

# Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system



Christopher J. Bennett, Rodney A. Stewart\*, Jun Wei Lu

Griffith School of Engineering, Griffith University, Gold Coast Campus 4222, Australia

## ARTICLE INFO

### Article history:

Received 10 August 2013

Received in revised form

6 January 2014

Accepted 9 January 2014

Available online 7 February 2014

### Keywords:

Expert system

Energy

Electricity

Demand forecast

Neural network

Low voltage

## ABSTRACT

The advent of distributed renewable energy supply sources and storage systems has placed a greater degree of focus on the operations of the LV (low voltage) electricity distribution network. However, LV networks are characterised by having much higher variability in time series demand meaning that modelling techniques solely relying on iterative forecasts to produce a next day demand profile forecast are insufficient. To cater for the complexity of LV network demand, a novel hybrid expert system comprised of three modules, namely, correlation clustering, discrete classification neural network, and a post-processing procedure was developed. The system operates by classifying a set of key variables associated with a future day and refining a recalled historical demand profile as the forecast. The expert system exhibited high hindcast accuracy when trained with a residential LV transformer's demand data with  $R^2$  values ranging from 0.86 to 0.87 and MAPE (mean absolute percentage error) ranging from 11% to 12% across the three phases of the network. Under simulated real world conditions the  $R^2$  statistic reduced slightly to 0.81–0.84 and the MAPE increased to 12.5–13.5%. Future work will involve integrating the developed expert system for forecasting next day demand in an LV network into a comprehensive distributed energy resource management algorithm.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

As the region-wide electricity generation and supply system steps down to the LV (low voltage) distribution network, the number of customers serviced by a transformer decreases, which in-turn, correlates with an increase in variability of time series electricity demand. Over short time periods (intraday to intra-week), demand is observed to have a greater degree of randomness, increased frequency of 'shocks' and less continuity between daily demand profiles of sequential days. The increase in variability can be attributed to the greater relative weighting of the behaviours of individual customers and influences of local phenomena e.g. weather, special events, etc.

The increase in demand variability poses a resource management problem for the operation of microgrids and DER (distributed energy resources). As an example, a time based heuristic energy management control system for an energy storage system would not be able to adequately meet its objectives due the times at which the system should optimally charge and discharge would be

changing on a daily basis. Similar to the operation of the conventional electricity generation and supply system, to overcome this resource management problem, control systems will need to rely on demand forecasts. Demand forecasts will enable the derivation of information such as how much power is required, the scheduling of charging and discharging of energy storage systems, and whether or not remedial measures are required to be employed.

This current research focusses on the development of a forecasting component for an energy management control algorithm for the purposes of scheduling DER in residential LV distribution networks. For the energy management control algorithm to achieve the optimal scheduling of DER, it is necessary for the demand profile for the next and subsequent days to be forecast as well as its key features, such as the time(s) of day when peak demand occurs and associated values.

ARIMA (autoregressive integrated moving average) modelling techniques have been shown to provide adequate forecasts when applied to systems with greater customer aggregation or longer forecast time intervals [1–6]. However, conventional modelling techniques such as ARIMA are sensitive to LV network prevalent uncharacteristic daily profiles and random shocks which will increase these models' propensity to produce naïve predictions. Applying iterative forecasting alone, the residuals of the random shocks would bias subsequent forecasts. To overcome some of the

\* Corresponding author. Tel.: +61 7 5552 8778.

E-mail addresses: [christopher.bennett2@griffithuni.edu.au](mailto:christopher.bennett2@griffithuni.edu.au) (C.J. Bennett), [r.stewart@griffith.edu.au](mailto:r.stewart@griffith.edu.au) (R.A. Stewart), [j.lu@griffith.edu.au](mailto:j.lu@griffith.edu.au) (J.W. Lu).

deficiencies of traditional time series forecasting techniques, forecasting researchers have begun to explore the application of Artificial NN (Neural Networks) to forecast demand and demand profiles [7–10]. The main benefits of the use of NNs include their ability to generalize, identify non-linear relationships and applicability to a wide range of applications [7].

To achieve the research goal to forecast demand profiles for high variance LV residential distribution networks, an expert system based on pattern recognition which incorporates a clustering algorithm and NN was developed. This paper describes the development and validation process for the expert system when applied to three phases of an LV transformer supplying power to 128 residential customers located in Brisbane, Australia.

## 2. Research background

Griffith University, Elevare, Ergon Energy and Energex are working on a joint project to assess the feasibility of the installation of STATCOM (static synchronous compensators) with BESS (battery energy storage systems) in the LV distribution network. Funding for this project has been provided by the Queensland State Government 2012–2014 Research Partnership Grant. STATCOMs are four quadrant synchronous inverters with the ability to correct frequency distortions and dampen harmonics. The combination of STATCOMs and BESS will enable the reduction of peak demand on network infrastructure and the active maintenance of power quality. The installation of this technology has the potential to reduce network expenditures through the replacement or deferral of other expenditures such as replacing transformers and/or upgrading lines.

The assessment of the feasibility involves the design and quantification of the effectiveness of STATCOMs with BESS. A number of subprojects were initiated to achieve the project's goals including determining the technical parameters, developing a STATCOM with BESS energy management control algorithm and performing economic analysis. The research reported in this paper denotes the completion of the demand forecasting component of the energy management control algorithm. Information generated from the determination of technical parameters, simulation of the STATCOM with BESS in the LV distribution network and physical trialling will be used as input variables in the economic analysis.

## 3. Literature review

### 3.1. Short-term electricity demand modelling

The most notable publications apply conventional modelling techniques including ARIMA, multivariate regression and machine learning techniques (e.g. support vector machines, fuzzy inference systems, NN, etc.). ARIMA(p,d,q) is the general model of the Box Jenkins set of time series modelling techniques. The 'p' represents that number of lagged parameters (autoregressive parameters); the 'd' represents the number of discrete differences; and the 'q' represents the number of lagged forecast error parameters in the model to account for a moving average in the time series. Regression models in the electricity demand space involve the addition of deterministic parameters to the use of the lagged forecast parameters. Additional parameters may include weather, economic, behavioural and time dependent variables. Many regression models can be considered ARIMAX (Auto Regressive Integrated Moving Average with Exogenous Input) models due to the combination of the ARIMA model with exogenous variables. NNs mimic how biological neural networks model systems. NNs are composed of two or more layers of artificial neurons with synapse (weights) linking each neuron of the previous layer to the next. Signals (inputs) are multiplied by weights connected to the neuron, summated, inputted into

the neuron's activation function and the output is sent to the neurons of the next layer. A training algorithm adjusts the weights throughout the network in order to model the desired system.

Engle et al. [1], Taylor [4], Mirasgedis et al. [5] and Taylor [6] developed network demand time series models based on the ARIMA or regression modelling techniques. Taylor [4] and Taylor [6] developed ARIMA models using the exponential smoothing, double seasonal exponential smoothing and triple seasonal algorithms. Taylor [4] used 30 min demand data from England and Wales. Taylor [6] used demand data from Britain and France. The research showed that the developed models were accurate and that accuracy increases as more seasonalities are included in the models. Engle et al. [1] and Mirasgedis et al. [5] developed time series models with autoregressive parameters and additional variables including heating and cooling days, RH (relative humidity) and day of the week dummy variables. It was noted that models performed well and models with weather variables performed better than models without.

Kassaei et al. [11], Darbellay and Slama [2], Abraham and Nath [12], Ringwood et al. [3] and Cavallaro [13] developed short-term electricity demand forecast models using NN. Darbellay and Slama [2] used Czech Republic demand data and Ringwood et al. [3] used Ireland's Electricity Supply Board's data to construct univariate NN models. The autocorrelation function was used to identify cyclical components in the demand time series and to structure the models accordingly. The models achieved a high level of accuracy and performed better than univariate ARIMA models. Cavallaro [13] constructed a multivariate NN with variables such as day of the week and average temperature and noted a high accuracy. Kassaei et al. [11] and Abraham and Nath [12] combined NN with fuzzy logic. Kassaei [11] used a univariate NN to model normal loads and fuzzy logic model to model weather dependent loads. The forecast is generated by the output of the NN and fuzzy logic model. It was found that the NN and fuzzy logic model performed better than the singular NN model. Abraham and Nath [12] applied an ARIMA, EFuNN (evolving fuzzy NN) and a NN to Victoria's demand data. The EFuNN approach differs from a conventional NN since the neurons in the network are performing functions such as 'fuzzification' of inputs, rule based transformations and defuzzification. The weights in the network are altered by a training algorithm. The EFuNN performed the best out of the set of developed models.

The above mentioned models were all short-term electricity demand models applied to networks with a large number of consumers. All achieved high levels of accuracy and there were not any clear distinctions regarding which modelling method yields the best results. The work conducted by Taylor [6] provided evidence that the greater number of seasonality variables included in the model accounted for increases in model accuracy. Engle et al. [1], Darbellay and Slama [2] and Mirasgedis et al. [5] indicated that the inclusion of deterministic variables such as weather variables improves model accuracy. An inference may be drawn from these studies that the inclusion of both ARIMA variables and deterministic variables would derive higher model accuracy. However, this inference becomes less applicable as the forecast period shortens due to the relationship between demand and deterministic variables not being apparent [4]. Forecast windows such as a day ahead or greater are more responsive to deterministic variables.

### 3.2. Demand profile forecasting

For the case of short-term demand forecasting models (i.e. 30 min ahead, an hour head, etc.) iterative forecasting techniques are essential. Iterative forecasting is where the forecast at time  $t$  is used as an input variable in the model to forecast at time  $t + 1$ . This process repeats itself until the desired number of forecasts has been made. A shortcoming of this technique to forecast demand profiles is that the

forecast errors in the initial forecast and each iterative forecast are compounded. Alternate modelling approaches have been developed including multivariate forecast NNs and ensembles of models and pattern recognition [7]. Multivariate forecast NN forecast the next day's demand profile through the use of multiple output neurons for each time interval of the next day's demand profile. The ensemble of models approach entails that there are independent models for each time interval of the next day's demand profile.

Beccali et al. [8], Hippert et al. [9] and Sousa et al. [10] used a multivariate forecast NN methodology in the development of short-term demand profile forecast models. Beccali et al. [8] used a SOM (self-organising map) algorithm to cluster similar demand profiles in order to provide demand profile indices and a NN with 24 output neurons for each hour of the day. Additional NN inputs to the indices included historical load data, historical weather data and day of the week dummy variables. Hippert et al. [9] developed a multivariate forecast NN and compared it against alternative techniques such as naïve forecasts, ensembles of smoothing filters and ensembles of regression models. Hippert et al. [9] found that the NN model performed better than the alternatives. Sousa et al. [10] adapted a demand profiling technique to a NN with 24 output neurons.

The ANNSTLF (NN short-term load forecast) system described by Khotanzad et al. [14] utilizes both an ensemble of models and NNs capable of producing multiple forecasts. The ANNSTLF system is comprised of four NNs. Each NN forecasts distinct multiple hours and are sequentially combined to produce the next day's demand profile. Input variables of the NNs include historical and forecasted temperature, historical and forecasted RH and historical demand data. The temperature is forecasted using an ensemble of NN models with multiple outputs and are combined using an adaptive scaling algorithm. RH is forecasted by the use of a moving average algorithm. The system was applied to data from ten utilities and displayed a high level of accuracy. Fan and Chen [15] developed a more advanced system which relied on a SOM to cluster demand profiles according to exogenous variables and multiple sets of 24 support vector regression models. When input variables are inputted into the trained SOM, it calls a specific set of 24 support vector regression models to forecast the demand profile.

Pattern recognition approaches involve an algorithm analysing a set of input variables and classifying according to a set of known or trained relationships. The SOMs employed by Beccali et al. [8] and Fan and Chen [15] are examples on the use of pattern recognition in the process of forecasting demand profiles. Beccali et al. [8] used the output of the SOM as an input variable for the NN; whereas Fan and Chen [15] used the output of the SOM to call a specific set of support vector regression models. Espinoza et al. [16], Konjic et al. [17] and Sousa et al. [10] clustered demand profiles of customer demographics and applied the known demand profiles to other customers in order to produce short-term demand forecasts. As noted by Hippert [7], machine learning techniques such as NNs can be used specifically for pattern recognition. Dai and Wang [18] furthered the use of NNs and pattern recognition to forecast the demand profiles of future days through load classification based on their associated input variables. The load sets are identified through clustering demand profiles against input variables. Each load set has a characteristic demand profile which in turn is used as the forecast.

### 3.3. Formulation of the expert system

An expert system is a program which has decision making capabilities based on reasoned knowledge. Many of the NN based demand forecast models discussed fulfil some properties of an expert system due to the way in which NNs model systems (i.e. self-

learning features). Systems containing pattern recognition faculties such as Beccali et al. [8], Fan and Chen [15] and Dai and Wang [18] best fulfil the criteria of being expert systems due to decision making via a classification being central.

In order for the demand profile of residential LV transformers to be forecast, the modelling method employed must not be sensitive to high variance, low continuity of demand profile patterns of sequential days and presence of shocks. This biases the selection of a modelling method away from the iterative, ensemble modelling and some derivations of the multivariate NN forecast methods due to their reliance on historical demand data. Pattern recognition approaches are more versatile due to their ability to classify demand profiles based on their association with exogenous or deterministic variables.

To advance the approach, demand profile properties such as total load or TEU (total energy use), peak demand and morning peak can be used as input variables to better the classification of future demand profiles and post-processing. These properties can be forecast for future days using conventional techniques such as ARIMAX variables due to relatively lower variance in the TEU, peak demand and morning peak time series and greater response to exogenous variables. Their role in post-processing is to modify the characteristic demand profile such that the most accurate forecasts are produced.

## 4. Research objective

The core research objective was to develop an expert system to forecast the next day's demand profile for the LV residential distribution network.

## 5. Data

### 5.1. Source

Energex (i.e. the power distribution company supplying South East Queensland, Australia) provided the data for the three phase LV transformer and the 128 residential customers the transformer supplied (i.e. case study area). The transformer is located in an inner northern suburb of Brisbane, Queensland, Australia. The metering of the transformer involved recording the voltage, current and power factor for each phase at 10 min intervals. The phases of the transformer are unbalanced. The provided data set contains data for the period covering the middle of January 2012 to the middle of February 2013. Weather statistics used in this research, such as temperature and RH, were collected by the Brisbane City weather station and made available online from the Australian Bureau of Meteorology. The data from 2012 was used as input variables of the coefficient estimation of ARIMAX (Auto Regressive Integrated Moving Average with Exogeneous Input) models and input variables for the training of the NN.

Brisbane has a subtropical climate and experiences mild winter and hot humid summers. Data sourced from the Australian Bureau of Meteorology [19] displays that the hottest month of the year is January which has average maximum and minimum temperatures of 29.0° and 21.2°. The coldest month of the year is July which has average maximum and minimum temperatures of 20.8° and 9.0°. Brisbane has a mean annual precipitation of 1028.2 mm with the greater amount of precipitation occurring over the months of November to March.

### 5.2. Overview

Fig. 1 displays 168 h (7 day) subsections of phase 1's smoothed demand time series denoting summer (a), autumn (b), winter (c)

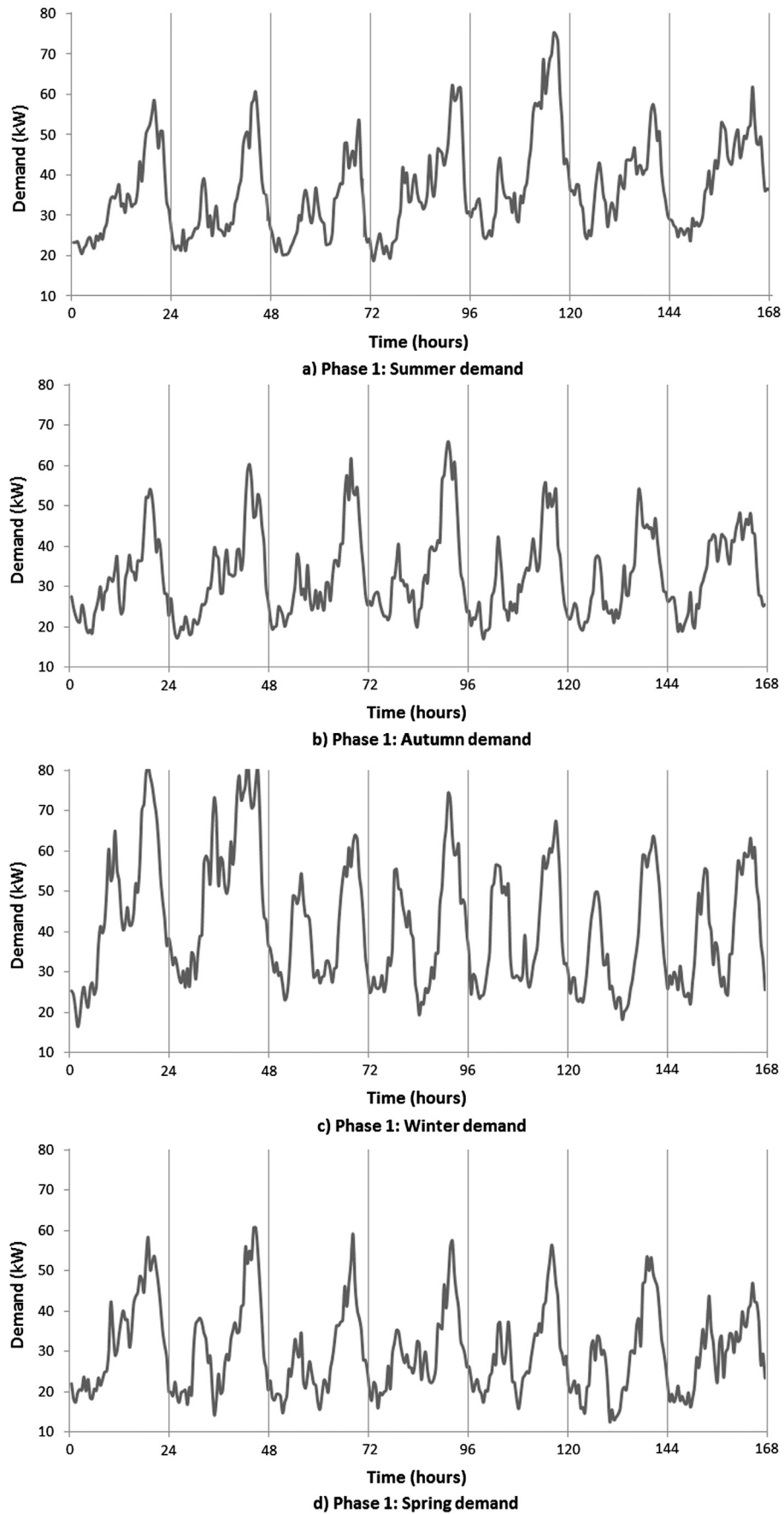


Fig. 1. Seasonal demand profiles.

and spring (d) demand profiles. What is observed is that the shape of the demand profile changes throughout the year according to changes in customer behaviour in response to changes in temperature. Three types of demand profiles can be identified. The first which typically occurs during summer is characterised by a small or

no peak during the morning and a large peak in the evenings. The second which occurs during autumn and spring periods has a small peak in the morning and a larger peak in the evenings. The third which occurs during winter has a large peak in the morning and a large peak in the evening.

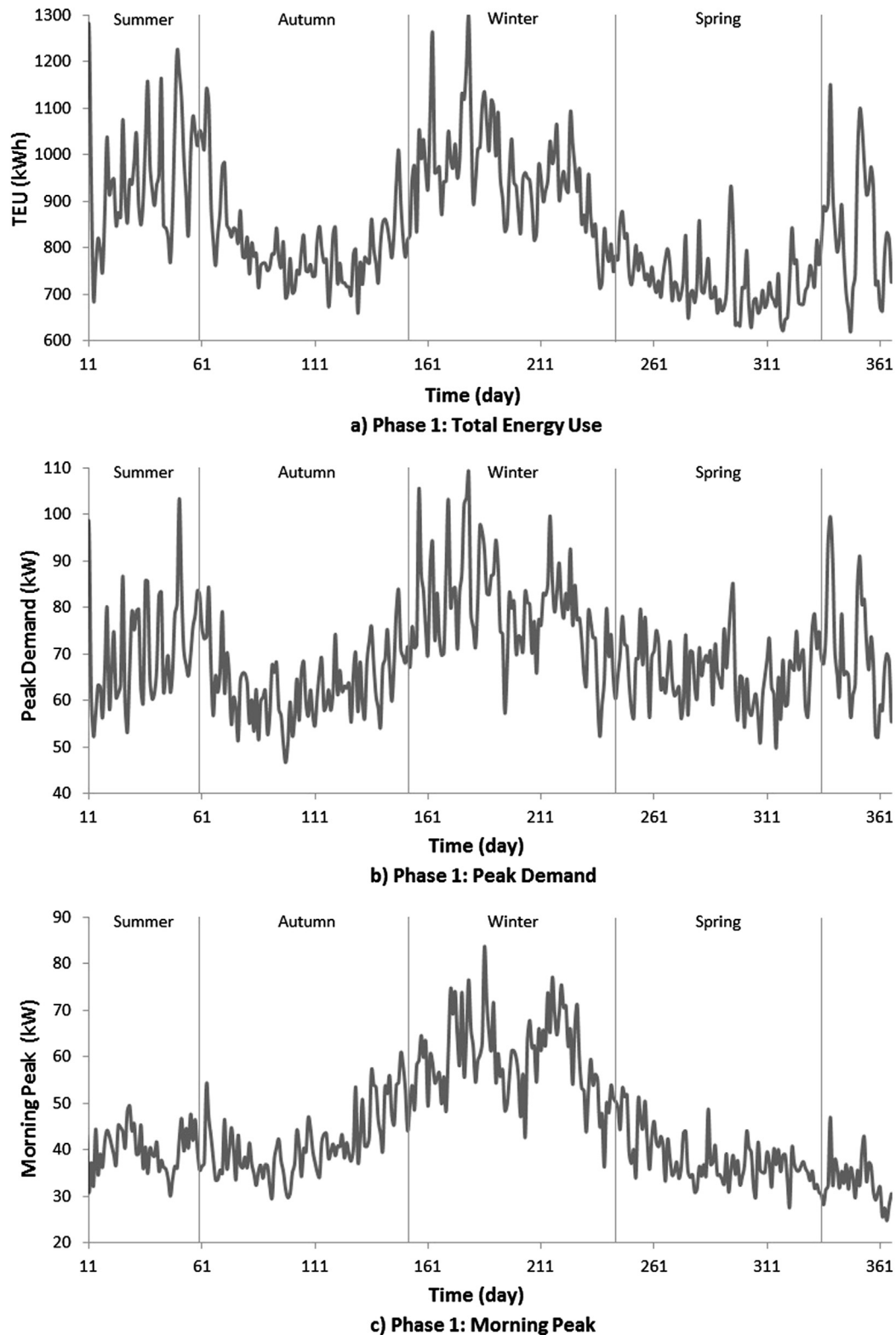


Fig. 2. Seasonal demand profile properties.

Fig. 2 displays phase 1's demand profile properties such as daily TEU (a), daily peak demand (b) and daily morning peak (c) for the 2012 component of the time series. The trend in the TEU time series is such that the greatest amount of energy is consumed during the summer and winter periods of the year. Winter has a consistently greater level of consumption in comparison to summer which is more volatile. The yearly peak energy consumed of 1.3 MWh occurred in winter, on day 179. The peak demand time series follows a similar pattern with the greatest demands occurring in summer and winter. The yearly peak demand of 108 kW occurred in winter, on day 179. The morning peak time series exhibits a different pattern. The morning peak is constant for the first third and last third of the year. As daily temperature starts to decrease in the middle third of the year, the daily demand profile begins to exhibit the two large peak demand profile patterns. These time series reaffirm the observation that the shape of the demand profile changes due to customers' responses to changes in temperature.

## 6. Method

### 6.1. Overview

#### 6.1.1. Formulation

It was noted that the transformer's demand profiles for consecutive days are inconsistent; differing in the profile's shape and magnitude. What was observed throughout the year is that a finite group of patterns in the demand profiles repeat themselves according to repetitions in the set of external variables that affect demand such as temperature and day of the week. Periods where the demand profiles are more consistent are generally due to the set of exogenous variables on each day being similar. Due to the coincidence of the finite set of patterns and the respected sets of exogenous variables, a pattern recognition solution is feasible for forecasting future demand profiles. The general operation of the pattern recognition solution is such that a future day will have a set of exogenous variables. Based on this set of variables an algorithm will select a pattern from the group of patterns which will best represent the demand profile which will occur on that day. The development and the operation of the solution are distinct processes. To develop the solution the following steps are required:

1. Identify repeating patterns within the time series by clustering demand profiles by how well they correlate with one another. The cluster that a demand profile is a member of is that demand profile's classification. The mean of each cluster and the CDP

(characteristic demand profile) form the underlying repeating patterns.

2. The exogenous variables associated with each demand profile are to be associated with the group that the demand profiles are classified under.
3. A pattern recognition algorithm is to be trained to associate sets of exogenous variables to respected groups of demand profiles. The trained pattern recognition algorithm takes a set of exogenous variables as input variables and selects a cluster which the set most likely occurs under.

Following the listed steps, the expert system based on pattern recognition was developed using a correlation clustering algorithm and a feed forward back propagation NN.

#### 6.1.2. Expert system development and operation

Fig. 3 displays the steps involved in the development and operation of the expert system. The development process begins with data pre-processing. The data is then passed into a correlation clustering algorithm that produces a set of identified clusters and variables associated with each cluster. This information is passed into a multivariate forecast NN. The NN is structured with an output vector with  $k$  number of elements matching the number of clusters. The first element represents cluster 1 and the  $k$ th element represents cluster  $k$ . The NN is trained under the back propagation algorithm. From the identification of clusters, the demand profiles for each cluster are averaged to produce the set of CDPs. There're  $k$  number of characteristic demand profiles. Characteristic demand profile 1 corresponds to cluster 1 and characteristic demand profile  $k$  corresponds to cluster  $k$ .

The operation of the expert system to forecast the demand profile for a future day starts with the acquisition of the weather forecast for that day. This data is then supplied to three ARIMAX demand profile property forecast models. The weather forecast data and demand profile property forecast are supplied to the NN as input variables. These are combined with additional "known" variables such as day, week, month, public holiday, etc. The NN produces the output vector with  $n$  number of elements. Each element will have a value ranging from 0 to 1. This is analogous to estimations of the likelihood of a given set of input variables to be classified under a specific cluster. The set of input variables are classified under the corresponding cluster with the element which contains the highest value or value closest to 1. If element  $x$  has the highest value then the set of input variables will be classified under cluster  $x$ . The associated characteristic demand profile with cluster

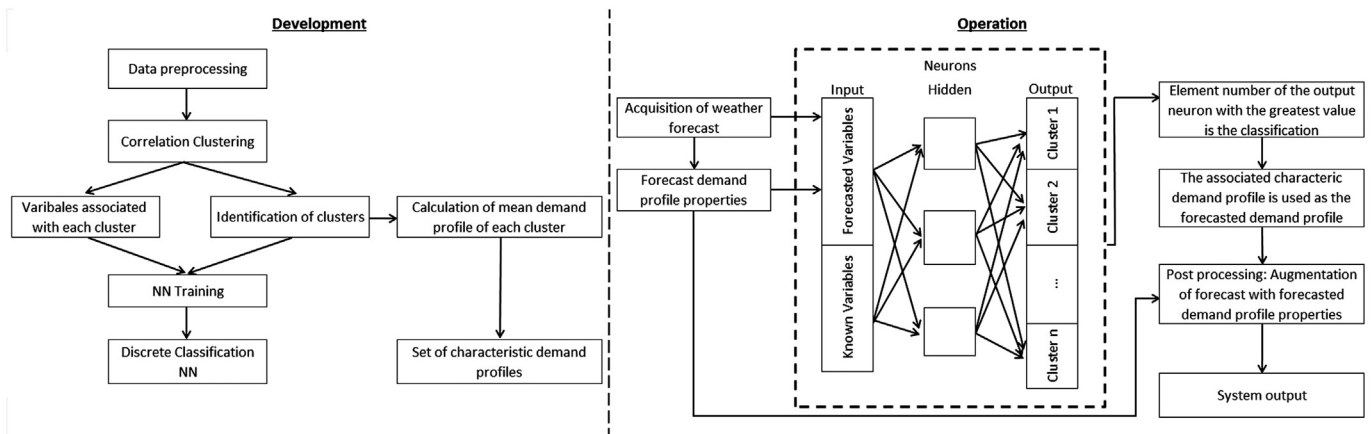


Fig. 3. Expert system construction and operation.



classification is then used as the demand profile forecast. The demand profile forecast is then improved by augmenting it according to the forecasted demand profile properties.

## 6.2. Pre-processing

Pre-processing of the data involved the following steps:

1. Removal of demand profiles from the set which contained 10 or more data points with missing data points.
2. Conversion of the 10 min average power recordings to 30 min average power recordings.

The removal of demand profiles with 10 or more missing points reduced the number of demand profiles from 366 to 349. The conversion from 10 min average power recordings to 30 min averages dampens the effect that random shocks have on the time series. This reduces the number of elements in the demand profile from 144 to 48. In turn, the demand profile forecast will be comprised of the average power value for 48 half hour intervals. Half hour intervals are sufficient to achieve the ultimate goal of this research.

## 6.3. Correlation clustering

The correlation clustering algorithm operates by clustering a set of vectors based on how well they correlate with different subsets of vectors. Correlation is calculated by the use of Pearson's correlation equation (eq. (1)). The output of the equation will range from 0 to 1. A value of 1 denotes a perfect correlation and a value of 0 denotes no correlation.

$$\rho(X, Y) = \frac{1}{n-1} \frac{\sum ((x - u_x)(y - u_y))}{\sigma_x \sigma_y} \quad (1)$$

where  $\rho$  is the correlation statistic,  $X$  and  $Y$  are vectors,  $n$  is the number of elements in each vector,  $x$  is an element within vector  $X$ ,  $y$  is an element in vector  $Y$ ,  $u_x$  is the mean of  $X$ ,  $u_y$  is the mean of  $Y$ ,  $\sigma_x$  is the standard deviation of  $X$  and  $\sigma_y$  is the standard deviation of  $Y$ .

Given  $k$  (user defined) set of clusters  $C = \{C_1, C_2, \dots, C_k\}$ , there will be  $k$  number of cluster means  $CDP = \{CDP_1, CDP_2, \dots, CDP_k\}$ . The calculation of the cluster means represent the underlying trends with the clusters that DP (demand profiles) can be correlated with. There're  $n$  number of demand profiles. For each iteration of the clustering process, where  $t$  is the iteration number, the algorithm calculates the cluster means (eq. (2)) and assigns demand profiles to each cluster (eqs. (3) and (4)). The algorithm continues until user defined number of epochs has been reached.

$$CDP_i^{(t)} = \text{mean}(C_i^{(t)}) \quad (2)$$

calculated for all clusters  $CDP_i$  for  $i = 1, 2, \dots, k$ .

$$P_j = \{\rho(DP_j, CDP_1), \rho(DP_j, CDP_2), \dots, \rho(DP_j, CDP_k)\} \quad (3)$$

where  $P$  is a set of correlation statistics between demand profile  $j$  and the set of cluster means. This is calculated for all demand profiles  $DP_j$  for  $j = 1, 2, \dots, n$ .

$$C_i^{(t+1)} = \{DP_j : P_{ij} = \max(P_j) \text{ for } 1 \leq j \leq n\} \quad (4)$$

which states if correlation statistic  $P_{ij}$  equals the maximum of the set then demand profile  $j$  will be classified under cluster  $i$ . This is calculated for all clusters  $CDP_i$  for  $i = 1, 2, \dots, k$ .

After the correlation clustering algorithm has completed, the clusters are subdivided by magnitude. In order to derive the

demand profiles within each cluster they are integrated, z-scored and separated by arbitrary intervals of deviations from the mean. The final number of clusters is  $k$  multiplied by the number of intervals of deviations from the mean.

## 6.4. Artificial neural network

### 6.4.1. Algorithm

The NN is constructed by layers of artificial neurons interconnected by synapse (weights) from one layer to the next. The NN chosen for discrete classification is a sigmoid activation back-propagation network. Each neuron receives input signals from a set of synapse and performs a function described by equations (5) and (6). The NN is trained, with adjustments of the weights throughout the network to achieve the optimal forecast, according to equations (7)–(10). Equation (7) is the training algorithm for weights connected to the output layer and equation (8) is the training algorithm for weights of preceding layers. For the network to operate the input and output data was normalized. To prevent over fitting the algorithm randomly selects and separates 20% of the data in order to validate the model. The algorithm continues until user defined number of epochs has been reached.

$$v_j = \sum_{h=1}^m w_{jh} x_h \quad (5)$$

where  $v_j$  is the summation of the weights multiplied by the outputs of each neuron (or inputs of the NN) respectively for neuron  $j$ . There're  $m$  number of neurons or inputs in the previous layer.

$$Y_j = \sigma(v_j) = \frac{1}{1 + \exp(-v_j \alpha)} \quad (6)$$

where  $Y_j$  is the output of neuron  $j$ ,  $\sigma(v_j)$  is the sigmoid function and  $\alpha$  is a constant which affects the gradient of the sigmoid function.

$$\delta_j = \sigma'(v_j) (\hat{Y}_j - Y_j) \quad (7)$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \omega \delta_j Y_i \quad (8)$$

where  $w_{ij}$  is the weight connecting neuron  $i$  to output neuron  $j$ ,  $\omega$  is the training rate,  $v_j$  is the result of the summation function for neuron  $j$ ,  $\hat{Y}_j$  is the forecasted value,  $Y_j$  is the observed value,  $\delta_j$  is the local gradient at neuron  $j$ ,  $Y_i$  is the output of neuron  $i$  of the previous layer and  $t$  is the training epoch.

$$\delta_i = \sigma'(v_i) \sum_{j=1}^k \delta_j w_{ij} \quad (9)$$

$$w_{hi}^{(t+1)} = w_{hi}^{(t)} + \omega \delta_i Y_h \quad (10)$$

where  $w_{hi}$  is the weight connecting neuron  $h$  to output neuron  $i$ ,  $v_i$  is the result of the summation function of neuron  $i$ ,  $\delta_i$  is the local gradient at neuron  $i$  and  $Y_h$  is the output of neuron  $h$  of the previous layer.

### 6.4.2. Variable selection

The input variables of the NN are selected by a combination of prior statistical analysis and directed trial and error. Prior statistical analysis involves observing how demand changes for different external variables. It has been observed that variables such as temperature and humidity have effects on demand [1,5]. Directed

trial and error involves the addition and removal of variables according to whether or not they increase the accuracy of the model. When a variable is added or removed the network is retrained multiple times to establish accuracy baselines for the training and validation sets. Each time the network is retrained, a new validation set is selected randomly from the data set. The variable is added or removed depending on whether or not the baseline accuracies improve.

### 6.5. Forecasting

Each element  $Y_i$  of the output vector of the NN  $Y$  corresponds to a cluster  $C_i$  within the set of clusters  $C$ . The future day's input variables are classified under the cluster  $i$  where  $Y_i$  equals the element with the maximum value within the vector  $Y$ . The forecasted demand profile  $DP_f$  will equal the cluster mean  $CDP_i$ .

### 6.6. ARIMAX models

To improve forecast accuracy, forecasted demand profile properties such as NDPD (next day peak demand), NDMP (next day morning peak) and NDTEU (next day TEU) can be used to augment the forecasted demand profile. The three forecast models, NDPD, NDMP and NDTEU, are ARIMAX models that were developed for each phase of the network and validated. The ARIMAX models are populated with exogenous variables (e.g. temperature, RH, day of the week, etc.), autoregressive terms and the double exponential smoothing algorithm. The peak demand, morning peak and TEU time series contain data for 403 days. The first 190 days of the time series were used for model training (coefficient estimation via regression) and the remaining 201 days were used for model validation.

The morning peak and the peak demand forecasts can be used to adjust the amplitudes of the morning peak and peak demand of the demand profile forecast. The area under the forecasted demand profile curve can be adjusted to match the TEU forecast. The morning peak and peak demand adjustments of the demand profile forecast will alter its integral. To mitigate the likelihood of greater deviations from observations for the NDPD and NDMP adjustments, the morning peak and peak day demand values will be adjusted first then the TEU forecast.

### 6.7. Post processing

#### 6.7.1. Peak adjustment

The adjustment of the morning peak and peak demand follow the same algorithms. The first is the peak decrease algorithm set denoted by equations (11)–(13). The second is the peak increase algorithm set denoted by equations (14)–(16). Both algorithms initialize by halving the demand profile forecast into two vectors; one containing the morning peak and one containing the evening peak. The following discussion outlines the steps for calculating both vectors.

The peak decrease algorithm finds elements which have a higher value than the peak forecast described by equation (11):

$$EL = \{j : DP_{j,f} = DP_{j,f} > P_f \text{ for } 1 \leq j \leq n\} \quad (11)$$

where EL (element location) is a set of element locations where the element has a greater value than peak forecast  $P_f$ , where  $j$  is an element number and  $n$  is the length of the vector. The third step, equation (12), in the algorithm set is to find the difference between the peak of the demand profile vector peak and the peak forecast:

$$\Delta P = \max(DP_{j,f}) - P_f \quad (12)$$

where  $\Delta P$  is the difference between the element with the maximum value in  $DP_j$  and the peak forecast  $P_f$ . The forth step of the peak decrease algorithm involves the adjustment of the values in the demand profile which are greater than the peak forecast:

$$DP_{EL,f} = DP_{EL,f} - \Delta P \quad (13)$$

which describes that the elements which have a value greater than the peak forecast  $P_f$  that were adjusted downwards while maintaining the prior shape of the demand profile curve.

The peak increase algorithm calculates the difference between the peaks of the demand profile forecast and the peak forecasts using equation (14):

$$\Delta P = P_f - \max(DP_{j,f}) \quad (14)$$

where the variables are described above. The algorithm proceeds to create a vector with values ranging from 0 to 1 and 1 to 0 and multiplies it with  $\Delta P$  as per equation (15):

$$\Delta DP = [0, 0.1, \dots, 1, 1, \dots, 0.1, 0] \times \Delta P \quad (15)$$

where  $\Delta DP$  is the peak adjustment vector which has a scaled response to maintain the shape of the peak period of the original demand profile forecast. The peak adjustment vector is then added to the pp (peak period) according to equation (16):

$$DP_{pp,f} = DP_{pp,f} + \Delta DP \quad (16)$$

which describes that the peak periods of the demand profile are adjusted according to the adjustment vectors. The adjustment vectors highest value overlaps with the peak demand element.

#### 6.7.2. Total energy demand adjustment

The adjustment of the demand profile forecast according to the forecasted TEU follows a series of steps. These steps include: (1) calculate the integral of the forecasted demand profile; (2) calculate the difference between the integral and the TEU forecast; and (3) convert the difference value  $\Delta TEU$  to an average power drawn over a 24 h time period and add it to the demand profile forecast. These three calculation steps are described by equations (17)–(19), respectively, as follows:

$$\Delta TEU = TEU_f - \int DP_f \quad (17)$$

where  $\Delta TEU$  is the difference between the forecast  $TEU_f$  obtained from the NDTEU model and the integral of the demand profile forecast  $DP_f$ .

$$\overline{\Delta Power} = \frac{\Delta TEU}{r} \quad (18)$$

where  $\overline{\Delta Power}$  is the average difference in power drawn over a 24 h time period and  $r$  is the number of discrete time intervals in the demand profile.

$$DP_p = DP_f + \overline{\Delta Power} \quad (19)$$

where  $DP_p$  is the post-processed forecast.



## 7. Results and discussion

### 7.1. ARIMAX models

The variables contained within the NDPD, NDMP and NDTEU models are outlined in Table 1. The variables which populate the models were selected on the basis that their inclusion improved accuracy statistics for both the training and validation sets. Demand  $t-1$ , Demand  $t-2$ , Temp., Temp.2, RH and DES forecast variables are included in all models. The double exponential smoothing algorithm accounts for the changing local mean throughout the times series and is analogous to a ARIMA(0,2,2) model. The two autoregressive terms mitigate autocorrelation in the error terms and make the ARIMA components ARIMA(2,2,2) within the ARIMAX models. Temp. and Temp.2 variables account for the parabolic response to temperature. NDPD and NDTEU models contain day of the week dummy variables and RH-temperature interaction terms. The morning peak time series was observed as not being responsive to different days of the week. NDPD has one additional variable which is the intercept.

Table 2 displays the models' coefficient of determination ( $R^2$ ) and MAPE (mean absolute percentage error) accuracy statistics for both the training and validation sets. The NDTEU models displayed the highest level of accuracy for the both the training and validation sets. The NDPD and NDMP models were less accurate than the NDTEU models. A possible explanation for this deviation in model accuracy could be that the peak demand and morning peak time series were more variable and had greater instances of random shocks than the TEU time series. The post-processing algorithm adjusts the demand profile forecast according to the TEU forecast after the morning peak and peak demand adjustments have been made. This enables a more accurate forecast to mitigate potential NDPD and NDMP forecast errors.

Fig. 4 presents Phase 3's ARIMAX models' hindcasts and forecasts. For the majority of the training and validation time series, the hindcasts and forecasts follow the pattern of the observed data. In line with the accuracy statistics displayed in Table 2, the NDTEU hindcasts and forecasts exhibit a better fit to observed data than the NDPD and NDMP models. The NDPD and NDMP models display instances where the hindcasts and forecasts significantly deviate from observations. Divergences on days 29, 58, 59 and 155 were attributed to spikes in demand on the day after a period of heavy rainfall. The Queen's Diamond Jubilee public holiday, day 163, had an abnormal spike in demand. The significant deviation on day 395 occurred the day after a period of heavy rainfall and minor flooding. The NDTEU models were observed to be less sensitive to the exogenous shocks.

**Table 1**  
ARIMAX model variables.

NDPD	NDMP	NDTEU
Intercept	Intercept	–
Demand $t-1$	Demand $t-1$	Demand $t-1$
Demand $t-2$	Demand $t-2$	Demand $t-2$
Temp.	Temp.	Temp.
Temp.2	Temp.2	Temp.2
RH	RH	RH
RH*Temp.	–	RH*Temp.
RH*Temp.2	–	RH*Temp.2
Sunday	–	Sunday
Saturday	–	Saturday
Monday	–	Monday
Tuesday	–	Tuesday
Wednesday	–	Wednesday
Thursday	–	Thursday
Friday	–	Friday
DES Forecast	DES Forecast	DES Forecast

**Table 2**  
Next day forecast models' accuracy statistics.

Model	$R^2$			MAPE		
Training						
Phase	1	2	3	1	2	3
NDPD	0.72	0.70	0.72	7.05	7.12	7.81
NDMP	0.80	0.60	0.66	8.12	9.05	8.65
NDTEU	0.84	0.77	0.87	4.51	4.21	4.44
Validation						
Phase	1	2	3	1	2	3
NDPD	0.58	0.56	0.65	7.90	7.48	8.13
NDMP	0.75	0.44	0.66	10.12	12.88	13.21
NDTEU	0.74	0.80	0.78	6.97	6.30	7.33

### 7.2. Correlation clustering

Table 3 displays the number of demand profiles classified under each cluster. The correlation clustering algorithm was run until the most optimal clustering solution was found by minimising the number of low correlations within each cluster. This was achieved by an initial selection of 3 cluster nodes and further subdividing the clusters according to where z-scores of demand profiles are located (i.e.  $z < -1$ ,  $-1 < z < 1$  or  $z > 1$ ). This resulted in a total of 9 clusters per phase. The number of demand profiles per cluster ranged from 21 to 24.

### 7.3. Neural network

Table 4 contains the NN classification accuracy statistics. Each phase of the network has a separate NN with 9 output neurons corresponding to each cluster. 'Accuracy' is the percentage of correctly made classifications. 'Miss' represents the number of invalid or incorrect classifications. The higher the accuracy percentage and lower the miss percentage entails the better the NN is at correctly classifying a set of input variables. The accuracy percentage ranges from 0% to 100%. The miss percentage ranges from 2.15% to 11.11%. The overall accuracy and miss percentages for each NN range from 70.66% to 77.98% and 4.69%–6.04% respectively. A possible explanation of the low classification accuracy of some clusters is that the cluster is too similar to another in terms of magnitude, shape and associated input variables. Post-hoc analysis of clusters having very similar characteristics resulted in some cluster mergers.

### 7.4. Hindcast accuracy statistics

Table 5 highlights the accuracy statistics of the expert system applied when using the training data set. The accuracy statistics include RMSE (root mean square error or standard error),  $R^2$ , correlation and MAPE. The RMSE displays the 1  $\sigma$  confidence interval related to each phase. Each phase's expert system had high levels of accuracy. Phase 1's expert system had the least accuracy with an  $R^2$  of 0.86, correlation of 0.93 and MAPE of 12%. Phase 2's expert system had the highest accuracy with an  $R^2$  of 0.88, correlation of 0.94 and MAPE of 11%. These reported high levels of accuracy indicate that the expert system's sensitivity to miss-classifications is low.

Fig. 5 displays the  $R^2$  statistic distribution for each phase's expert system when the accuracy of each day's hindcast is calculated independently. It is to be noted that after pre-processing of the data, the number of days used to train the systems were reduced to 349. Phase 2's expert system had the greatest number of high accuracy forecasts with 143 days having  $R^2$  statistics greater than 0.9. Phase 1's and Phase 3's expert systems had a similar number of high accuracy hindcasts with 97 and 91 days with  $R^2$  statistics being

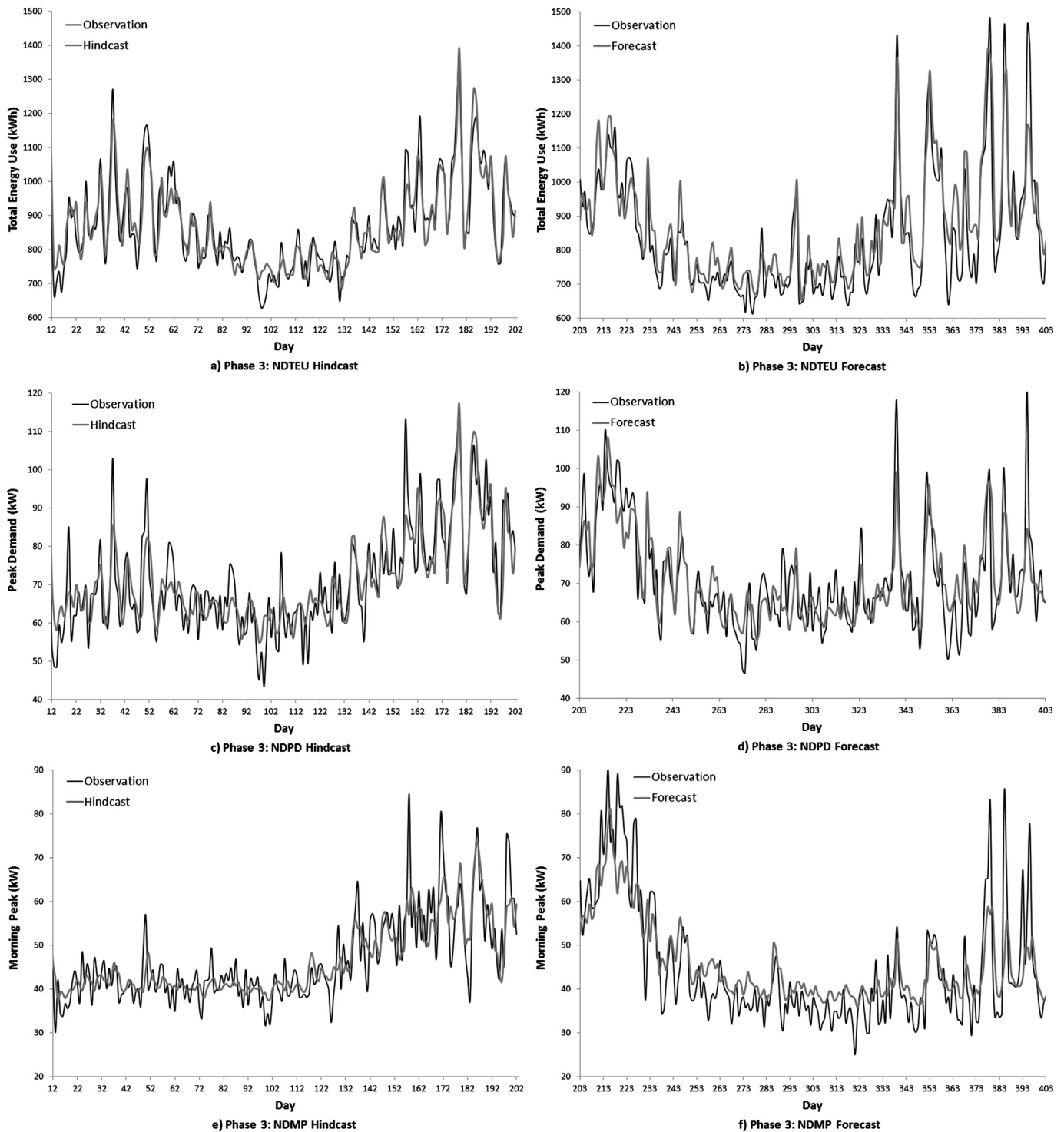


Fig. 4. Phase 3 ARIMAX models.

**Table 3**  
Correlation clustering results.

Cluster	1	2	3	4	5	6	7	8	9
Phase 1	33	38	33	37	34	30	57	54	33
Phase 2	21	23	36	36	43	32	51	64	33
Phase 3	58	67	36	34	46	29	27	25	27

greater than 0.9, respectively. As the aggregated  $R^2$  statistics increases the performance of the models coincide. For each phase's expert system there were between 6 and 14 days (approx. 2–4% of sample) where the accuracy of the forecasts deviated from observations such that the forecasts were considered to be of low accuracy ( $R^2 < 0.5$ ).

Fig. 6 displays a 7 day period comparing Phase 1's observations verses hindcast. Each day, highlighted by vertical lines, is an

**Table 4**  
NN classification accuracy.

Cluster		1	2	3	4	5	6	7	8	9	Overall
Phase 1	Accuracy (%)	42.85	77.41	89.28	77.41	68.00	100.00	86.36	87.50	73.07	77.98
	Miss (%)	6.45	3.58	2.15	5.01	3.22	2.86	5.73	9.31	3.94	4.69
Phase 2	Accuracy (%)	0.00	61.90	93.75	74.19	89.47	74.07	100	85.41	95.65	74.93
	Miss (%)	9.31	5.01	2.15	4.30	2.50	2.50	7.88	7.16	4.30	5.01
Phase 3	Accuracy (%)	81.63	81.13	66.66	50	82.85	62.96	31.57	100	79.16	70.66
	Miss (%)	5.37	11.11	7.16	6.45	9.67	4.65	4.65	3.22	2.15	6.04

**Table 5**  
Hindcast accuracy.

	$R^2$	RMSE (W)	Correlation	MAPE
Phase 1	0.86	5210	0.93	12%
Phase 2	0.88	4346	0.94	11%
Phase 3	0.87	5231	0.94	12%

independent hindcast. What can be observed is that the expert system is able to classify a set of input variables and use the characteristic demand profile of the corresponding cluster to produce a hindcast with a reasonable degree of accuracy. Differences between the expert systems' hindcasts and observations can be attributed to

randomness in the observation and the smooth curve of the characteristic demand profiles.

### 7.5. Validation accuracy statistics

To simulate real world performance, forecasted information was used as input variables over the training set period. Each of the forecasts incorporated as input variables contain error. This necessitates that the performance of the expert system will be hampered due to the compounding effect of the errors. Table 6 contains the accuracy statistics for simulated real world performance. In comparison to the expert systems' hindcasts, its accuracy decreased slightly which is in line with expectations. The  $R^2$  statistics

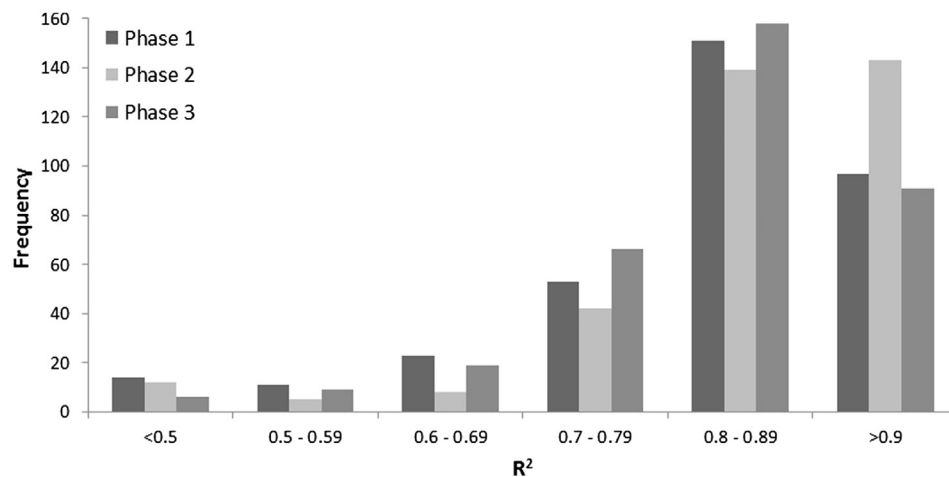


Fig. 5. Demand profile hindcast accuracy distribution.

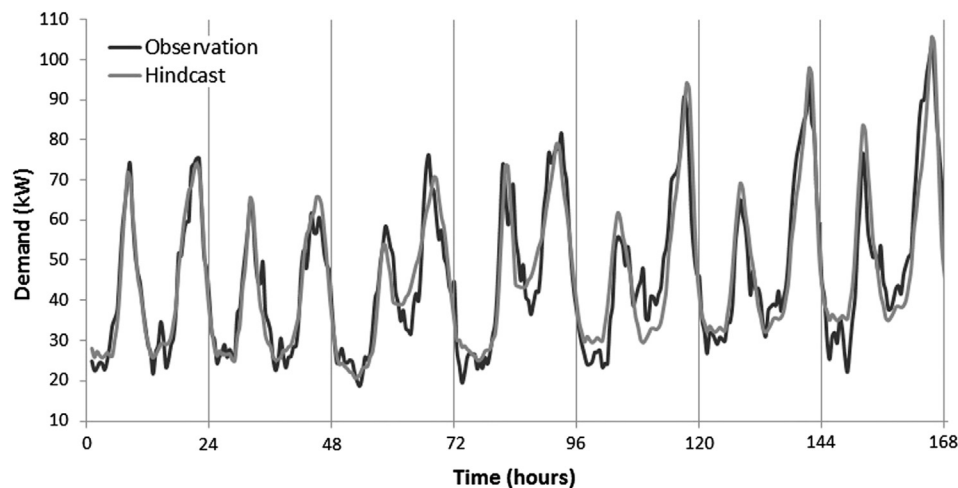


Fig. 6. Demand profile hindcast verses observation overlay.

**Table 6**  
Expert system demand forecast validation statistics.

	$R^2$	RMSE (W)	Correlation	MAPE
Phase 1	0.81	6104	0.90	13.5%
Phase 2	0.84	4990	0.92	12.5%
Phase 3	0.82	6139	0.91	13.5%

decreased by 4–5 points; the RMSE increased by 500 W–900 W; correlation decreased by 2–3 points; and, MAPE increased by 1.5%. Nonetheless, with  $R^2$  statistics ranging from 0.81 to 0.84, correlations ranging from 0.90 to 0.92 and MAPE ranging from 12.5% to 13.5%, the expert systems exhibit a reasonable level of accuracy.

Fig. 7 displays the  $R^2$  statistic distribution for each phase's expert system when the accuracy of each day's forecast is calculated independently. There were 349 forecasts per phase. As expected, the number of high accuracy forecasts ( $R^2 > 0.9$ ) decrease by 46–60 in the simulated real world performance. The number of low accuracy forecasts increased by 23–34.

The ability of the expert system to correctly forecast a demand profile for a future day is dependent on the accuracy of the forecasted information. Fig. 8 compares the normalised error of Phase 2's NDPD model against  $R^2$  statistics of the demand profile forecasts for the last 50 days of 2012. As the error of the forecasted

information increases (days 334–340) the ability for the system to select a characteristic demand profile and post-processes it such that it produces a reasonable forecast decreases. In turn, it can be stated that the forecast accuracy of the expert system can be improved if the error of the forecasted information decreases.

## 8. Conclusion

This paper presented an expert system that was developed to forecast demand profiles in residential LV distribution networks in order to overcome potential issues created by high variance and frequent random shocks. The expert system was constructed by the combination of demand profile property connection forecasts (i.e. TEU, peak demand and morning peak), correlation clustering, NN discrete classification and post-processing. The expert system operates by classifying input variables and calling a corresponding characteristic demand profile as the forecast. The post-processing component adjusts the forecast to better conform to demand profile property forecasts.

The expert system was trained using demand data from an LV residential transformer supplying 128 customers located in Brisbane, Australia. The expert system exhibited high hindcast accuracy with  $R^2$  ranging from 0.86 to 0.87 and MAPE ranging from 11% to

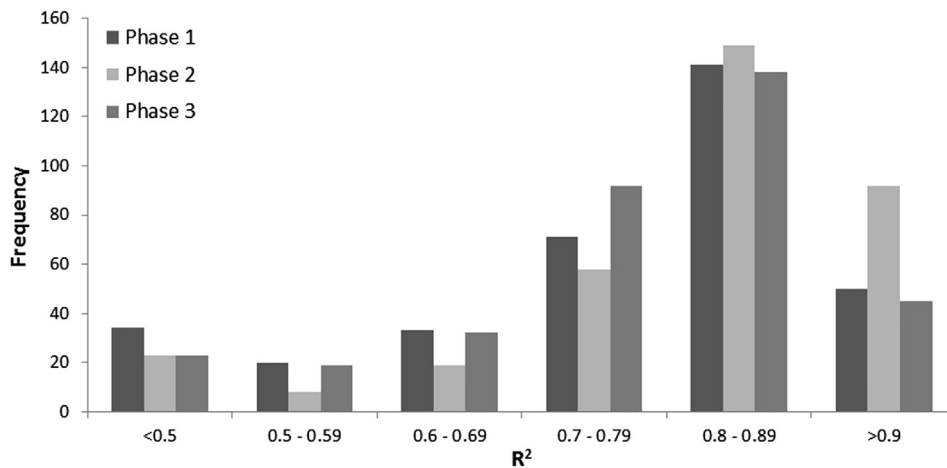


Fig. 7. Demand profile forecast accuracy distribution.

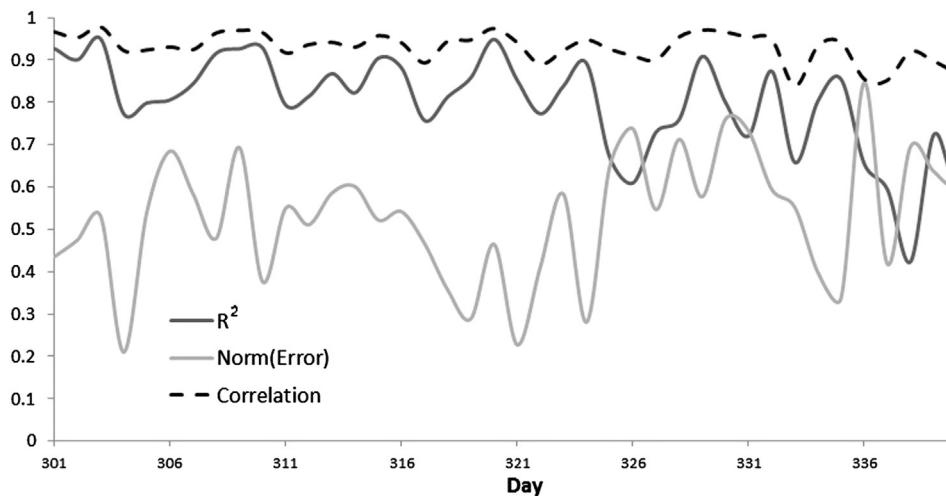


Fig. 8. Error analysis.

12% across the network's phases. An overlay (Fig. 5) of the hindcasted demand profiles and actual observations displayed that the expert system can correctly replicate the shape and magnitude of observed data. When analysing simulated real world performance by using forecasted input variable information, the  $R^2$  statistic was reduced to a respectable 0.81 to 0.84 and MAPE increased to 12.5%–13.5%. The accuracy of the expert system only slightly decreased in comparison to the hindcast, which is typical. When analysing the contributors to the accuracy decrease, it was noted that the system was susceptible to instances of poor demand profile property connection forecasts. This effects both the NN discrete classification and post-processing components of the system.

To improve the expert system's future accuracy requires increases in accuracy of the demand profile property forecasts. This process would involve further detailed analysis of the existing data and further collected information to ascertain whether or not periods of poor property connection demand profile forecasts are due to exogenous shocks or anomalous behaviour. The authors intend to collect further data from the LV transformer and the 128 residential smart meters to provide a more comprehensive dataset for refining the expert system.

Future work will integrate the demand forecasting expert system into an energy management control algorithm to schedule the charging and discharging of BESS in LV residential distribution networks.

### Acknowledgements

The authors are grateful to ENERGEX for providing the data which made this research possible.

### References

- [1] Engle R, Mustafa C, Rice J. Modelling peak electricity demand. *J Forecast* 1992;11(3):241–51.

- [2] Darbellay G, Slama M. Forecasting short-term electricity demand: do neural networks stand a better chance? *Int J Forecast* 2000;16:71–83.
- [3] Ringwood J, Bofelli D, Murray F. Forecasting electricity demand on short, medium and long time scales using neural networks. *J Intell Robot Syst* 2001;31:129–47.
- [4] Taylor J, Buizza R. Using weather ensemble predictions in electricity demand forecasting. *Int J Forecast* 2003;19:57–70.
- [5] Mirasgedis S, Sarafidis Y, Georopoulou E, Lalas D, Moschovits M, Karagiannis F, et al. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy* 2006;31:208–27.
- [6] Taylor J. Triple seasonal methods for short-term electricity demand forecasting. *Eur J Oper Res* 2010;204:139–52.
- [7] Hippert H, Pedreira C, Souza R. Neural networks for short-term load forecasting: a review and evaluation. *IEEE Trans Power Syst* 2001;16(1):44–55.
- [8] Beccali M, Cellura M, Brano V, Marvuglia A. Forecasting daily urban electric demand profiles using artificial neural networks. *Energy Convers Manag* 2004;45:2879–900.
- [9] Hippert H, Hunn D, Souza R. Large neural networks for electricity load forecasting: are they overfitted? *Int J Forecast* 2005;21:425–34.
- [10] Sousa J, Neves L, Jorge H. Assessing the relevance of demand profiling information in electricity load forecasting based on neural network models. *Electr Power Energy Syst* 2012;40:85–93.
- [11] Kassaei H, Keyhani A, Woung T, Rahman M. A hybrid fuzzy neural network bus load modelling and prediction. *IEEE Trans Power Syst* 1999;14(2):718–24.
- [12] Abrahm A, Nath B. A neuro-fuzzy approach for modelling electricity demand in Victoria. *Appl Soft Comput* 2001;1:127–38.
- [13] Cavallaro F. Electric load analysis using an artificial neural network. *Int J Energy Res* 2005;29:377–92.
- [14] Khotanzad A, Afkhami-Rohani A, Lu T, Abaye A, Davis M, Maratukulam D. ANNSTLF – a neural network based electric load forecasting system. *IEEE Trans Neural Network* 1997;8(4):835–46.
- [15] Fan S, Chen L. Short-term load forecasting based on an adaptive hybrid method. *IEEE Trans Power Syst* 2006;21(1):392–401.
- [16] Espinoza M, Joye C, Belmans R, Morr B. Short-term load forecasting, profile identification and customer segmentation: a methodology based on periodic time series. *IEEE Trans Power Syst* 2005;20(3):1622–30.
- [17] Konjic T, Miranda V, Kapetanovic I. Fuzzy inference systems applied LV substation load estimation. *IEEE Trans Power Syst* 2005;20(2):742–9.
- [18] Dai W, Wang P. Application of pattern recognition and artificial neural network to load forecasting in electric power system. In: *Third International conference on natural computation*; 2007, vol. 1. IEEE; 2007. pp. 381–5.
- [19] Australian Bureau of Meteorology. Climate statistics for Australian locations. Australian Bureau of Meteorology; 2013 [last accessed 22.07.13]. URL link available at: [www.bom.gov.au/climate/averages/tables/cw\\_040842.shtml](http://www.bom.gov.au/climate/averages/tables/cw_040842.shtml).