak, in excess of the peak which the MPC optimization expects over its horizon ($E_{\text{bill}} + E$), then c_1 is adjusted to prevent this. More formally, the implemented charging decision during the current interval is: $\hat{c_1} = \min(c_1, E_{\text{bill}} + E - d_1)$, where d_1 is the actual demand during the current time-period, and $\hat{c_1}$ is then subjected to the batteries rate and state of charge constraints.

2.5. Set-Point Controller

A naive set-point (SP) controller benchmark is considered, and its performance is compared to that of the forecast-based controller. Subject to state-of-charge and rate-of-charge constraints, in any interval it implements the following battery charging decision $\hat{c}_1$:

$$\hat{c}_1 = E_{\text{bill}} - d_1 \tag{12}$$

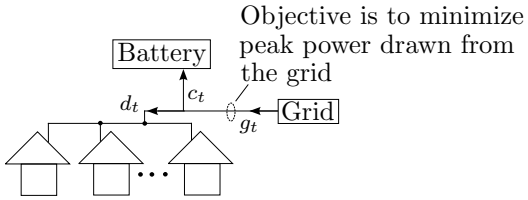

Figure 3: Block diagram showing control problem considered, and the positive sign convention for energy flows.

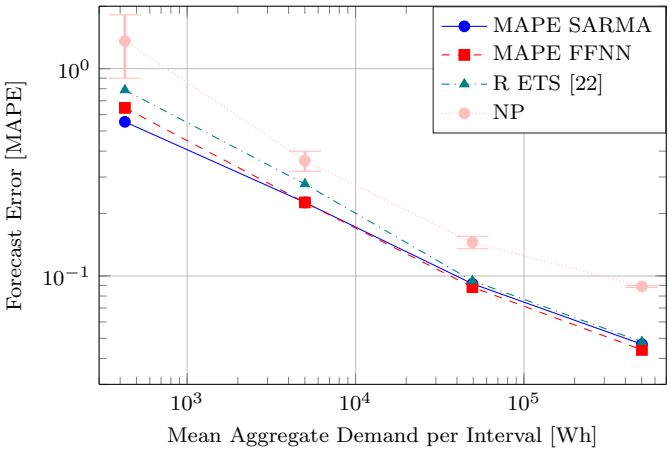

Figure 4: Test one-step MAPE for FFNN and SARMA forecast models trained to minimize MAPE, plotted against mean load of aggregation. Each point shows the average of 6 aggregates of {1, 10, 100, 1000} households. Performance of the NP forecast is plotted for reference (error bars showing $\pm$ 1.0σ).

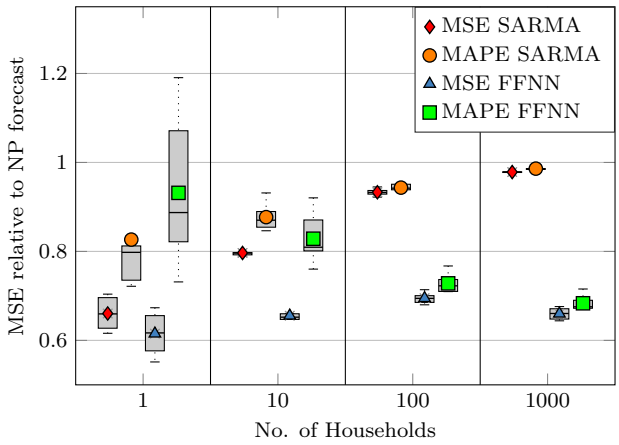

Figure 5: Test MSE for FFNN and SARMA forecast models trained to minimize MSE and MAPE. The MSE relative to that achieved using a NP forecast is plotted. Symbols show the mean, the box shows the inter-quartile range.

3. Results and Discussion

3.1. Forecasting

Fig. 4 presents the forecasting results achieved by SARMA and FFNN models optimized to minimize MAPE; showing that at aggregation levels of 5×10^3 Wh/interval (10 houses) and 5×10^5 Wh/interval (1000 houses), MAPE is approximately 22% and 4.5% respectively. These values

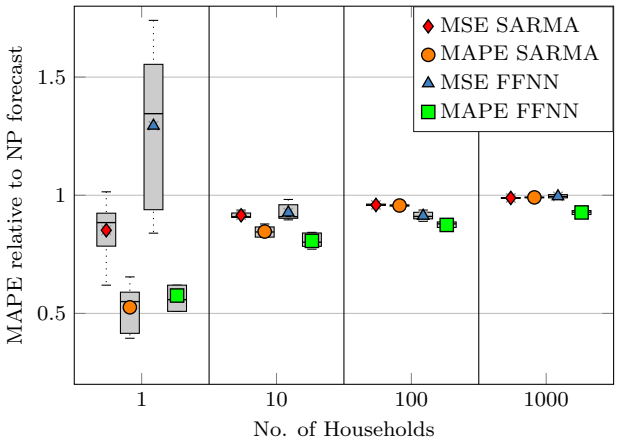

Figure 6: Test MAPE for FFNN and SARMA forecast models trained to minimize MSE and MAPE respectively. The MAPE relative to that achieved using a NP forecast is plotted. Symbols show the mean, the box shows the inter-quartile range.

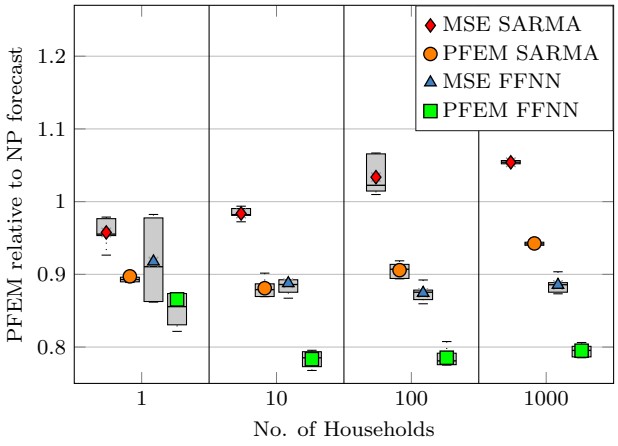

Figure 7: Test PFEM with arbitrary parameters ($\alpha = 2, \beta = 2, \gamma = 2, \delta = 1$) for FFNN and SARMA forecast models trained to minimize PFEM and MSE. The PFEM is shown relative to PFEM achieved using a NP forecast.

are consistent with the performance of methods presented in [1][3], demonstrating that the implementations of the forecasting methods presented here provide performance consistent with the literature (as measured by MAPE), over a range of aggregation levels. Also shown, for reference, in Fig. 4 is the performance of an automated time-series forecasting package[22], although it has not been customized to minimize MAPE.

Figs. 5 to 8 show the performance of the SARMA and FFNN forecast models trained to minimize, and evaluated against, four different error metrics. The labeling of the forecast models is as follows: "MSE SARMA" indicates that a SARMA model was used, and that its parameters were selected to minimize MSE on the training dataset.

[3]Fig. 6 of [1] shows MAPE ranges of 12.8% to 32% at an aggregation of 1×10^4 Wh, and 2.6% to 4.5% at an aggregation of 1×10^6 Wh at a 1-hour interval. The MAPEs achieved here at equivalent aggregation levels are 22% and 4.5% which lie in these ranges.

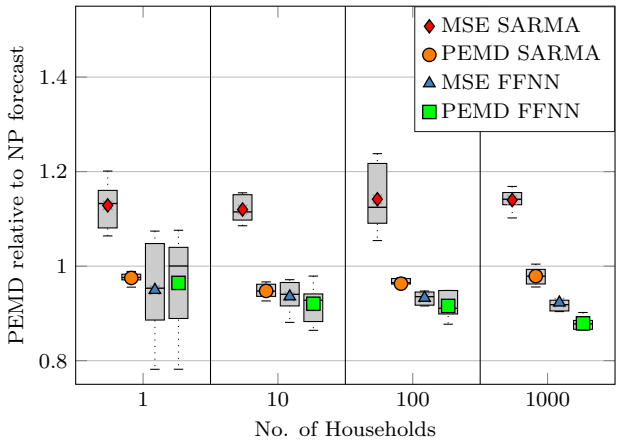

Figure 8: Test PEMD with arbitrary parameters ($a = 10, b = 0.5, c = 0.5, d = 4$) for FFNN and SARMA forecast models trained to minimize PEMD and MSE. The PEMD is shown relative to PEMD achieved using a NP forecast.

The data presented in Figs. 5-8 show the performance of the resulting model on an unseen test dataset, as measured by the error metric on the y-axis of each plot, relative to the performance of the NP forecast measured by that metric. At all levels of aggregation the forecasts trained to minimize a given error metric on the training data, perform well as measured by that metric on the test data. Figs. 7 and 8 demonstrate that forecast models trained to minimize PFEM/PEMD (with a particular set of parameter values) will produce low errors using metrics with the same parameter values on an unseen test data-set. FFNN forecast models were used for the remainder of the study, given their better test performance as measured using the PFEM and PEMD metrics.

Note that Figs 5-8 do not show the superiority of any of the metrics considered over any of the others; they only show that for a given metric (*e.g.* a given parameterization of the PEMD or PFEM metric), forecasts trained to minimize that metric on historic training data, perform well according to that metric on unseen test data. This is important to establish before we attempt to customize the parameters of the PFEM/PEMD metrics to particular applications.

3.2. Application to Model Predictive Control

Fig. 9 shows the performance of the MPC in reducing the peak load of a collection of customers (plotted as the peak reduction ratio, *i.e.* $1 - E_{\text{bill}}/\max(d_t)$, where E_{bill} is the maximum power drawn from the grid over the billing period[4], and $\max(d_t)$ is the maximum local demand over the same period), when provided forecasts from a FFNN model trained to minimize the error metrics shown in the legend. The objective of the controller is to minimize peak

[4]A billing period of 7-days is considered, the results show average peak reduction over 8 weeks of operation.

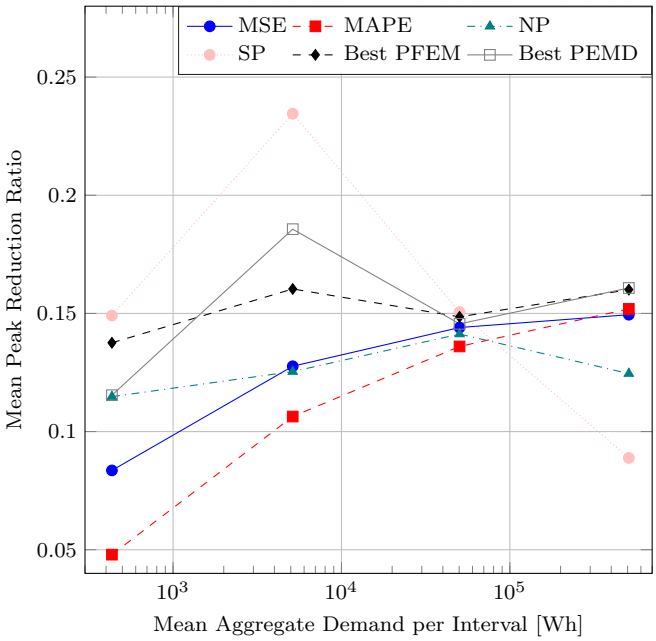

Figure 9: Performance of MPC provided with forecasts from FFNN models trained to minimize various error metrics, with mean peak reduction ratio plotted against mean load of aggregation. The 'Best PFEM/PEMD' forecasts are trained to minimize PFEM/PEMD, with parameters selected based on each forecast's performance on part of the training data. A grid-search was performed over parameters $\alpha = \{0.5, 1, 2\}$, $\beta = \{1, 2\}$, $\gamma = \{1, 2\}$, $\delta = \{0, 1\}$, and $a = \{50, 100\}$, $b = \{0.5, 1, 2\}$, $c = \{0.5, 1\}$, $d = \{5, 10\}$.

power demand (here measured as peak energy within a single interval), we plot the peak reduction ratio to allow us to compare results between different customers and across aggregation levels. The battery capacity is set at 5% of the average daily demand of the selected aggregation. Increasing battery capacities were considered and showed similar relative performance of the methods until peak reduction ratio began to saturate at the peak-to-average ratio of the demand aggregation. The PFEM and PEMD-minimizing forecasts (which were trained on a training data set) were evaluated with a simulation of the controller on an error metric parameter selection dataset, and that which performed best was chosen. The performance of this selected error metric (and its associated forecast model) was then evaluated on an unseen test dataset, and this is the performance shown in Fig. 9. The division of training data for conventional and parametrized forecast error metrics is shown in Table 1. The parameter values considered in the error-metric grid-search were chosen based on judgment; the number of values for each parameter had to be small, to keep the time required to train the forecast models manageable. Note that typical current practice is to arbitrarily assume that a conventional error-metric (such as mean-squared-error) is appropriate, hence consideration of even a small set of forecast error metric parameterizations is an improvement.

The following observations can be made regarding Fig. 9:

Table 2: Mean PFEM Parameter Values Selected at Different Aggregation Levels

No. of Customers	α	β	γ	δ
1	1.7	2.0	1.8	1.0
10	1.8	1.7	1.7	0.33
100	1.8	1.5	1.5	0.33
1000	1.8	1.7	1.3	0.83

- The naive set-point (SP) controller performs well for small aggregations of demand, where forecasts are least accurate, causing forecast-based methods to be less effective;
- A controller provided with a naive periodic (NP) forecast also performs well, despite those forecasts having larger errors according to conventional metrics (*e.g.* MSE and MAPE);
- A controller provided with a Best PFEM-minimizing forecast consistently out-performs that provided with an MSE-minimizing forecast (by 45% on average);
- Similarly, a controller provided with a Best PEMD-minimizing forecast out-performs that provided with an MSE-minimizing forecast (by 38% on average);
- MAPE is a poor choice of forecast error metric for this application, especially at smaller demand aggregations, because it applies a heavier penalty to over-forecasts (so minimizing it produces models which under-forecast).

For clarity only mean peak reduction ratios at each aggregation level are plotted in Fig. 9. There is variation in the performance improvement offered between the 6 different random aggregations of customers (of each size) considered. A one-sided Wilcoxon signed rank test rejects a null hypothesis (that Best PFEM-minimizing forecasts do not improve peak reduction ratio compared to MSE-minimizing forecasts) with a p-value of 0.0003.

Table 2 shows the mean parameter values selected for the Best PFEM forecasts at different levels of demand aggregation. The PFEM parameter α is consistently greater than 1, suggesting under-forecasts should be more heavily penalized for the peak-reduction problem, as might be expected. This is consistent with the poor performance of controllers provided with MAPE-minimizing forecasts.

4. Conclusion

This paper has presented arguments for, and an empirical demonstration of, the benefits of customizing the error metric of forecast models to the control problem their forecasts will be used by. Using two new parametrized forecast error metrics; PFEM and PEMD, it has been shown that, using some of the available training data, one can select suitable parameters of a forecast error metric in order to produce more useful forecasts for a given on-line control problem.

This forecast error metric customization approach has been shown to offer a performance improvement compared to a MPC provided with a forecast from a model trained to minimize a conventional error metric. In the application considered, a controller attempts to minimize the peak power drawn by an aggregation of customers over a billing period. Appropriate selection of the PFEM parameters improved the peak-reduction achieved by a MPC by 45%, averaged over demand aggregation levels between 1 and 1000 households, compared to a MPC provided with a MSE-minimizing forecast.

5. Further Work

The concept of customizing forecast error metrics is generalizable to problems where a forecast is produced for a single on-line controller which can be accurately simulated, and it needs to be empirically evaluated against more problems to understand when and where it is most effective. Future work will extend the approach to consider RES supply forecasts, multi-variate forecasts, and probabilistic forecasts.

This work has not considered the impact of operating the energy storage system on the distribution network. A study considering this, and the impact on energy storage operation of being subjected to network constraints such as voltage magnitude, and capacity limits would be a useful extension of this work.

6. Appendix

Figs. 10 & 11 demonstrate the performance of the error-metric specific forecasts for a particular parameterization of PFEM and PEMD error-metrics, compared to forecasts which minimize a conventional (MSE) metric, when applied to the net demand of several homes in Australia which have roof-top PV systems [23]. It illustrates that the forecasts with customized error metrics still perform well for demand signals which are influenced by the presence of behind-the-meter generation, which will be an increasingly important part fo the generation mix moving forwards.

Acknowledgment & Source Code

This work was supported through access to computing resources provided by Research Platform Services at the University of Melbourne. Khalid did this work with the support of MIRS & MIFRS scholarships. We would like to thank two anonymous reviewers for their valuable comments.

All source code used in this study has been uploaded to an open source repository, [24] (the data needs to be accessed separately [19], as it is available only on request).

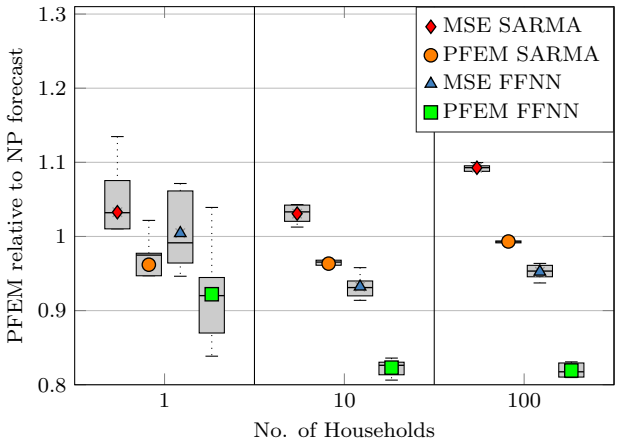

Figure 10: Test PFEM with arbitrary parameters ($\alpha = 2, \beta = 2, \gamma = 2, \delta = 1$) for FFNN and SARMA forecast models trained to minimize PFEM and MSE. The PFEM is shown relative to PFEM achieved using a NP forecast. Results are for forecasting the net demand of households with rooftop PV systems.

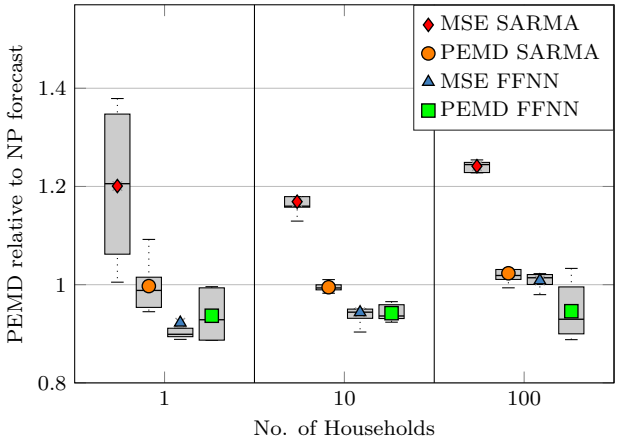

Figure 11: Test PEMD with arbitrary parameters ($a = 10, b = 0.5, c = 0.5, d = 4$) for FFNN and SARMA forecast models trained to minimize PEMD and MSE. The PEMD is shown relative to PEMD achieved using a NP forecast. Results are for forecasting the net demand of households with rooftop PV systems.

References

[1] R. Sevlian, R. Rajagopal, Short term electricity load forecasting on varying levels of aggregation, arXiv preprint arXiv:1404.0058v2.
URL https://arxiv.org/abs/1404.0058v2

[2] P. Mirowski, S. Chen, T. K. Ho, C.-N. Yu, Demand forecasting in smart grids, Bell Labs Technical Journal 18 (4) (2014) 135–158. doi:10.1002/bltj.21650.

[3] N. Amjady, F. Keynia, H. Zareipour, Short-term load forecast of microgrids by a new bilevel prediction strategy, IEEE Transactions on Smart Grid 1 (3) (2010) 286–294. doi:10.1109/TSG.2010.2078842.

[4] A. H. Murphy, What is a good forecast? an essay on the nature of goodness in weather forecasting, Weather and Forecasting 8 (2) (1993) 281–293. doi:http://doi.org/bbpk8s.
URL http://doi.org/bbpk8s

[5] S. Haben, J. Ward, D. Vukadinovic Greetham, C. Singleton, P. Grindrod, A new error measure for forecasts of household-level, high resolution electrical energy consumption, International Journal of Forecasting 30 (2) (2014) 246–256. doi:10.1016/j.ijforecast.2013.08.002.

[6] S. Kolassa, R. Martin, Percentage errors can ruin your day (and rolling the dice shows how), Foresight: The International Journal of Applied Forecasting Fall (2011) 21–27.

[7] B. F. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, D. J. Maratukulam, Analysis of the value for unit commitment of improved load forecasts, IEEE Transactions on Power Systems 14 (4) (1999) 1342–1348. doi:10.1109/59.801894.

[8] R. J. Hyndman, A. B. Koehler, Another look at measures of forecast accuracy, International Journal of Forecasting 22 (4) (2006) 679 – 688. doi:http://dx.doi.org/10.1016/j.ijforecast.2006.03.001.

[9] B. E. Flores, A pragmatic view of accuracy measurement in forecasting, Omega International Journal of Management Science 14 (2) (1986) 93–98. doi:10.1016/0305-0483(86)90013-7.

[10] G. Elliott, A. Timmermann, Optimal forecast combinations under general loss functions and forecast error distributions, Journal of Econometrics 122 (1) (2004) 47–79. doi:10.1016/j.jeconom.2003.10.019.

[11] P. F. Christoffersen, F. X. Diebold, Optimal Prediction Under Asymmetric Loss, Econometric Theory 13 (06) (1997) 808. doi:10.1017/S0266466600006277.

[12] T. Hong, Short term electric load forecasting, (thesis) (September 2010).
URL http://www.lib.ncsu.edu/resolver/1840.16/6457

[13] J. W. Taylor, L. M. de Menezes, P. E. McSharry, A comparison of univariate methods for forecasting electricity demand up to a day ahead, International Journal of Forecasting 22 (1) (2006) 1–16. doi:10.1016/j.ijforecast.2005.06.006.

[14] R. E. Edwards, J. New, L. E. Parker, Predicting future hourly residential electrical consumption: A machine learning case study, Energy and Buildings 49 (2012) 591 – 603. doi:10.1016/j.enbuild.2012.03.010.

[15] H. S. Hippert, C. E. Pedreira, R. C. Souza, Neural networks for short-term load forecasting: a review and evaluation, IEEE Transactions on Power Systems 16 (1) (2001) 44–55. doi:10.1109/59.910780.

[16] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012, International Journal of Forecasting 30 (2) (2014) 357–363. doi:10.1016/j.ijforecast.2013.07.001.

[17] M. Rowe, T. Yunusov, S. Haben, C. Singleton, W. Holderbaum, B. Potter, A peak reduction scheduling algorithm for storage devices on the low voltage network, IEEE Transactions on Smart Grid 5 (4) (2014) 2115–2124. doi:10.1109/TSG.2014.2323115.

[18] M. Rowe, T. Yunusov, S. Haben, W. Holderbaum, B. Potter, The real-time optimisation of DNO owned storage devices on the LV network for peak reduction, Energies 7 (6) (2014) 3537. doi:10.3390/en7063537.

[19] ISSDA, CER smart meter customer behaviour trials data, accessed via the Irish Social Science Data Archive, ver. CER Electricity Revised March 2012.
URL www.ucd.ie/issda

[20] Y. Rubner, C. Tomasi, L. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision 40 (2) (2000) 99–121. doi:10.1023/A:1026543900054.

[21] O. Pele, M. Werman, Fast and robust earth mover's distances, in: Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 460–467. doi:10.1109/ICCV.2009.5459199.

[22] R. J. Hyndman, forecast: Forecasting functions for time series and linear models, R package version 7.1 (2016).
URL http://github.com/robjhyndman/forecast

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

[24] K. Abdulla, FEMC: Forecast error metric customization, https://github.com/khalida/femc (2015).