Novel autoregressive basis structure model for short-term forecasting of customer electricity demand

I. INTRODUCTION

To reduce long-run capital infrastructure costs and to mitigate the negative ramifications of the introduction of renewable energy production technologies in electricity generation, transmission and distribution networks, Static Synchronous Compensators (STATCOM) can be integrated into the transmission and distribution network. STATCOMs, are four quadrant pure sine wave synchronous inverters. STATCOMs are able to import and export real and reactive power, modulate frequency distortions and eliminate harmonics. Combined with battery energy storage, STATCOMs can store energy produced by wind generators and solar PV arrays and release power when it is demanded and mitigate LV distribution network voltage non-conformance.

The integrated energy management control algorithm is to take into account the specifications and functionality of the STATCOM, physical characteristics of the battery energy storage system, the extent at which peak demand is to be reduced, the instantaneous conditions of the LV distribution network and also electricity demand forecasts. Given the continued trend towards smart-grid technology implementation, electricity demand information in the future will be recorded throughout the network including high resolution residential and commercial electricity demand data. The energy management control algorithm will be required to read electricity demand data from each connected component, then in real-time formulate and continuously refine future electricity demand forecasts in order to optimise energy resource allocation. This paper will cover recent progress on the development of an electricity demand statistical modelling technique that will form the energy management control algorithm's forecasting component.

II. BACKGROUND

Griffith University (GU) in conjunction with ENERGEX (electricity distribution company) are working on a joint project to examine the feasibility of installing STATCOMs with battery energy storage in the LV distribution network in South East Queensland, Australia. The objective of the this joint project is to determine the effectiveness of STATCOMs to reduce expenditures associated with upgrading LV distribution network infrastructure and costs associated with infrastructure construction at greenfield sites. It is anticipated that the installation of STATCOMs with energy storage will achieve this objective by reducing peak demand through load smoothing and by also improving power quality issues (e.g. reactive power) in the network. STATCOMs with battery energy storage will be considered to be feasible if they prove to be a financially viable alternative to the current regime of infrastructure upgrade triggers and current methods of providing infrastructure at greenfield sites.

ENERGEX selected a subset of the LV distribution network for a pilot project. The pilot project involves high resolution electricity metering of a large number of residential premises, the metering of a number of distribution transformers, integrating STATCOMs with battery energy storage and installing a communication network. Larger grid-based STATCOMS with energy storage as well as numerous distributed smaller residential STATCOMS with energy storage (10-20kVA) will be trialed in the pilot project. Data collected from the participating residential premises, transformers and STATCOMs is stored centrally which allows for data analysis and control operations.
A. Short-term Electricity Demand Forecasting

The duration of time that the STATCOM can intelligently discharge energy resources is limited by the capacity of the energy storage. The capacity of the battery energy storage component of the STATCOM is limited by the high cost of robust lithium-ion batteries. This economic constraint limits the cost-effective sizing of energy resources, thus forecasting methods applicable to the problem at hand are related to the immediate to short-term horizon. Typically forecasts ranging from 5 minutes ahead to a week ahead are required. The most common statistical modelling approaches used to develop short-term forecast models include time series analysis techniques and artificial neural networks (NN) [1 – 5].

B. Representative Publications

Reference [1] employed the use of NN to forecast electricity price in a competitive market. The NN employed was a three layered feed-forward network trained by the Levenberg-Marquardt algorithm. An ARIMA model was developed to compare results. Both models were constructed by use of hourly historical price of 42 days prior to the week being forecasted. Error in the NN ranged from 5.23% to 13.65% and error in the ARIMA ranged from 6.32% to 13.78%. For the forecasted periods it was found that the NN outperformed the ARIMA model.

Reference [2] engaged in an inquiry to determine whether or not NN yield better short-term (hourly) time series forecasts than more conventional methods such as the Autoregressive Integrated Moving Average (ARIMA). Both modelling techniques (NN and ARIMA) were examined in the inquiry. The method that was employed was first to identify whether or not there were linear or non-linear correlations between the dependent and independent variables. The autocorrelation function was used identify linear correlations and a normalisation of the mutual information function was used to determine non-linear correlations. It was observed that the system was non-stationary and dominated by linear correlations. The ARIMA model was differenced multiple times to achieve stationarity. The inquiry found that the forecasting ability of the ARIMA and the NN were not notably different with normalised mean square errors at the 12th horizon of 2.8% and 3.3%.

Reference [3] published a literature review of the application of neural networks for short-term load forecasting. The review identified that the most common input variables were electricity load, temperature, humidity, weather and load parameters. They also found that ANN could be used to calculate short-term electricity demand in a variety of ways including:

- Forecasting next hour's electricity demand.
- Forecasting electricity demand for a sequential set of hours by use of multiple output neurons.
- Forecasting next day's electricity demand or peak load.

Reference [4] developed a method of forecasting southeast Brazil's hourly electricity demand and compared the results against a benchmark model. The development of the forecast model involved treating each hour of the day as a separate time series. Each model is based on a two-step decomposition of the time series. The first step is comprised of Fourier series, dummy variables and a linear trend. The purpose of the first step is to integrate seasonality, day of the week and special event effects into the forecast model. The second step involved estimating linear autoregressive models by regression and NN. The benchmark model in this study was a modified version of the Seasonal Integrated Autoregressive Moving Average (SARIMA) model. It was found that their proposed model outperformed the benchmark model and that the second step NN estimation performed comparatively poorly at horizon forecasting.

Reference [5] used the ARIMA method with double seasonal exponential smoothing to develop a short-term electricity demand forecast model. The construction of the double seasonality ARIMA model involves the addition of differenced seasonal modifiers such that the model takes on the ARIMA(p,d,q)(P,D,Q)(p,D,Q) form. The (P,D,Q) modifiers denote the seasonality. Double exponential smoothing of seasonality involves using Holt-Winters method of smoothing the ratio of observed values to the differenced local level such to estimate two seasonality indexes. The ARIMA double seasonal exponential smoothing was compared against permutations of seasonal ARMA and found that the ARIMA double seasonal exponential smoothing performed better.

IV. AIMS

The aim of this research was to develop a 30 minute ahead electricity demand forecasting method for individual residential premises and distribution transformers for use as a forecasting component as part of the energy management control algorithm.

V. DATA

To develop the forecasting faculties of the energy management control algorithm, electricity demand data from 127 residential premises and 1 LV distribution transformer was analysed. This electricity demand data was collected and supplied by ENERGEX's smart-grid research division. The 127 residential premises and the low voltage distribution transformer reside in an inner-north suburb of Brisbane, Australia (trial area). Each residential premise had high resolution electricity meter installed which records 1-minute averaged voltage, current and phase angle at 1 minute intervals. The LV distribution transformer has its voltage, current and phase angle recorded at 30 minute intervals, averaged over the 30 minute period, for each phase of the Y-configuration. For purposes of describing the development of the prototype method of creating forecast models, Q1, 2012 (i.e Jan-Mar 2012) data from one residential premises and a low voltage distribution transformer will be used.
The data collected from residential premises in the trial area is in 1 minute periods. For the construction of the forecast models, the 1 minute intervals are averaged over 30 minute intervals. For an example residential daily electricity demand profile, Figure 1 below displays the 1 minute interval electricity demand profile and the 30 minute interval electricity demand profile. Averaging electricity demand over thirty intervals decreases the variance of the electricity demand profile by 14%. The reduction in variance is accounted by the smoothing of the frequent peaks (i.e. short term high power consumption devices in homes such as kettles) in the electricity demand profile. In other terms, the diversity factor is reduced. The integrals of both electricity demand profiles result in the same value of 30.96 kWh. This result indicates that energy consumed information is not lost when the data is averaged.

![Demand Profile Comparison](image)

### VI. PROTOTYPE FORECAST MODELLING METHOD

#### A. Model

Autoregressive time series forecast models based solely on the lagged dependent variable result in a model where the forecasted dependent variable fluctuates around a general trend conforming to a constant variance. An autoregressive time series forecast models can be built upon a basis function. Under such a model, the lagged dependent variable based coefficients determine the deviation from the basis function. By observation, the daily electricity demand profile is a periodic function. The method of identifying the underlying periodic functions that would comprise the observed demand profile is Fourier Analysis. Regression or another method is then used to calculate the value of the coefficients within the forecast model.

Engaging in the Fourier Analysis process, as stipulated above, will only result in describing what is already known, the basis. In addition, it needlessly increases the complexity of the modelling process. The prototype method of creating forecast models will be founded on the use of data comprising a daily electricity demand profile as the basis structure. The forecasting component will involve forecasting the deviation of future electricity demand from the basis structure. The forecasting component's models will be determined by the Autoregressive method. Once the future deviation has been forecast, the deviation will be added to basis structures in order forecast future demand.

The general forecast model produced by the forecast model prototype method is as denoted by equation 1.

\[ y_t = BS_t + \text{diff}(x_{t-1}, x_{t-2}, \ldots, x_{t-k}) \]  

(1)

where \( y_t \) is the electricity demand forecasted at time \( t \), \( BS \) is the basis structure of the forecast model, \( BS \) is the electricity demand of the basis structure at time \( t \), \( \text{diff}() \) is the electricity demand difference forecast model, \( x_t \) is the electricity demand difference between the basis structure and daily electricity demand profile at time \( t \), \( x_{t,l} \) is the electricity demand difference between the basis structure and daily electricity demand profile at time \( t \) lagged by \( l \) and \( k \) is the maximum number of lags.

The underlying hypothesis of the general forecast model is that there is a general trend of electricity demand within the residence. In other terms, it is assumed that deviations in electricity demand in the past will be positively correlated with deviations in electricity demand in the future. If electricity demand is greater than basis structures it is assumed that the increased electricity demand will continue for the next 30 minute interval. A similar pattern is assumed for a lesser electricity demand.

#### B. Method

The following four steps outline the general process of the prototype method for creating forecast models:

1. Separation of data into weekday and weekend daily electricity demand profile categories. This necessitates the condition of integrating multiple forecast models into the energy management control algorithm.
2. Selection of basis structures.
3. Calculation of the differences between the recorded daily electricity demand profiles and the basis frames.
4. Use of the autoregressive method to produce electricity demand difference forecast models.

#### C. Selection of Basis Structure

The selection of basis structures is determined by first defining an electricity demand profile matrix, \( DP \), as containing each daily electricity demand profile for either the weekday or weekend category. Each row of the array denotes a different daily demand profile and the column of the array is denoted by the 30 minute interval time step. The next stage in the process is to calculate the \( f \)-statistic and Pearson's correlation statistic (\( \rho \)) between each daily electricity demand profile and every other electricity demand profile resulting in matrices containing the results. The \( f \)-statistic is calculated by equation 2 and the Pearson's correlation statistic being calculated by equation 3.

\[ f_{\text{stat}}(X, Y) = \frac{\text{variance}_{\text{larger}}}{\text{variance}_{\text{smaller}}} \]  

(2)

where in this specific equation \( X \) is a vector containing a particular daily electricity demand profile, \( Y \) is a vector containing a different daily electricity demand profile, the variance for both \( X \) and \( Y \) is calculated, the larger variance
is the numerator and the smaller variance is the
denominator. The resultant \( f \)-stat matrix, \( F \), is an \( n \) by \( m \)
matrix where \( n \) is the number of rows and \( m \) is the number
of columns. Each specific row and column index
permutation pertains to the \( f \)-stat calculation result.

\[
\rho(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{(x_i - \mu_x)(y_i - \mu_y)}{\text{std}(X)\text{std}(Y)} \right)
\]

where in this specific equation \( X \) is a vector containing
a particular daily electricity demand profile from \( DP \), \( Y \) is a vector containing a different daily electricity demand
profile from \( DP \), \( x \) is the electricity demand value for an
individual time step in \( X \), \( u_x \) is the mean of \( X \), \( u_y \) is the
mean of \( Y \) and \( \text{std}(•) \) is the standard deviation function.
The resultant correlation matrix, \( P \), is an \( n \) by \( m \) matrix,
where \( n \) is the number of rows and \( m \) is the number
of columns. Each specific row and column index permutation
pertains to the correlation calculation result.

The rows of the \( f \)-stat matrix and correlation matrix are
summed resulting in two summation vectors as described
by equations 4 and 5. The purpose of summing the result
matrices is to identify the daily electricity demand profile
which is most similar to every other daily electricity
demand profile.

\[
[f_{ik}]_{k=1}^m = \sum_{j=1}^n F_{jk}, \text{ for } k = 1,2, \ldots , m \tag{4}
\]

where \( f_i \) is the summation of columns in \( F \), \( F \) is the \( f \)-
stat matrix, \( n \) is the number of rows in \( F \) and \( m \) is the
number of columns in \( F \).

\[
[\rho_{ik}]_{k=1}^m = \sum_{i=1}^n P_{ik}, \text{ for } k = 1,2, \ldots , m \tag{5}
\]

where \( \rho_i \) is the summation of the columns in \( P \), \( P \) is the
correlation matrix, \( n \) is the number of rows in \( P \) and \( m \) is the
number of columns in \( P \).

The location index of the element in \( f_i \) which has the
least value denotes the daily electricity demand profile
which has the most similar variance with every other daily
electricity demand profile. The index of the element in \( \rho_i \)
which has the greatest positive value denotes the daily
electricity demand profile which most positively correlates
with every other electricity demand profile. For each daily
electricity demand profile selected from \( f_i \) and \( \rho_i \), the
summed \( f \)-statistic and summed Pearson’s correlation
statistic (\( p \)) are recorded:

\[
f_{rec} = [a, b] \tag{6}
\]

\[
\rho_{rec} = [a, b] \tag{7}
\]

where \( a \) is the recorded summed statistic for the daily
electricity demand profile selected from \( f_i \), denoted as \( A \),
and \( b \) is the recorded summed statistic for the daily
electricity demand profile selected from \( \rho_i \), denoted as \( B \).

The selection of the daily electricity demand profile
which is most similar to every daily electricity demand
profile to form the basis structure, \( BS \), is conducted by the
following algorithm:

\[
f_{diff} = \left( f_{rec}(a) - f_{rec}(b) \right) / \left( f_{rec}(b) \right)
\]

\[
\rho_{diff} = \left( \rho_{rec}(a) - \rho_{rec}(b) \right) / \left( \rho_{rec}(b) \right)
\]

\[
\text{if } |f_{diff}| > |\rho_{diff}|
\]

\[
BS = A
\]

\[\text{else}\]

\[
BS = B
\]

end

D. Autoregressive Model

The choice of the autoregressive method to estimate
model parameters over other method such as NN relied on
the fact that use of the regression function is more
computationally efficient and less complex. To calculate the \( \text{diff}(•) \) electricity demand difference forecast function,
the autoregressive method is undertaken. The first step of
the autoregressive method is the use of the autocorrelation
function and partial autocorrelation function to identify the
structure of the model.

Once autocorrelation and partial autocorrelation
functions have been calculated, the results are tested for
statistical significance (equation 8) to determine what lags
are to form the basis of the coefficients in the
autoregressive model.

\[
\rho(\bar{X},\bar{Y}) = \frac{1}{n^3 - 1} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)
\]

\[
\text{std}(X)\text{std}(Y)
\]

\[
\pm \frac{z_{1-\alpha/2}}{n^{1/2}} \tag{8}
\]

where \( z \) is a two-tailed score from Student’s \( t \) –
distribution with a level of significance of \( \alpha \) and \( n \) is the
number of time steps in a daily electricity demand profile.
There is a typical assumption of a 95% confidence
interval, therefore \( \alpha \) is 0.05.

The data is normalised then the coefficients of the
autoregressive model are estimated by multivariate
regression as described by equation 9.

\[
\beta = \left( \bar{X}' \bar{X} \right)^{-1} \bar{X}' \bar{Y}
\]

where \( \beta \) is a vector of the estimated coefficients, \( \bar{X} \) is
the matrix of observations pertaining to the independent
variables and \( \bar{Y} \) is a vector of observations pertaining to the
dependent variable which \( \bar{X} \) is regressed against.

VII. RESULTS AND DISCUSSION

A. Basis Structure

The basis structure as a forecast model itself has a
coefficient of determination of 0.19 and a standard error
(root mean squared error) of 561.08 W.

B. Autocorrelation and Partial Autocorrelation

The threshold to determine statistical significance
correlation at the 95% confidence interval is 0.28. As
observed by the result of the partial autocorrelation
function (Figure 2) the correlations between the electricity
demand at time \( t \) and its subsequent lags cease to be
significant after two lags. This sets the electricity demand
difference forecast function with two electricity demand
lag coefficients, lag 1 and lag 2 from \( t \). The demand
difference forecast model demand is described by equation 10.

\[
diff_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2}
\]  

(10)

where \( \text{diff}_t \) is the electricity demand difference forecasted at time \( t \), \( x_{t-1} \) is the electricity demand difference between the basis structure and daily electricity demand profile at time \( t \) lagged by 1, \( x_{t-2} \) is the electricity demand difference between the basis structure and daily electricity demand profile at time \( t \) lagged by 2, \( \beta_0 \) is the value of the y-intercept, \( \beta_1 \) is the value of lag 1 coefficient and \( \beta_2 \) is the value of lag 2 coefficient.

**C. Analysis**

Table 1 below displays the results of the regression estimation of the coefficients in the electricity demand difference forecast function and statistical analysis. The calculated f-statistic on the regression was greater than the level of significance at the 95% confidence interval. This suggests that the calculated regression coefficients are not equivalent to 0. The coefficients were tested for statistical significance at the 95% confidence interval. Each coefficient was observed to be statistically significant indicating that the estimated coefficients are not equivalent to 0. From a comparison of the coefficients, the lag 1 coefficient is the dominating coefficient in the forecast mode.

![Partial Autocorrelation](image)

**Figure 2. Partial Autocorrelation**

**TABLE I. COEFFICIENT ANALYSIS**

<table>
<thead>
<tr>
<th>Test</th>
<th>Critical 95% C.I.</th>
<th>Statistic</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>f-test</td>
<td>2.76</td>
<td>1911</td>
<td>Yes</td>
</tr>
<tr>
<td>t-test</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Value</th>
<th>Statistic</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>0.07</td>
<td>450.76</td>
<td>Yes</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.83</td>
<td>2552.76</td>
<td>Yes</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.12</td>
<td>373.62</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2 below displays the accuracy statistics for the electricity demand difference forecast function, \( \text{diff}(\bullet) \), and the general forecast model, basis structure + \( \text{diff}(\bullet) \). Both forecast models have the same standard error which is equivalent to 31% of the average electricity demand observed and the same confidence interval at the 95% confidence level. The calculated Durbin-Watson Statistics value of 2 suggests that there is no autocorrelation of error in both models. The coefficient of determination of the general forecast model is greater than the basis structure alone by 0.46 and has a smaller standard error by 192.53 W. These statistics describe that the general forecast model is better apt in replicating the observed variance in weekday daily electricity demand.

**TABLE II. MODEL ANALYSIS**

<table>
<thead>
<tr>
<th>Model</th>
<th>Standard Error (W)</th>
<th>Confidence Interval (W)</th>
<th>( R^2 )</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_t )</td>
<td>1.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d_t(\bullet) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS + ( d_t(\bullet) )</td>
<td>368.55</td>
<td>± 722.36</td>
<td>0.56</td>
<td>2.00</td>
</tr>
</tbody>
</table>

D. Observed vs. Forecasted Electricity Demand

Figure 3 displays two curves; the observed daily electricity demand profile and the forecasted daily electricity demand profile based on the 30 minute ahead general forecast model. From the comparison of the two profiles, it can be seen that there are instances of the forecasted profile that deviate from the observed profile. The most noticeable deviations occur between minutes 1100 and 1300. The first discrepancy between forecasted profile from the observed profile results from a lack of input information in the model in order to forecast large spikes in future demand, as denoted between minutes 1140 to 1170. The second discrepancy between the forecasted profile and the observed profile is the appearance that the forecasted profile follows the observed profile shifted one time interval when large demand spikes occur. This is a result of the lag 1 coefficient being greater in magnitude than the other coefficients in the electricity demand difference forecast function.

![Observed vs. Forecasted Electricity Demand](image)

**Figure 3. Observed vs. Forecasted Electricity Demand**

The deviation between the observed and forecasted electricity demand is a result of the developed general forecast model not being specifically suited to this type of daily electricity demand profile i.e. a high peak during the peak demand period. As published in the literature, electricity demand is correlated with variables such as season, temperature, humidity, and load characteristics [3, 5]. There variables have not yet been considered in the development of the electricity demand forecast component.
A possible avenue of mitigating the observed deviation and developing a sound forecast component includes the following steps:

1. Implementation of a more robust technique of clustering daily electricity demand profiles that takes into account additional variables such as season, month, temperature, humidity, weather, days of the week, etc. rather than just the weekday-weekend dichotomy. This can be achieved through a multivariate clustering algorithm.

2. General forecast models will be developed for each cluster identified in step 1 according to the method outlined in this paper.

3. For the purpose of forecasting electricity demand, a general forecast model will be selected from a NN classifier based on the variables used to cluster the daily electricity demand profile.

4. To increase model accuracy, the ARMA or ARIMA time series models can be replace the autoregressive method for purposes of populating the demand difference forecasting function.

It is anticipated that similar electricity demand profiles will be clustered according to different permutations in the additional set of variables. For a hypothesised high demand day with large peaks, a suitable electricity demand forecast model would be called. In such a circumstance, the deviation between electricity demand and the basis structure would not be as great which should result in improved forecasts.

VIII. CONCLUSION AND FUTURE WORK

This paper set out to describe the method of a prototype forecasting component of the energy resource management control algorithm for STATCOMs with battery energy storage. It is desired for forecasting of electricity demand of each load throughout the LV distribution network. This mandates a large number of models to be estimated. The more computationally efficient and less complex method of developing forecast models, the less demand there is on hardware requirements. To meet these requirements a tiered general forecast model was selected. The forecast model is comprised of a basis structure selected from observed electricity demand data and an electricity demand difference forecasting component estimated by the autoregressive method.

The prototype forecasting component was demonstrated by the application of the method on one residential dwelling. The selected basis structure was found to account for observed variance with a coefficient of determination of 0.19. The addition of the electricity demand difference forecasting function increased the model's ability to replicate observed variance with a coefficient of determination of 0.65. Further analysis of the produced general forecast model found that there were discrepancies between the forecasted electricity demand profile and the observed electricity demand profile such that the forecast followed the observed electricity demand when large demand spikes occur. This is due to the developed general forecast model not being specifically suited to the different types of electricity demand profiles.

To overcome present forecasting model limitations, future work will entail the development of an expert system based on a NN classifier. The expert system will involve clustering electricity demand profiles against a number of variables. Forecast models will be developed for each identified cluster. The NN classifier will forecast what type of daily electricity demand profile future demand is likely to occur derived from the identified clusters. In turn, the classification will enable a general forecast model to be called. Further improvements in model accuracy may be attained through the ARMA or ARIMA time series models.

In conclusion, the work presented in this paper serves as an important first step in the development of the energy management control algorithm’s forecasting component. It is anticipated that future work as listed will overcome the limitations of the prototype method through the development of forecasting models suited to different types of daily electricity demand profiles.

ACKNOWLEDGMENT

I would to acknowledge ENERGEX for supplying the data used in the development of the prototype forecast method.

REFERENCES


