Forecasting High Frequency Intra-Day Electricity Demand using Temperature

James McCulloch* Katja Ignatieva†

August 5, 2017

Abstract

This paper introduces a Generalised Additive Model (GAM) to link high frequency intra-day (5-minute) aggregate electricity demand in Australia to the time of the day and intra-day temperature. We show a superior model fit when using Daylight Saving Time (DST), or clock time, instead of the standard (solar) time. We also introduce the time weighted temperature model that relates instantaneous electricity demand sensitivity to temperature as a function of the daily activity cycle. The results on DST and time weighted temperature modelling are novel in the literature and are important innovations in high frequency electricity demand forecasting. The overall accuracy of the proposed GAM specification in predicting demand is comparable to the accuracy of the commercial demand forecasting model used by the Australian Energy Market Operator (AEMO). The parsimonious GAM model provides a solid foundation for the development of more elaborate models for forecasting high frequency electricity demand.

JEL Classification

Keywords: High Frequency, Electricity, Instantaneous Demand, Temperature, Generalised Additive Model (GAM).

*Quantitative Finance Research Centre (QFRC), University of Technology, Sydney, NSW 2007, Australia. Email: james.duncan.mcculloch@gmail.com (corresponding author)

†UNSW Australia, Business School, School of Risk and Actuarial Studies, Sydney, NSW 2052, Australia. Email: k.ignatieva@unsw.edu.au

1 Introduction

Electricity consumption and demand largely depend on two variables - economic and climate conditions. We explore both of these variables using 5 minute intra-day data. Weather conditions have been widely explored in the past decade, and documented to play a crucial role when dealing with forecasting electricity consumption and demand. A non-exhaustive list of recent literature dealing with the impact of weather variables on electricity consumption includes Pardo et al. (2002), Manera and Marzullo (2005), Giannakopoulos and Psiloglou (2006), Bessec and Fouquau (2008), Beccali et al. (2008), Lam et al. (2008), Miller et al. (2008), Wangpattarapong et al. (2008), Psiloglou et al. (2009), Akil and Miyauchi (2010), Pilli-Sihvola et al. (2010), Włodarczyk and Zawada (2010) and Bašta and Helman (2013). Among all variables that can be used to explain potential variations in electricity consumption, the outside temperature is demonstrated to be the most important weather variable, see e.g. Lam et al. (2009) and Moral-Carcedo and Pérez-García (2015). During cold winter months, electricity demand increases due to electrical heating, whereas during hot summer months air conditioners and coolers increase electricity consumption. Other variables, such as sunshine hours, rainfall, wind speed, humidity, cloudiness etc. are documented to have a much lower impact on demand, see e.g. Molnár (2011), Bašta and Helman (2013) and Moral-Carcedo and Pérez-García (2015). Furthermore, focusing exclusively on the temperature allows us to avoid potential collinearity problems when simultaneously employing several weather variables as explanatory variables in the regression modelling[1], see e.g. Lam et al. (2009) and Moral-Carcedo and Pérez-García (2015).

In order to study the impact of temperature on electricity demand, the introduced and documented approaches primarily deal with heating degree days (HDD) and cooling degree days (CDD), refer to Al-Zayer and Al-Ibrahim (1996), Sailor and Munoz (1997), Valor et al. (2001), Sailor (2001), Pardo et al. (2002), Amato et al. (2005), Xiao et al. (2007). This has proven to be efficient since it takes into account the non-linear relationship observed between electricity consumption and outside temperature. The other strand in the literature models these non-linearities by splitting the data into four different seasons, and modelling linear relationships within each season individually. This approach has been utilised, in combination with wavelet analysis, in e.g. Bašta and Helman (2013). The results documented in the literature are typically mixed, and depend on the region or country under consideration, the frequency of the data, as well as whether the analysis is performed at an aggregate level, or split by sector (residential vs. commercial). The results from existing literature are detailed below.

[1]For example, temperature is expected to be correlated with sunshine hours and cloudiness.

For the *residential sector* Blázquez et al. (2013), using annual data for electricity consumption in Spain find low impact of HDD and CDD on electricity demand. Also, using annual data, Dergiades and Tsoulfidis (2008) find similar result for the U.S. and Hondroyiannis (2004) for Greece, both papers concluding that short-run demand elasticity to temperature is lower than long-run demand elasticity to temperature. Zachariadis and Pashourtidou (2007) document an opposite result for Greece, finding weather fluctuations to be the most significant cause of short-term variation in electricity consumption. Rhodes et al. (2014) use a sample of 103 Texas (U.S.) homes, and demonstrate that electricity demand fluctuates differently during the day, depending on the time of year, weather, occupant behaviour (at home) and schedules. Similarly, Sandels et al. (2014) show for the case of Sweden that electricity consumption is more sensitive to temperature during hot summer months than during cold winter months.

For the *commercial sector*, Psiloglou et al. (2009) use hourly data for electricity consumption in Greece (proxied by Athens) and U.K. (proxied by London) to find that in both regions the dependence between the outside temperature and demand is non-linear, but for Greece, the relationship is U-shaped with a minimum around 20 degrees celcius and two maxima; while for UK, there is only one maximum. Lam et al. (2008) use Hong Kong data and separate the data into commercial and residential. The authors show that commercial sector tends to have a longer cooling season than the residential sector, and that electricity consumption in residential sector increases strongly from May to October. Asadoorian et al. (2008), on the other hand, document that non-residential electricity demand in mainland China is not responsive to the mean temperature. Zachariadis and Pashourtidou (2007) find that in case of Greece electricity demand in commercial sector is not much affected by changes in weather, which is the opposite to the result they find for the residential sector. Similarly, Moral-Carcedo and Pérez-García (2015) study disaggregated data by sector using daily data for Spain and show that firms' aggregate electricity demand is rather insensitive to temperature with slight variations across firms in different service sectors. Sailor (2001) uses monthly per capita electricity consumption for the U.S. and documents significant impact of weather conditions on per capita electricity consumption. Hong et al. (2013) and Zhou et al. (2014) find similar results for the U.S. when using state-level energy demand per building.

At an *aggregate level*, Tung et al. (2013) use daily aggregate energy demand for Taiwan and show that an increase in temperature leads to an increase in average electricity demand. Hekkenberg et al. (2009) use data for the Netherlands and show that electricity demand peaks during cold winter months and declines to its minimum during hot summer months, but predict the development of an additional peak during the summer. Bessec and Fouquau (2008) use monthly demand data

on electricity consumption for 15 European countries. Using a panel threshold regression model, the authors confirm the non-linearity of the link between electricity consumption and temperature, which is more pronounced in warm rather than in cold countries. Mirasgedis et al. (2006) use high frequency demand data sampled at an hourly frequency for Greece and show that especially in the hot summer months electricity demand sees a significant increase. Moral-Carcedo and Vicéns-Otero (2005) use daily data for Spain and show that an increase in consumption is attributed to an extensive use of heating and cooling equipment during winter and summer, respectively. Similar results are found in Pardo et al. (2002) for the daily demand in Spain.

This paper introduces a parsimonious Generalised Additive Model (GAM), refer to Wood (2006), to link the intra-day (5-min) aggregate demand to the time of the day, temperature and time of the year. High frequency data has given us interesting and novel insights (as detailed below) into demand forecasting. In particular, our results allow us to characterise the high frequency relationship between electricity consumption and temperature. To our knowledge, none of the papers in the existing literature model demand and temperature data at such high (5-minute) frequency.

We model demand using daylight saving time (DST), i.e. clock time, and standard (astronomical) time and show that using the DST time provides a significant improvement to the model fit. We explain how and why model fit improves even further when we introduce the *time weighted temperature model*, which assigns different temperature signal weighting based on the DST time. This relates the magnitude of the temperature demand signal with the daily activity cycle. The motivation behind using the time weighted temperature model is the observation that electricity demand attributed to temperature variation away from the maximum comfort temperature (20.0 degrees celcius[2]) is time sensitive. Our proposed methodology suggesting to weight temperature demand signal depending on the DST time (daily activity cycle) is confirmed when using cross-sectional regressions estimated at each (5-minute) time interval, resulting in cross-sectional daily time dependent demand. We observe that the minimum morning sensitivity is at 4:00am, the morning maximum is reached at 9:00am and the night decline begins at 18:30pm. The results on DST and time weighted temperature modelling are novel in the literature and are important innovations in high frequency electricity demand forecasting.

The overall accuracy of the our parsimonious GAM model is evaluated against the commercial demand forecasting model used by the Australian Energy Market Operator (AEMO) for modelling expected intra-day electricity demand in the New South Wales (NSW)/Sydney market. The forecast

[2]This threshold is chosen empirically to provide the optimal model fit.

of this commercial model is used by electricity generators to price bids for expected intra-day wholesale demand the next day. The specification of the AEMO forecasting model is commercial property and is not public, however the AEMO intra-day demand forecasts are published online. The standard deviation of the AEMO's forecast error is 2.6%. Our parsimonious GAM model leads to a comparable standard deviation of the forecast error corresponding to 4.3%. This result is excellent given the limitation of temperature data to only one temperature figure for the entire state of NSW. The parsimonious GAM model provides a solid foundation for the development of more elaborate and accurate models for forecasting high frequency electricity demand.

The remainder of the paper is organised as follows. Section 2 describes data used in our analysis. GAM model specification and its variations tested in the paper are introduced in Section 3. An extensive empirical analysis demonstrating the quality of fit of the proposed models to the entire data set is presented in Section 4. Estimation results for the seasonal demand models that fit data by month are summarised in Section 5. Section 6 deals with the prediction results for the electricity demand, and Section 7 concludes the paper.

2 Data Description

The data we use is an intra-day electricity demand in megawatts (Mw), available at 5 minute intervals for the Australian state of New South Wales (NSW) and the Australian Capital Territory (ACT) for the year 3-February-2014 to 2-February-2015. This is aggregate data (i.e. including households, companies, industrial and public sectors) that has been downloaded from the Australian Energy Market Operator (AEMO) website.[3] The electricity demand observations are merged with instantaneous temperature data over the same period and frequency. The temperature data was obtained from the Australian Government Bureau of Meteorology.[4]

Preliminary analysis of this data shows that time dependent intra-day variation of electricity is significantly different for weekends and public holidays compared to normal business days. This is an intuitive result if we regard time during the day as a proxy for economic and personal activity. We did not find significant variation in intra-day demand for different days of the week (Monday to Friday) for business days. Therefore, because we are primarily interested in modelling the high

[3]Current and archived forecast reports are available from http://`www.nemweb.com.au/REPORTS/CURRENT/` and `http://www.nemweb.com.au/REPORTS/ARCHIVE/`, respectively. Python software has been used to automatically download these forecasts.

[4]Available from `http://www.bom.gov.au`.

frequency relationship between temperature and demand in this paper, we have restricted the data to business days only. The restriction to business days gives 250 days of data and each day of data has 288 five minute demand observations, from 00:00-00:05 until 23:55-24:00. Thus, a total of 72,000 five-minute demand and temperature data points will be used for the empirical analysis.

The 5-minute temperature data is recorded in the Sydney suburb of Homebush. The location of this suburb is close to the population centre of the Greater Sydney urban area. Greater Sydney is 12,367.7 square kilometres and extends east to west from the Pacific ocean coast to the Blue Mountains (100km) and north to south from Berowra to Picton (110km). Therefore, if we assume that the Homebush temperature observations represent the instantaneous temperature in Greater Sydney, then this temperature observation is valid for 61% of population of the NSW/ACT electricity demand area. However, it should be noted that the assumption that we can represent temperature related electricity demand in the NSW/ACT demand area with a single temperature is a deliberate simplification. The goal of this paper is to introduce a rigorous parsimonious forecasting model that can be used as the foundation for more elaborate forecasting models. An obvious improvement to the accuracy of modelling electricity demand as a function of temperature would be to use multiple temperature (and potentially humidity) measurements from different suburban and urban areas.

The NSW/ACT electricity demand area combines both the NSW and ACT populations for a total population of 7.95 million (2015). Within the NSW/ACT electricity demand region there are four large urban areas: Greater Sydney with population 4.84 million (2015 estimate; 2011 census 4.39 million), Greater Newcastle with population 550,000 (2015 forecast), Canberra/ACT with population 385,000 (2015 estimate; 2011 census 356,000) and Wollongong with population 290,000 (2015 forecast). Wollongong and Newcastle are coastal port cities like Sydney, with Wollongong 100km south of Sydney and Newcastle 200km north of Sydney. The population of the NSW state is 7.55 million (2015 estimate; 2011 census 7.21 million). Canberra/ACT is 300 km inland from Sydney and has a continental climate with a much greater variation in temperature than Sydney and the other NSW coastal cities (Newcastle and Wollongong).

To provide an idea of the relationship between demand and temperature data, we show in Figure 1 winter and summer patterns of demand (top panel) and temperature (bottom panel) over the five business days of a typical week. In both, the summer (red line, 12-January-15 to 16-January-15) and winter (blue line, 14-July-14 to 18-July-14) graphs we observe a cyclical pattern in evolution of temperature and demand throughout the week. The winter demand graph experiences two daily peaks, which correspond to an increasing usage of heating during the cold morning and evening hours, while the demand during the day (when the outside temperatures are relatively

high) drops to a lower level. The summer demand graph shows peak demand during the hot afternoon hours caused by the use of coolers and air-conditioners. The winter and summer demand patterns suggest a minimum demand 'comfort' temperature of $20°C$. This result is well known in the literature and is based on Heating Degree Day (HDD) and Cooling Degree Day (CDD) temperature dependent energy derivatives. The magnitude of the difference between the current temperature and the 'comfort' temperature (abs$[Temp - 20]$) is strongly correlated with demand. For example, Monday January 12, 2015 was a hot day and the corresponding demand graph shows a pronounced spike mid-afternoon. The double morning and evening peaks of the winter pattern are pronounced because temperatures are well below $20°C$, with a dip in demand in the afternoon as the temperature rises. However, it is interesting to note that these peaks do not correspond to the minimum temperature early in the morning, but correspond to *lower temperatures when people are active*. This is clearly seen by examining the winter demand pattern in the early morning. The minimum demand for Tuesday, July 15, 2014 at 4am is essentially the same as the minimum demand for Thursday, July 17, 2014 at 4am, even though Tuesday 4am is considerably colder than Thursday 4am. The reason for this is intuitive: there are low levels of activity at 4am and the effect of the stronger temperature demand signal on Tuesday is attenuated by low personal and economic activity. Therefore, the *temperature demand signal is time dependent*. The relationship between temperature and demand shown in Figure 1 is our primary motivation for modelling intra-day electricity demand as a function of temperature.

3 Model Specification

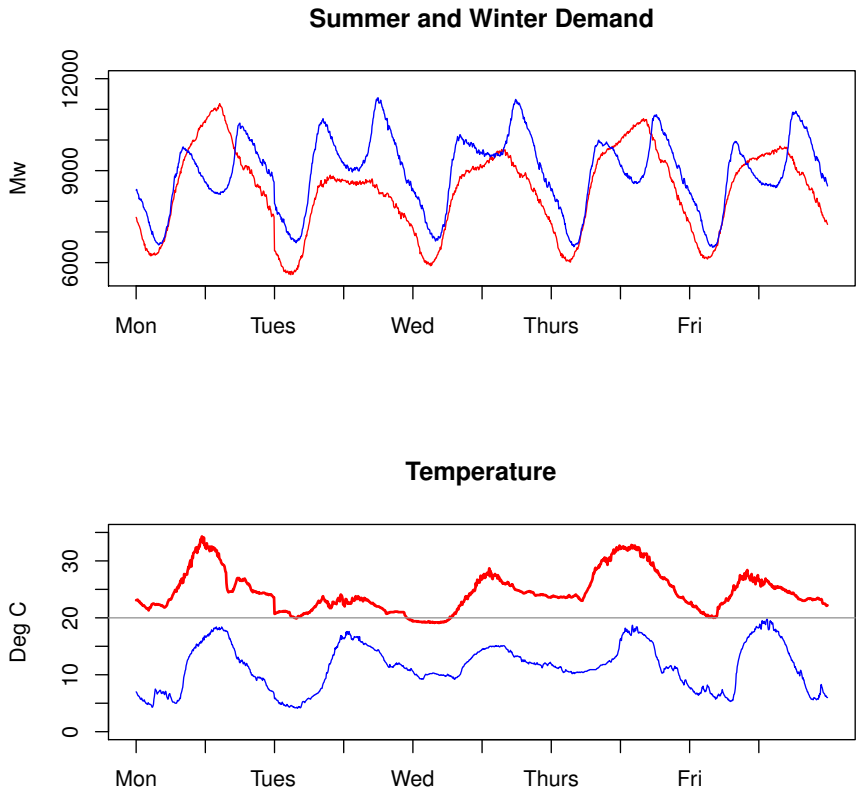
As stated in the introduction, the objective of our analysis is to develop a comprehensive high frequency modelling framework to link electricity demand to the outside temperature. For these purposes we introduce the Generalised Additive Model (GAM), see e.g. Hastie and Tibshirani (1990), Wood (2006). Specifically, at each time point t, $t = 1, ..., N$ with N being the total number of observations, we link the demand to the temperature in the following way:

$$\textbf{Model 1} \quad D_t = \beta_0 + s(Time_t) + \beta_1 |Temp_t - 20.0| + \beta_2 Year_t + \varepsilon_t. \tag{3.1}$$

The dependent variable D_t is the instantaneous electricity demand in Megawatt/hour (Mw/h). To be precise; if the integrated demand in each 5 minute record were extended to 1 hour by multiplying by 12, then the power consumed would be D_t.

It is intuitive and true (and will be shown below) that personal and economic activity is linked to

Figure 1: Summer (red line 12-Jan-15 to 16-Jan-15) and Winter (blue line: 14-Jul-14 to 18-Jul-14) demand (top panel) and temperature (bottom panel).



Note. Top panel: Summer demand (red line) and winter demand (blue line); bottom panel: temperature in summer (red line) and winter (blue line). The 'comfort' (black line) in the temperature graph is $20^{\circ}C$. The graph shows that the temperature demand signal is time dependent.

daylight saving time (DST), i.e. clock time rather than the actual (standard) time. The standard time is astronomical time. Daylight saving time (+1 hour) in Sydney commences at 2am on the first Sunday in October and the change from daylight saving (-1 hour) to standard time is 3am on the first Sunday in April. In our sample period (from February 3, 2014 to February 2, 2015) daylight saving ends on April 6, 2014 (clocks turned back from 3am to 2am) and starts on October 5, 2014 (clocks turned forward from 2 am to 3 am).

Following the convention commonly used in spreadsheets; $Time_t$ is a number in the interval $[0, 1)$ where 0 is the time recorded for the electricity demand in the 5 minute period 00:00-00:05, $1/(12 * 24) = 0.0034722$ is the time recorded in the period 00:05-00:10 and $((12 * 24) - 1)/(12 * 24) = 0.9965278$ is the time recorded for the 5 minute period 23:55-00:00. There are two time fields for

each demand record in the data; standard (astronomical) time and the DST. During the period when DST is active, the DST field is advanced by 1 hour or $0.041667 = 1/24$. For example, for the 5 minute time period on the 3rd of February 2014 (DST is active) where the standard time is recorded as 00:00-00:05 ($Time_t = 0.0$), the DST time is recorded as 01:00-01:05 ($Time_t = 0.041667$). The instantaneous electricity demand (D_t) recorded for this 5 minute period was 7135.67 megawatt/hour. Outside of the period where DST is active, the DST and standard time fields are equivalent.

It is assumed that the *temperature independent electricity demand* $s(Time_t)$ is a daily periodic cyclic empirical function of $Time_t$ over the sample period. We use GAM regression to determine the periodic function $s(\cdot)$ of $Time_t$. This periodic function is a *cyclic cubic spline*. A cyclic cubic spline function is a piecewise cubic function continuous up to second derivatives at the knots. At the endpoints of each daily cycle, the function values and derivatives up to the second order are equal, which creates a smoothed periodic function.[5] We notice that the function $s(\cdot)$ can be specified with a smoothing parameter (number of spline knots), or the number of degrees of freedom (df, the number of spline knots - 2), which is assumed to be larger than one (with $df = 1$ corresponding to a linear fit). The df parameter is chosen in such a way that it leads to the best goodness-of-fit measured by the Akaike Information Criterion (AIC), and we observe that $df = 10$ (see below) is optimal. The GAM in Equation (3.1) can be thought of as a Generalized Linear Model in which part of the linear predictor is specified in terms of a sum of smooth functions of predictor variables (Wood (2006)). This technique is particularly suited to modelling intra-day actual electricity demand as a function of the time dependent electricity demand due to daily personal and economic activity.

The *temperature dependent electricity demand* in Equation (3.1), $|Temp_t - 20.0|$, is the absolute value of the difference of the recorded temperature and $20°C$. We refer to the constant temperature of $20°C$ as the 'comfort' or minimum demand temperature. The value of $20°C$ is determined empirically below and is entirely consistent with the literature on CDD & HDD.[6].

The third term $Year_t$ is the linear long term drift in average electricity demand. The independent variable $Year_t$ is the scaled time of the year with values in the interval $[0, 1)$, where 0 corresponds to the first data record of electricity demand for the 5 minute standard time period 3-February-2014 00:00-00:05 (DST 4-February-2014 00:55-01:00) and $((365 * 288) - 1)/(365 * 288) = 0.999990487$ corresponds to the final electricity demand record for the standard time period 2-February-2015

[5] R function *s(Time, bs = "cc")* is used.

[6] Although some of the literature suggests that this temperature is slightly lower, corresponding to $18°C$ (see Alaton et al. (2002)).

23:55-24:00 (DST 3-February-2015 00:55-01:00).[7]

The GAM regression models assume the residual term ε_t to be Gaussian with zero mean.[8]

3.1 The Weighted Temperature Demand Signal

Another model, referred to as Model 2, which is aimed to improve the fit of Model 1, is given by the following equation:

$$\textbf{Model 2: } D_t = \beta_0 + s(DST_t) + \beta_1(w(DST_t) * |Temp_t - 20.0|) + \beta_2 Year_t + \varepsilon_t. \tag{3.2}$$

The difference between Model 1 and Model 2 is the term $w(DST_t) * |(Temp_t - 20.0)|$, where the function $w(\cdot)$ is a piecewise continuous sinusoidal function of DST and returns values between 0 and 1. As can be readily seen from the formulation above, this function weights the temperature demand signal where 1 represents the 'full' temperature signal ($|Temp_t - 20.0|$) and 0 completely attenuates the temperature signal. It is intuitive and reasonable that the demand sensitivity to the temperature signal $|Temp - 20.0|$ depends on the human and economic activity. This daily activity cycle can be readily determined by examining the sample daily electricity demand cycle with the temperature demand component removed. In other words, $w(DST_t)$ has approximately the same shape as $s(DST_t)$ and, therefore, like $s(DST)$, is driven by the daily activity cycle. The activity weighting of the temperature demand signal is modelled empirically using cross sectional regressions where the demand sensitivity to temperature is fitted (250 day data points) for each 5 minute period during the day. The cross sectional regressions performed in Section 4.3 clearly show the sensitivity to the daily exogenous demand cycle. The empirical cross sectional results and the approximating function $w(\cdot)$ will be discussed below (in particular, refer to Figure 4). We refer to this model as the *time weighted temperature model*.

The third model uses non-periodic splines to model the non-linear relationship between temperature and demand:

$$\textbf{Model 3: } D_t = \beta_0 + s(DST_t) + h(Temp_t) + \beta_2 Year_t + \varepsilon_t. \tag{3.3}$$

In addition to the cyclic spline function of time $s(DST_t)$ (that will be present in all models), we incorporate a non-periodic (non-cyclic) spline function of temperature $h(Temp_t)$ instead of using the function from Model 1 (second term on the right hand side of Equation (3.2)).

[7] The value of 288 corresponds to the number of observations per day: $12 * 24 = 288$.

[8] For the residual term, using the R GAM regression software, one can select any distribution from the exponential family of distributions.

The fourth model uses non-periodic splines to model the non-linear relationship (interaction) between *time weighted* temperature and demand:

$$\textbf{Model 4: } D_t = \beta_0 + s(DST_t) + h\big(Temp_t * w(DST_t)\big) + \beta_2 Year_t + \varepsilon_t. \tag{3.4}$$

The fifth and most sophisticated model that will be used for the yearly data (i.e. when modelling the entire sample) applies non-periodic splines to model the non-linear relationship of long-term change in demand as a function of the $Year$ using the fitted spline term $k(Year_t)$:

$$\textbf{Model 5: } D_t = \beta_0 + s(DST_t) + h\big(Temp_t * w(DST_t)\big) + k(Year_t) + \varepsilon_t. \tag{3.5}$$

3.2 Seasonal Demand Model

We notice that Models 1 to 5 use the entire data (one year) for the estimation, thus, they all have the $Year_t$ term included in the model. In our empirical section we will also use the seasonal demand model, which caters for non-stationarity of the exogenous time dependent demand, thus, fitting the following regressions for each calendar month:

$$\textbf{Model 6: } D_t = \beta_0 + s(DST_t) + \beta_1 |w(DST_t) * Temp_t - 20.0| + \varepsilon_t. \tag{3.6}$$

The seasonal demand Model 6 in Equation (3.6) is a simple two-term version of Model 2 in which the yearly regression term ($\beta_2 Year_t$) has been removed. This seasonal model will be fitted for each calendar month and is expected to better capture demand fluctuations compared to the yearly demand models (Models 1 to 5).

4 Empirical Analysis

In this section we perform empirical analysis using the demand and temperature data described in Section 2 and models introduced in Section 3.

4.1 Time Dependent Demand is Indexed by DST or Standard Time

We fit Model 1 in Equation (3.1) using DST and standard (astronomical) time and tabulate the results in Tables 1 and 2, respectively.

From both tables we observe that all terms are highly statistically significant. In particular, the regression shows a long-term decline in electrical load demand (the *Year* term in the regression),

Table 1: Model 1 with daylight saving time

	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
Intercept	7795.2***	4.08	1908.55		0.837
$s(DSTime)$				35513 (0.000)	
$\|Temp - 20\|$	108.146***	0.5134	211.29		
$Year$	-319.01***	6.00	-53.15		

Note. Estimation results for regression in Eq. (3.1) (Model 1) where $Time$ variable is given by the standard time; ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

Table 2: Model 1 with standard time

	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
Intercept	7804.9***	4.39	1776.69		0.811
$s(Time)$				29693 (0.000)	
$\|Temp - 20\|$	106.05***	0.51	209.36		
$Year$	-318.8***	6.46	-49.37		

Note. Estimation results for regression in Eq. (3.1) (Model 1) where $Time$ variable is given by the DST; ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

with daily demand falling 319 megawatt/hour over a one year period (-3.5%). This decline in electricity demand is in line with the media release from the Australian Energy Market Operator (AEMO) (2014).[9] [10] The temperature regression term ($|Temp - 20|$) shows the expected positive relationship between the absolute value of the difference of the recorded temperature and $20°C$. Note that this minimum demand temperature is slightly higher than the $18°C$ used as the reference minimum demand temperature in energy derivatives[11]. We notice that using DST as the independent variable for the time indexed daily periodic demand consumption function gives a small but highly significant improvement of fit compared to using standard time, which is reflected in the higher R^2_{adj} and larger value of the F statistic. This result is intuitive as the daily personal and economic demand cycle depends on clock (DST) time rather than standard time.

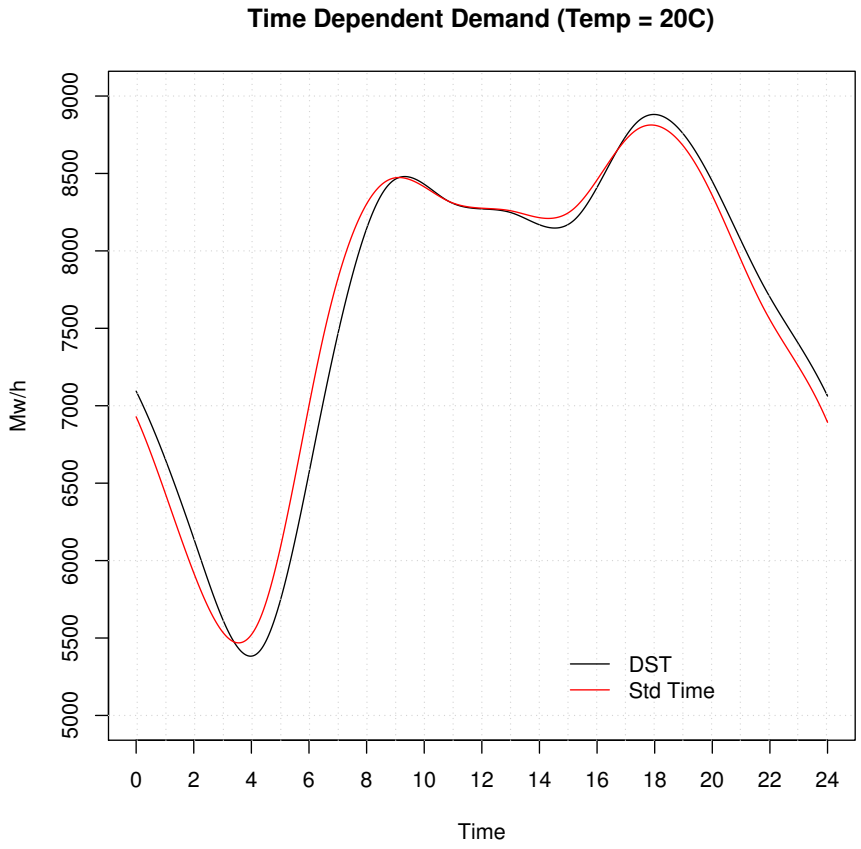
Figure 2 shows daily time dependent electricity demand using DST (black line) and standard time (red line) as indexing variables. This is the cyclic time dependent component of demand fitted with splines, $s(Time_t)$ using Equation (3.1) (Model 1). The $Temp$ variable is set to $20°C$ (no signal) and

[9] *"The 2014 NEFR [National Electricity Forecasting Report] shows reduced residential and commercial consumption in most NEM regions due to strong growth in rooftop photovoltaic (PV) system installations and ongoing energy efficiency savings in response to high electricity prices over recent years."*

[10] *"In 2013-14, rooftop PV results in a 2.9% reduction in consumption from the grid."*

[11] For a description of the over-the-counter (OTC) weather derivatives traded on the Chicago Board of Trade (CBOT) refer to Alaton et al. (2002)

Figure 2: Daily time dependent electricity demand



Note. Daily time dependent electricity demand using DST (black) and standard time (red) as indexing variables. This is the cyclic time dependent component of demand fitted with splines, $s(Time_t)$, using Equation (3.1) (Model 1). The $Temp$ variable is set to $20°C$ (no signal) and $Year$ is set to 0.5. The data covers all business days (250 days) from February 3, 2014 to February 2, 2015.

$Year$ is set to 0.5. The data is all business days (250 days) from 3-Feb-2014 to 2-Feb-2015. The two graphs are very similar in form. However, as expected, the standard time graph (red) leads the DST time graph (black) by about about half an hour showing the effect of Daylight Savings Time, which is in place from February 3, 2014 to April 6, 2014, and from October 5, 2014 to February 2, 2015. Note that if the two regressions where restricted to just the DST period then we would see two identical graphs exactly one hour apart.

4.2 The Minimum Demand 'Comfort' Temperature

We fit Model 1 in Equation (3.1) using different (from $20°C$ used above) minimum demand temperatures $Temp_{min}$ in the temperature dependent term of the demand regression, $\beta_1|Temp_t - Temp_{min}|$

where $Temp_{min} = \{17, 18, 19, 20, 21, 22, 23\}$. The objective is to analyse which minimum demand temperature results in a better fit. The regression results (DST is used as the time index) for the temperature dependent demand term $\beta_1|Temp_t - Temp_{min}|$ are reported in Table 3.

Table 3: Model 1 with different minimum demand temperatures

	β_1 Estimate	Std.error	t-value	R^2_{adj}
$\beta_1\|Temp - 17\|$	95.2***	0.58	163.9	0.795
$\beta_1\|Temp - 18\|$	108.4***	0.54	202.4	0.820
$\beta_1\|Temp - 19\|$	112.0***	0.49	226.2	0.835
$\beta_1\|Temp - 20\|$	108.1***	0.47	229.3	0.837
$\beta_1\|Temp - 21\|$	99.3***	0.46	215.5	0.829
$\beta_1\|Temp - 22\|$	89.4***	0.46	194.9	0.816
$\beta_1\|Temp - 23\|$	79.4***	0.46	172.7	0.801

Note. Estimation results for regression in Eq. (3.1) (Model 1) with different minimum demand; ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

The results show that $20°C$ is the *marginally* optimal constant for the minimum demand temperature since the regression temperature dependent coefficient $\beta_1|Temp_t - 20|$ has a slightly higher t-stat and the regression has a slightly higher R^2_{adj}. However, the difference between $20°C$ and $19°C$ or $21°C$ is very small, which implies that temperature dependent demand is a non-linear function of the difference between the minimum demand temperature and the measured temperature, with small differences producing little or no change in temperature dependent demand. This possibility will be examined in Section 4.5 where we fit Model 3 (Equation (3.3)) that captures the relationship between the temperature and demand via a non-linear function modelled by non-periodic splines.

4.3 Cross Sectional Regressions

One advantage of using high frequency temperature and demand data is that it gives us a large amount of data: We have 250 days, and 288 5-minute observations on each day, which results in a total of 72,000 data points. This allows us to perform cross sectional regressions using 250 observations at each individual 5 minute data point, thus, leading to a total of 288 regressions. Unlike models 1 through 5 specified above which are designed to generalize the characteristics of the data to allow us to predict demand on out-of-sample data, the cross sectional regressions will over-fit the data. Thus, cross sectional regressions are not considered to be a model but instead, are used as a tool to examine data. In fact, the cross sectional regressions provide valuable insights about the characteristics of the data to help us develop the more sophisticated models below.

Specifically, we fit the following linear model 288 times to each 5 minute period during the day:

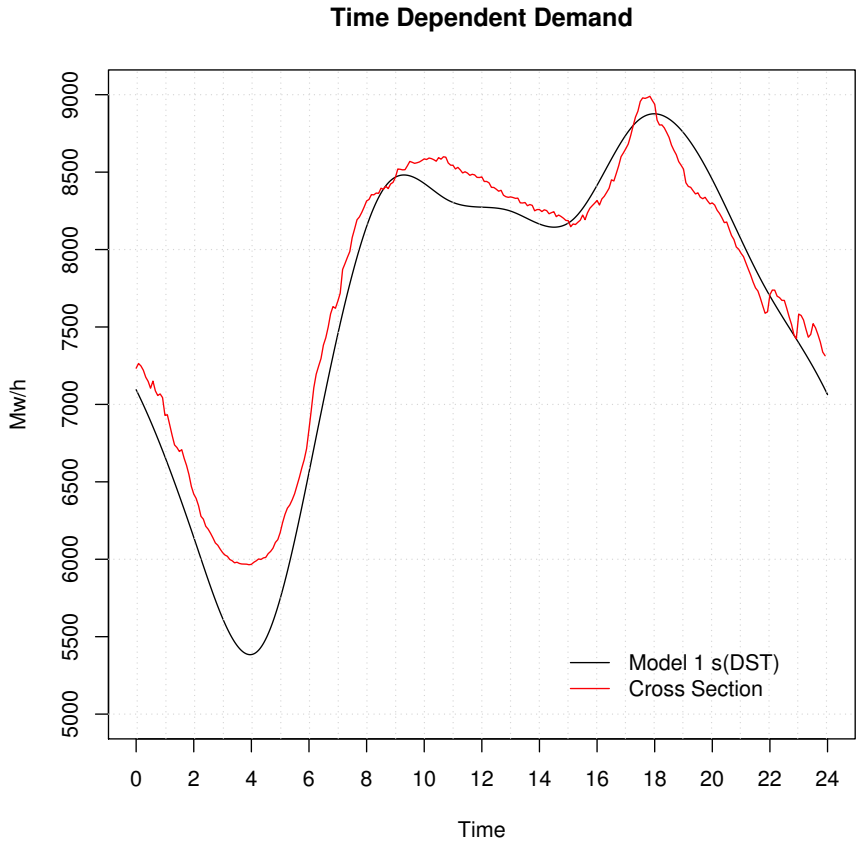
$$\textbf{Cross Section: } D = \alpha^0 + \alpha^1 |Temp - 20.0| + \alpha^2 Year + \varepsilon. \tag{4.1}$$

The cross sectional regressions gives us a 288 point vector of time indexed coefficients. Here, $\alpha_t^0, t \in [1, \ldots, 288]$ is the empirical exogenous daily demand cycle. The empirical time indexed demand sensitivity to temperature is given by coefficients $\alpha_t^1, t \in [1, \ldots, 288]$. The temperature demand weighting function $w(.)$ used in Models 2, 4, 5 and 6 is based on the empirical time indexed demand sensitivity to temperature. Finally, a 288 point vector of year coefficients from the cross sectional regressions $\alpha_t^2, t \in [1, \ldots, 288]$ shows that the yearly decline in demand has been concentrated during the daylight hours (demand replacement with solar power), and in particular, the greatest demand falls have been in peak periods. This indicates that peak demand pricing is causing time insensitive electricity consumers to shift demand to non-peak periods. Figures 3, 4 and 5 summarise coefficient estimates α_t^0, α_t^1 and α_t^2, respectively; they are discussed in details below.

Figure 3 shows the 288 time indexed coefficients $\alpha_t^0, t \in [1, \ldots, 288]$ from the cross sectional regression in Equation 4.1. Each of the 288 fitted coefficients α_t^0 was highly significant with a median t-statistic of 130 and a median standard error of 59. For comparison, the fitted time dependent exogenous component from Model 1, $s(DST)$, is displayed (black line). The comparison shows that Model 1 does a good job of generalizing the time dependent demand cycle with the exception of the daily low demand at 4am, which is underestimated by Model 1 by about 500Mw/h. This is due to the fact that Model 1 does not use a time weighted temperature demand signal $w(.)$ and overestimates the demand effect of cold temperatures at 4am.

Figure 4 shows the cross sectional empirical time demand sensitivity to temperature $\alpha_t^1, t \in [1, \ldots, 288]$ (red line) obtained using cross sectional regressions in Equation 4.1. The y-axis is additional demand in Mw/h generated by each degree variation from the 'comfort' temperature ($|Temp - 20.0|$). Each of the 288 fitted coefficients α_t^1 was highly significant with a median t-statistic of 15.1 and a median standard error of 8.0. For comparison, the piecewise sinusoidal temperature weighting function $w(\cdot)$ used in Model 2 (black line) is also displayed ($w(.) \in [0, 1]$ and is scaled up by $\max[\alpha_t^1] = 235.8$ to facilitate comparison). We are careful not to over-fit the data and deliberately choose a piecewise function to approximate the cross-sectional time demand sensitivity rather than a smoothed version of $\alpha_t^1, t \in [1, \ldots, 288]$. We observe that *electricity demand due to temperature variation away from the maximum comfort temperature (20 C) is time sensitive.* This result, to our knowledge, does not appear in the literature and is a major result in this paper. The pattern

Figure 3: Exogenous time dependent demand: $\alpha_t^0, t \in [1, \ldots, 288]$

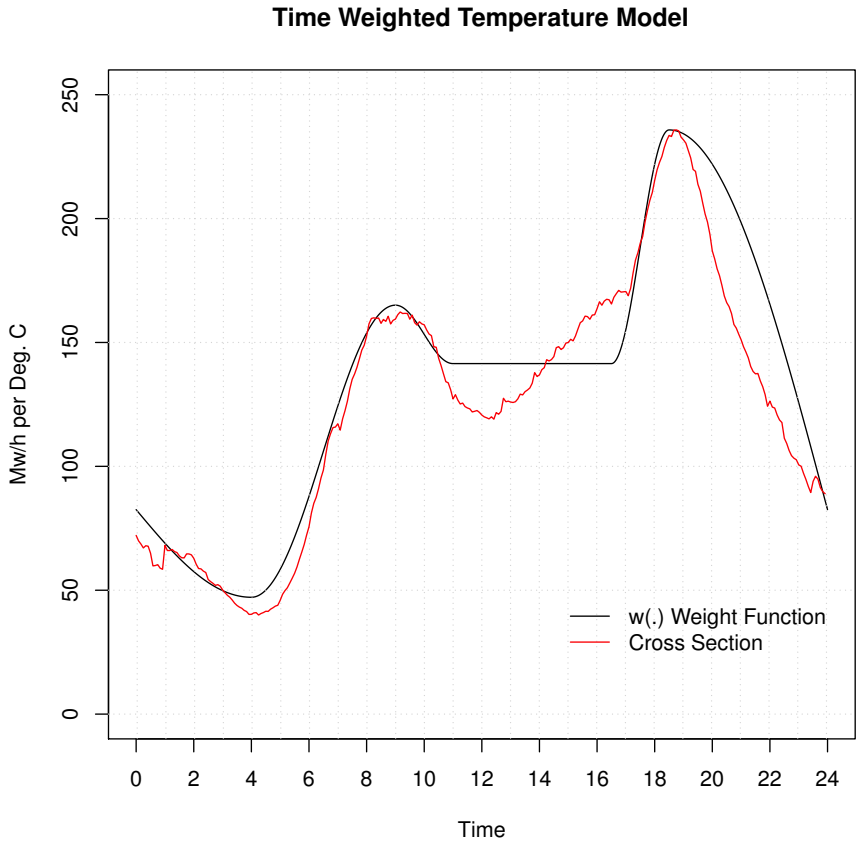


Note. This graph displays the 288 time indexed coefficients $\alpha_t^0, t \in [1, \ldots, 288]$ (red line) from the cross sectional regression in Equation (4.1); black line represents the comparison to Model 1. All coefficients α_t^0 are highly significant. Model 1 performs well in generalizing the time dependent demand cycle with the exception of the daily low demand at 4am.

of the cross-sectional empirical time demand sensitivity to temperature is intuitive and relates to the daily activity cycle. The minimum morning sensitivity is at 4:00am, the morning maximum is reached at 9:00am and the night decline begins at 18:30pm. Thus, an unweighted regression (Model 1) will underestimate the effect of cold temperatures on electricity consumption. We refer also to Figure 6 (discussed below) for different temperature sensitivities using weighted (Model 2) and unweighted (Model 1) regressions.

The Australian Energy Market Operator commented on the decline in electricity demand in 2014 with the following press release. Australian Energy Market Operator (AEMO) (2014): *"[Headline] NEM ELECTRICITY DEMAND CONTINUES DOWNWARD TREND. The 2014 NEFR [National Electricity Forecasting Report] shows reduced residential and commercial consumption*

Figure 4: Time weighted temperature model: $\alpha_t^1, t \in [1, \ldots, 288]$



Note. This graph displays the 288 time indexed coefficients $\alpha_t^1, t \in [1, \ldots, 288]$ (red line) from the cross sectional regression in Equation (4.1). The y-axis is additional demand in Mw/h generated by each degree variation from the 'comfort' temperature ($|Temp - 20.0|$). All fitted coefficients α_t^1 are highly significant. For comparison, the time weighted temperature sensitivity curve $w(\cdot)$ (black line, scaled by 235.8) is also presented. Electricity consumption is less sensitive to temperature variation away from the 'comfort' temperature late at night and early in the morning.

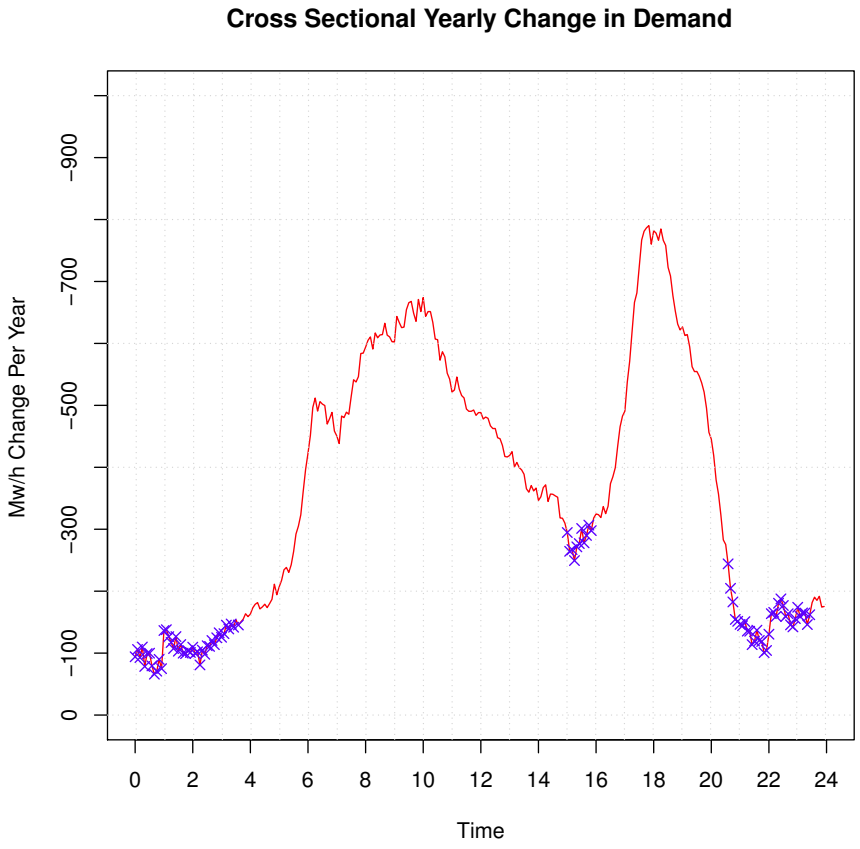
in most NEM regions due to strong growth in rooftop photovoltaic (PV) system installations and ongoing energy efficiency savings in response to high electricity prices over recent years."

If the decline in demand was mainly due to "rooftop photovoltaic (PV) system installations" then we would expect to see the decrease in demand as a sinusoid across the daylight hours with a peak decrease in demand around noon (maximum solar radiation). However, Figure 5 summarising the results for the cross sectional change in yearly demand given by α_t^2, shows that the yearly decline in demand has been concentrated during periods of peak demand. This supports the alternative explanation for the decline proposed by AEMO; "energy efficiency savings in response to high electricity prices over recent years". In particular, peak demand pricing is causing time insensitive

electricity consumers to shift demand to non-peak periods.

The fact that the decline in demand is concentrated in the peak demand periods and not spread more uniformly across the day means that the daily pattern of exogenous time dependent demand ($s(DST)$ in Figure 3) is evolving throughout the year with the morning and afternoon peak demands becoming less pronounced. This has very important implications for regression in Models 1 through 5. An implicit assumption of these models is that exogenous time dependent demand $s(DST)$ is stationary across the year. We now know this is not the case. Thus, the evolution of exogenous time dependent demand $s(DST)$ may be better captured by Seasonal Demand Models developed in Section 5.

Figure 5: Cross sectional change in yearly demand: $\alpha_t^2, t \in [1, \ldots, 288]$



Note. This graph displays the 288 time indexed coefficients $\alpha_t^2, t \in [1, \ldots, 288]$ (red line) from the cross sectional regression in Equation (4.1). The 288 fitted coefficients α_t^2 have a median standard error of 87.7. The coefficients marked with a blue cross on the graph are *not significant* at the 99% confidence level. The y-axis is (inverted) decline in demand in Mw/h for the year. The graph shows that the yearly decline in demand has been concentrated in the morning and afternoon peak periods. This indicates that peak demand pricing is causing time insensitive electricity consumers to shift demand to non-peak periods.

4.4 Model 2 Using Time Weighted Temperature

In this section we report the results from fitting Model 2 (Equation (3.2)), where temperature is time weighted using a constant piecewise continuous smooth function of DST, $w(DST_t)$. The results are reported in Table 4. We observe that Model 2 produces a superior fit compared to Model 1 (Tables 1 and 2), which is evident from the higher value for R^2_{adj} and more significant temperature term (higher value for the t-stat for $|(w(DSTime) * Temp) - 20|$).

Table 4: Model 2 using time weighted temperature

	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
Intercept	7772.4***	3.7	2100.91		0.862
$s(DSTime)$				28250 (0.000)	
$w(DSTime) * \|Temp - 20\|$	216.5***	0.7884	274.57		
Year	-360.55***	5.52	-65.28		

Note. Estimation results for regression in Eq. (3.2) (Model 2); ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

4.5 The Nonlinear Temperature Dependent Demand

To examine the first possibility suggested in Section 4.2 that the relationship between temperature difference and demand is non-linear, we fit Model 3 (Equation (3.3)) where the relationship between the temperature and demand is a non-linear function modelled by non-periodic splines. The results of this regression are reported in Table 5. We observe from the table that the term $h(Temp)$ is highly significant. However, R^2_{adj} from Model 3 is marginally lower compared to Model 2, which is due to the fact that we do not time weight the temperature signal in Model 3 using $w(.)$. Figure 6 shows using the black line the shape of the non-linear relationship between temperature difference and demand. It is clear from this graph that the relationship is 'U' shaped at the minimum demand 'comfort' temperature and this explains the *marginally* optimal $20°C$ minimum demand temperature, as discussed in Section 4.2.

4.6 Time Weighted Temperature Model

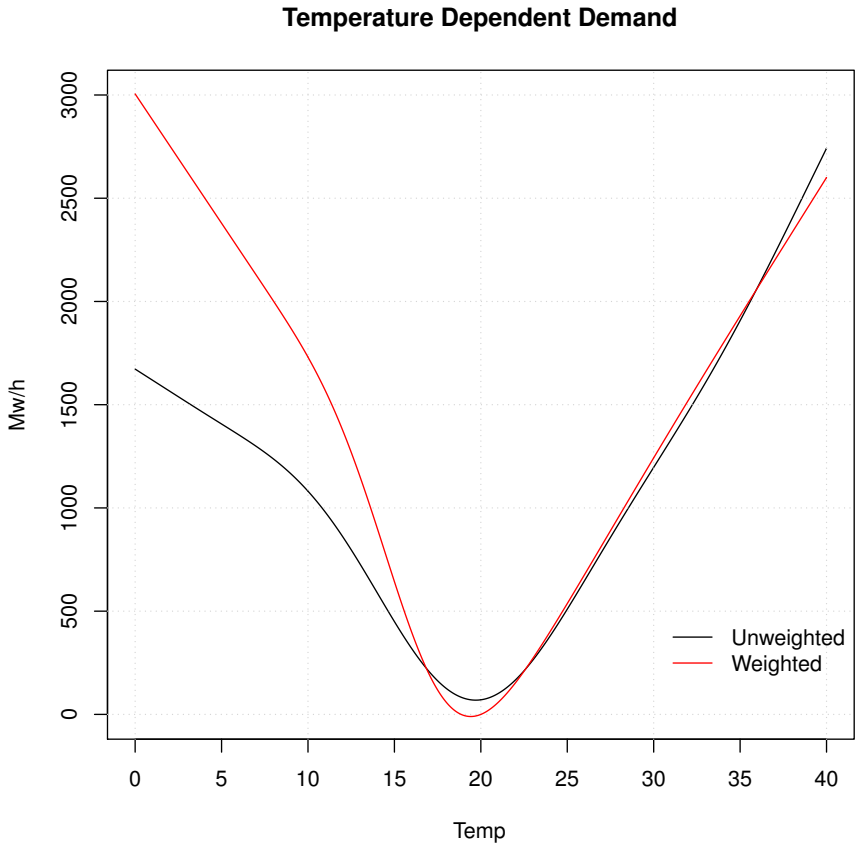
In this subsection we present the results for the time weighted temperature model (Model 4) given in Equation (3.4). Table 6 reports estimation results. The results from the t-tests indicate that both variables, temperature and year are highly statistically significant, and the value of the F-statistic

Table 5: Model 3 using nonlinear temperature.

	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
Intercept	8323.4***	3.4	2435.38		0.841
$s(DSTime)$				30722 (0.000)	
$h(Temp)$				7987 (0.000)	
Year	-338.4***	6.0	-56.17		

Note. Estimation results for regression in Eq. (3.3) (Model 3); ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

Figure 6: Nonlinear temperature dependent electricity demand using non-periodic splines.



Note. Non-linear temperature dependent electricity demand using non-periodic splines. Black line shows the resulted fitted demand when using Model 3 (Equation (3.3)) with the unweighted temperature spline function $h(Temp_t)$. Red line shows the resulted fitted demand when using Model 4 (Equation (3.4)) with the time weighted spline temperature function $h(Temp_t * w(DST_t))$.

for the weighted term $h\big(Temp * w(DST)\big)$ is larger compared to the F-statistic for the unweighted term $h(Temp)$ from Model 3 (Table 5). We also observe the largest (across all considered Models 1 through 4) value for the R^2_{adj} corresponding to 0.869, which again points towards a superior fit

for the weighted temperature model.

Table 6: Model 4 using *time weighted* nonlinear temperature.

	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
Intercept	8328.6***	3.1	2687.42		0.869
$s(DSTime)$				22115 (0.000)	
$h\big(Temp * w(DST)\big)$				11754 (0.000)	
Year	-349.0***	5.5	-63.92		

Note. Estimation results for regression in Eq. (3.4) (Model 4); ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

The red line in Figure 6 corresponds to the non-linear temperature dependent demand function which results from fitting Equation (3.4) (Model 4) that uses a spline function $h(Temp_t * w(DST_t))$. Similarly to the curve resulting from Model 3 which uses unweighted temperature function $h(Temp_t)$, this function is a rounded 'U' shaped function, with an optimal minimum demand 'comfort' temperature corresponding to *slightly below* $20°C$. We notice that for both, weighted and unweighted functions, fitted demand takes nearly identical values for larger temperatures, and the curves deviate from each other for small temperatures. We observe that in the case of a time weighted temperature, the demand increases with decreasing temperatures at a faster rate compared to when using an unweighted function.

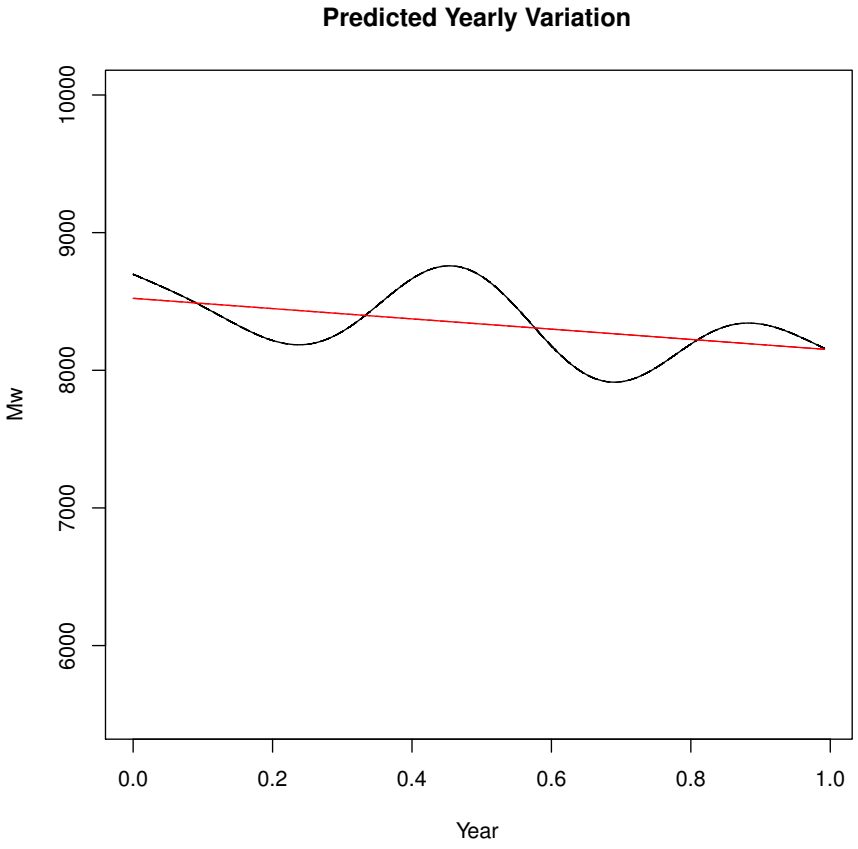
Finally, we assume that changes in demand as a function of the $Year$ variable are non-linear. This is achieved, as suggested above, by fitting a non-periodic spline function. We fit Model 5 (Equation (3.5)), which uses the spline term $k(Year_t)$. The results are reported in Table 7. We observe that the addition of the non-linear term $k(Year_t)$ produces a $R^2_{adj} = 0.898$, which is significantly higher than the one obtained for Model 4 ($R^2_{adj} = 0.869$). Figure 7 shows the fitted functional form of the term $k(Year_t)$ (black line). This was fitted with DST set to noon ($DST = 0.5$) and temperature set to the optimal 'comfort' temperature (no temp signal; $Temp = 20°C$). The red line is a linear approximation of $k(Year_t)$ and shows a decline in demand of -373.4 Mw/h across the year. This is consistent with the linear fitted decline in Model 4. The independent variable $Year_t$ is scaled $[0, 1)$, where 0 corresponds to the first data record of electricity demand for the 5 minute *standard* time period February 3, 2014 00:00-00:05 (DST February 4, 2014 00:55-01:00) and 0.999990487 corresponds to the final electricity demand record for the *standard* time period February 2, 2015 23:55-24:00 (DST February 3, 2015 00:55-01:00).

Table 7: Model 5 using nonlinear yearly variation in demand.

	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
Intercept	8156.8***	1.36	6000		0.898
$s(DSTime)$				34731 (0.000)	
$h\big(Temp * w(DST)\big)$				8971 (0.000)	
$k(Year)$				4190 (0.000)	

Note. Estimation results for regression in Eq. (3.5) (Model 5); ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

Figure 7: Fitted spline function of yearly changes in demand.



Note. The fitted functional form of the $k(Year_t)$ term (black line) in Model 5 (Equation (3.5)), fitted using DST set to noon ($DST = 0.5$) and temperature set to the optimal 'comfort' temperature (no temp signal; $Temp = 20°C$). The red line is a linear approximation of $k(Year_t)$.

5 Seasonal Demand Models

The yearly electricity demand models in the previous section are essentially a three-term models where each of the three terms on the right hand side of the regression equation (Models 1 through

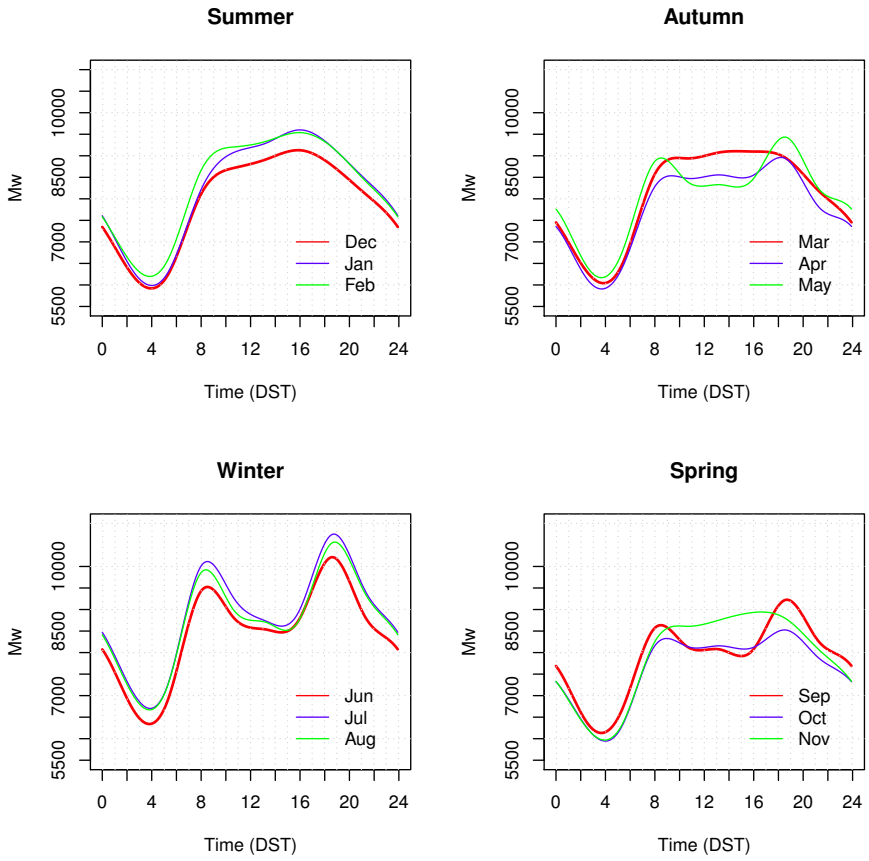
5) could be described as follows. The first term, $s(DST_t)$, is a fitted cyclic/periodic spline function that models demand as a function of a DST (activity). The second term $h(\cdot)$ models demand innovation due to temperature. We notice that the function $h(\cdot)$ in Models 1 and 2 can be thought of as a linear function (terms $\beta_1|Temp_t - 20.0|$ and $\beta_1(w(DST_t) * |Temp_t - 20.0|)$ on the right hand side of Equations (3.1) and (3.2)). The function $w(DST_t)$ can be thought of as a constant in Model 3 (Equation (3.3)) since there is no weighing of the temperature demand signal. Finally, the third term $k(Year_t)$ models long term demand seasonality and drift, and can be thought of as a linear function ($\beta_2 Year_t$) in Models 1, 2, 3 and 4.

The key assumption of the (yearly) Models 1 through 5 is that the daily (DST)/activity demand cycle is homogeneous and does not change across weekdays or seasons. This assumption of daily demand cycle homogeneity is the reason we excluded weekends and holidays from the analysis. When the cross sectional regression was examined in Section 4.3 it was shown that the term $s(DST_t)$ is not stationary across the year. The solution to the non-stationarity of $s(DST_t)$ is to fit regressions over shorter periods, thus, we fit the regression in Model 6 (Equation (3.6)) for each calendar month. This model is referred to as a seasonal demand model.

The daily smoothed (periodic splines) electricity demand curves, $s(DST_t)$, for different months of the year that result from fitting Model 6 are shown in Figure 8. It is important to notice that even if the DST (activity) demand cycle is homogeneous across months, the $s(DST_t)$ term will change each month (see Figure 8) because of the change of the daily temperature cycle between months. It is not the purpose of our paper to disentangle that part of the daily $s(DST_t)$ term that is due to intrinsic DST (activity time) and that part that is due to the daily temperature cycle. However, performing the regression each month has two implications:

- Any seasonal change in the daily intrinsic DST (activity time) related demand will be captured in the $s(DST_t)$ term, resulting in a better fit to the data.
- The seasonal daily temperature variation will also be captured in the $s(DST_t)$ term. This effect is obvious when we examine the empirical $s(DST_t)$ functions in Figure 8. The winter $s(DST_t)$ functions (in June, July August) show twin peaks of demand in the morning and evening when the temperature is significantly colder than the 'comfort' temperature ($20°C$). The daytime demand is lower as the temperature rises towards the 'comfort' temperature. Conversely, in summer (December, January, February) the peak demand reaches its maximum in the afternoon when temperatures are above the 'comfort' temperature.

Figure 8: Electricity demand curves for different months

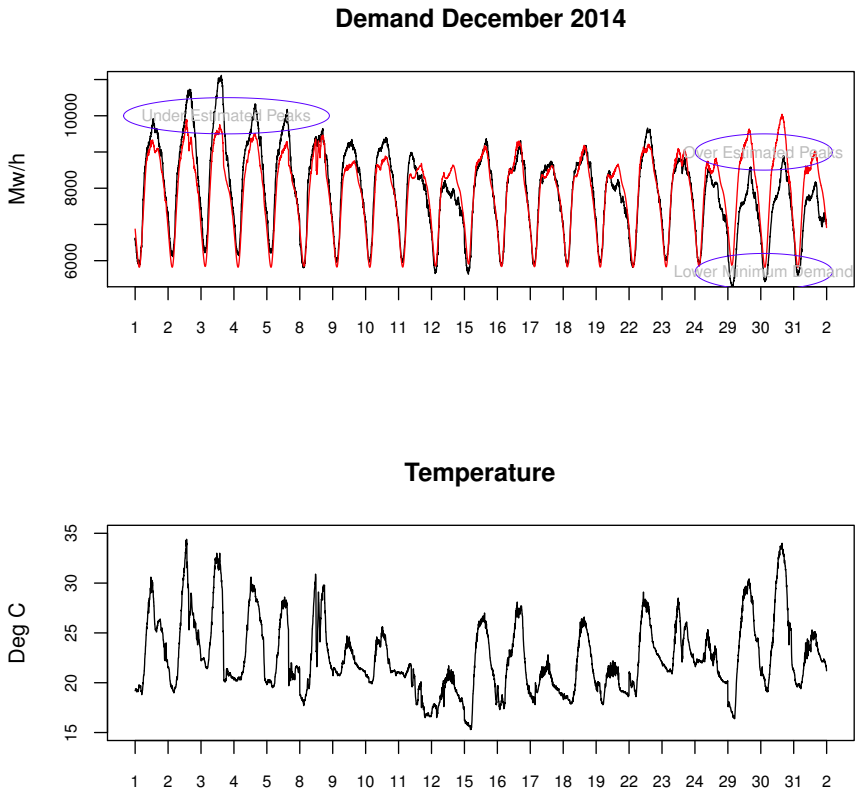


Note. Daily smoothed (periodic splines) electricity demand curves for different months of the year (indexed by DST).

Estimation results for Model 6 are summarised in Table 8. We observe that all variables are highly statistically significant, with R^2_{adj} ranging between 0.803 to 0.963. It is worth noticing that the smallest R^2_{adj} of 0.803 is observed for the month of December. If however, we exclude the last three days of the year (29, 30, 31 December), the R^2_{adj} increases to 0.906. In the following we provide the explanation for this observation, referring to Figure 9 that shows actual demand (black line) and predicted demand (red line) for December in the top panel, and associated temperatures in the bottom panel. The period between Christmas and New Year (December 29, 30, 31) is a 'defacto' holiday period with many businesses closed (even though they are officially business days). Although this period is not formally a holiday, it is characterized with very low electricity demand (see the low minimum demand marked on the graph). An implicit assumption in the regression is that the DST (activity) based demand cycle is homogenous across days and that demand innovations are driven by temperature. However, by including the 'defacto' holiday period, the underlying demand in the data is not homogeneous and the regression fitted using this low

demand data is 'biased' downwards causing temperature driven electricity peaks earlier in the month (December 2: Max 34°C; December 3: Max 33°C) to be underestimated. Conversely, the regression overestimates temperature driven electricity peaks during the low demand 'defacto' holiday period (December 29: Max 30°C; Decembers 30: Max 34°C).

Figure 9: Actual (black) and predicted (red) demand for December 2014



Note. Top panel: actual (black line) and predicted (red line) demand for December 2014; bottom panel: the associated temperatures.

6 Predicting Electricity Demand

In the previous sections we specified the input parameter temperature $Temp_t$ as the *actual* temperature, i.e., we assumed that the actual temperature is known. We refer to the resulting model as the 'Oracle', or the benchmark model as it contains all the information about the temperature variable. This assumption is however unrealistic for any predictive model of electricity consumption, as clearly the actual temperatures are unknown at the time of prediction.

Table 8: Model 6: Monthly seasonal regressions

Month	Coefficent	Estimate	Std.error	t-value	F test (p-value)	R^2_{adj}
February 2014	Intercept	7961.7***	7.3	1095.7		0.928
	$s(DST_t)$				3174 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	197.3***	3.5	56.9		
March 2014	Intercept	8018.1***	5.5	1446.0		0.963
	$s(DST_t)$				8855 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	46.2***	2.7	17.0		
April 2014	Intercept	7775.9***	6.7	1167.2		0.919
	$s(DST_t)$				5872 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	27.5***	3.2	8.5		
May 2014	Intercept	7792.0***	6.9	1134.5		0.952
	$s(DST_t)$				12085 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	109.0***	2.6	41.6		
June 2014	Intercept	8031.1***	12.6	637.3		0.939
	$s(DST_t)$				6574 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	129.3***	3.5	37.3		
July 2014	Intercept	8223.1***	11.7	700.7		0.941
	$s(DST_t)$				6683 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	147.9***	2.5	58.3		
August 2014	Intercept	7956.1***	13.1	607.5		0.932
	$s(DST_t)$				3964 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	184.2***	2.9	62.9		
September 2014	Intercept	7496.7***	8.2	914.7		0.870
	$s(DST_t)$				3179 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	139.6***	2.6	53.4		
October 2014	Intercept	7363.4***	4.8	1529.4		0.934
	$s(DST_t)$				6765 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	126.9***	1.7	76.4		
November 2014	Intercept	7509.0***	6.3	1192.6		0.900
	$s(DST_t)$				2848 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	233.0***	2.4	95.2		
December 2014	Intercept	7615.2***	10.8	703.7		0.803
	$s(DST_t)$				1108 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	139.3***	4.1	34.1		
December 2014	Intercept	7585.7***	8.0	948.0		0.906
December 29,30,31	$s(DST_t)$				1839 (0.000)	
excluded	$w(DST_t) * \lvert Temp_t - 20\rvert$	231.0***	3.3	70.3		
January 2015	Intercept	7498.6***	8.3	898.1		0.930
	$s(DST_t)$				1862 (0.000)	
	$w(DST_t) * \lvert Temp_t - 20\rvert$	279.8***	2.9	97.9		

Note. Estimation results for monthly seasonal regression model in Eq. (3.6) (Model 6); ***, ** and * indicate significance at 0.001, 0.01 and 0.05 significance level, respectively.

At the other extreme we can use calculated temperatures based on long-term seasonal temperature variations and a physics based model of intra-day temperatures. This is a 'zero knowledge' model and we refer to this model as the 'Zero' model. The model of intra-day temperatures of the 'zero' model is based on the physics of daytime solar radiation (sinusoidual) and night-time cooling (exponential decline) using a model developed by Göttsche et al. (2001). Full details on the intra-day physics based model of intra-day temperature are detailed in Appendix A. However, the 'zero' model is also unrealistic since any forecaster of next day electricity demand will have access to meteorological temperature forecasts.

The most realistic electricity forecasting model uses next day forecast maximum and minimum temperatures and then applies the Göttsche et al. (2001) physics model to interpolate intra-day temperatures for the next day. We refer to this model as the 'Forecast' model. Specifically, we use the publicly available 16:20 DST ('the 6 o'clock news forecast') Australian Bureau of Meteorology forecast of next day maximum and minimum temperatures at the Sydney suburb of Paramatta.[12] We note that a commercial electricity demand forecaster would use more accurate and frequent subscription (non-public and expensive) forecasts available to Aviation and other commercial users.

The empirical results obtained from forecasting electricity demand using the 'Oracle' model (complete temperature information), the 'Zero' model (no temperature information) and the 'Forecast' model (one day ahead forecast for the maximum and minimum temperatures) are presented in Table 9. We analyse the data using the seasonal Model 6 (Equation (3.6), refer also to Section 5), which is fitted independently to monthly data.[13] The model does not have a long term $Year_t$ component. For comparison we also analyse the best performing model for yearly data, Model 5 (Equation (3.5), refer also to Section 4).

The Australian Energy Market Operator (AEMO) also publishes next day electricity forecasts and these are tabulated below for comparison with the models we have developed above.[14] The AEMO forecasts used are published at 12:00pm DST the previous day for 48 half hour periods beginning

[12]Paramatta is the closest suburb (approx 10 kilometres) to Homebush where Australian Bureau of Meteorology forecasts are available. The median difference between the forecast minimum at Paramatta and the actual minimum at Homebush was $-0.2^{\circ}C$ (std dev. $1.4^{\circ}C$) and the difference between the forecast maximum at Paramatta and the actual maximum at Homebush was $-0.1^{\circ}C$ (std dev. $1.4^{\circ}C$).

[13]Days December 29, 30 and 31 have been removed from the December 2014 data, refer to the discussion in Section 5.

[14]These forecasts can be found on the AEMO web server at the URLs: `http://www.nemweb.com.au/REPORTS/CURRENT/Short_Term_PASA_Reports/` and `http://www.nemweb.com.au/REPORTS/ARCHIVE/Short_Term_PASA_Reports/` for current and archived reports, respectively.

at 04:00am the following day. The date of the AEMO demand forecasts was all business days from July 1, 2014 to June 30, 2015.[15] Thus, our forecasts and comparisons to AEMOS's forecasts are based on the sample from July 1, 2014 to February 2, 2015. The AEMO forecast data can be regarded as the best publicly available commercial electricity demand forecasts. However, the AEMO forecasting model and methodology is not in the public domain. Demand forecasts are used by generators and distributors to price the delivery of power into the National Electricity Market (NEM) and it is likely that there are additional non-public prediction models, particularly for price sensitive demand peaks, developed by generators and distributors to assist with pricing electricity supply in the NEM.

Table 9: Electricity demand prediction statistics, $n = 71136$

Model 6 (Seasonal)		
	Median Error	Std Dev. Error
'Oracle'	-0.06%	3.52%
'Forecast'	0.09%	4.20%
'Zero'	-0.37%	4.73%
Model 5 (Yearly)		
	Median Error	Std Dev. Error
'Oracle'	-0.12%	4.03%
'Forecast'	0.15%	4.31%
'Zero'	-0.35%	4.67%
AEMO Prediction		
	Median Error	Std Dev. Error
'Forecast'	0.06%	2.64%

Note. Electricity demand prediction statistics for the 'Oracle' model (complete temperature information), the 'Zero' model (no temperature information) and the 'Forecast' model (one day ahead forecast for the maximum and minimum temperatures). The results are contrasted against AEMO's forecasting model.

As expected, using 'Oracle' temperature data results in the best prediction statistics, leading to the smallest median forecasting error and the smallest standard deviation of the forecasting error. The 'Forecast' (realistic) temperature data leads to the intermediate performance, and the 'Zero' (no knowledge) temperature data has the least predictive power (refer to Table 9). We note that the yearly Model 5 (Equation (3.5)) has good predictive power and is very similar in performance to Model 6 (Equation (3.6)). We also note that the standard deviation of the 'Forecast' prediction error (4.31%) is approximately 1.65% higher than the comparable AEMO's standard deviation of the forecast error (2.64%). Given the limitation of one representative temperature for the entire

[15]Unfortunately, we were unable to obtain AEMO forecasting data to exactly match our February 3, 2014 to February 2, 2015 data period.

NSW demand area of our seasonal and yearly models, this is an excellent result. We discuss the direction of further forecast model development in Conclusion (Section 7).

7 Conclusion

This paper introduces a parsimonious Generalised Additive Model (GAM) to relate electricity demand in Australia to the time of the day, temperature and time of the year. We notice that using Daylight Savings Time (DST) as the independent variable for the time indexed daily periodic demand consumption function provides a small but highly significant improvement of fit compared to using standard (astronomical) time. We establish a minimum demand 'comfort' temperature of $20°C$. However, the difference between $20°C$ and $19°C$ or $21°C$ is very small and this implies that temperature dependent demand is a non-linear function of the difference between the minimum demand temperature, with small variations away from the comfort temperature producing very small changes in demand. This "U" shaped temperature sensitivity curve is shown in Figure 6.

The major novel result of this paper is that the temperature demand signal is time weighted. This relates the magnitude of the temperature demand signal to the daily activity cycle based on DST. This result is intuitive: A cold morning at 04:00am generates a much lower increase in demand compared to a cold morning with the same temperature at 09:00am due to the difference in personal and economic activity between the two times. Regression models using time weighted temperature demand outperform models that were not time weighted (see Section 4). The result is also empirically modelled using cross-sectional regressions of the change in demand as a function of the temperature for all 5 minute periods across the day. This empirical time weighted temperature demand signal relationship is shown in Figure 4.

The overall accuracy of our parsimonious GAM model is evaluated against the commercial demand forecasting model used by the Australian Energy Market Operator (AEMO). The specification of the AEMO forecasting model is commercial property and is not public. However, the AEMO intra-day demand forecasts are published online. The standard deviation of the AEMO forecast error is 2.6%. The comparable result for our parsimonious GAM model is a standard deviation of the forecasting error corresponding to 4.3% (see Section 6). This result is excellent given the single temperature limitation of our model (discussed below and in Section 2). The parsimonious GAM model is a solid foundation for the development of more elaborate and accurate models for forecasting high frequency electricity demand.

Further development of the parsimonious GAM model to increase forecasting accuracy will depend on improved temperature and climate data. Temperature variations of $10°C$ or more between coastal and inland suburbs of Sydney are common. Demand weighted temperature data from different suburbs would increase model accuracy. In addition, there are several substantial cities several hundred kilometres from Sydney. A more accurate model would also include demand weighted temperatures from these cities. Finally, a commercial electricity demand forecaster would use more accurate and frequent subscription (non-public and expensive) temperature forecasts available to Aviation and other commercial users.

References

Y. Akil and H. Miyauchi. Elasticity coefficient of climatic conditions for electricity consumption analysis. *2010 International Conference on Power System Technology (POWERCON)*, pages 1–6, 2010.

J. Al-Zayer and A. Al-Ibrahim. Modelling the impact of the temperature on electricity consumption in eastern province of saudi arabia. *Journal of Forecasting*, 15:97–106, 1996.

P. Alaton, B. Djehiche, and D. Stillberger. On modelling and pricing weather derivatives. *Applied Mathematical Finance*, 9(1):1–20, 2002.

A. Amato, M. Ruth, P. Kirshen, and J. Horwitz. Regional energy demand response to climate change: methodology and application to the commonwealth of massachusetts. *Climate Change*, 71:175–201, 2005.

M. Asadoorian, R. Eckaus, and C. Schlosser. Modeling climate feedbacks to electricity demand: the case of china. *Energy Economics*, 30:1577–1602, 2008.

Australian Energy Market Operator (AEMO). NEM electricity demand continues downward trend. *Media Release*, 2014. http://www.aemo.com.au/News-and-Events/News/2014-Media-Releases/NEM-Electricity-Demand-Continues-Downward-Trend.

M. Bašta and K. Helman. Scale-specific importance of weather variables for explanation of variations of electricity consumption: The case of Prague, Czech Republic. *Energy Economics*, 40: 503–514, 2013.

M. Beccali, M. Cellura, V. Lo Brano, and A. Marvuglia. Short-term prediction of household electricity consumption: Assessing weather sensitivity in a mediterranean area. *Renewable and Sustainable Energy Reviews*, 12:2040–2065, 2008.

M. Bessec and J. Fouquau. The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. *Energy Economics*, 30:2705–2721, 2008.

L. Blázquez, N. Boogen, and M. Filippini. Residential electricity demand in Spain: New empirical evidence using aggregate data. *Energy Economics*, 36:648–657, 2013.

T. Dergiades and L. Tsoulfidis. Estimating residential demand for electricity in the United States, 1965-2006. *Energy Economics*, 30(5):2722–2730, 2008.

J. A. Duffie and W. A. Beckman. *Solar engineering of thermal processes*. Fourth Edition, New York,USA: Wiley, 2013.

C. Giannakopoulos and B. Psiloglou. Trends in energy load demand for Athens, Greece: Weather and non-weather related factors. *Climate Research*, 31:97–108, 2006.

F. M. Göttsche, F. S. Olesen, and Schädlich. Influence of land surface parameters and atmosphere on meteosat brightness temperatures and generation of land surface temperature maps by temporally and spatially interpolating atmospheric correction. *Remote Sensing of Environment*, 75:39–46, 2001.

T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 1990.

M. Hekkenberg, R. Benders, H. Moll, and A. Schoot Uiterkamp. Indications for a changing electricity demand pattern: the temperature dependence of electricity demand in the netherlands. *Energy Policy*, 37(4):1542–1551, 2009.

G. Hondroyiannis. Estimating residential demand for electricity in Greece. *Energy Economics*, 26(3):319–334, 2004.

T. Hong, W. Chang, and H. Lin. A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data. *Applied Energy*, 111:333–3350, 2013.

J. Lam, H. Tang, and D. Li. Seasonal variations in residential and commercial sector electricity consumption in Hong Kong. *Energy*, 33:513–523, 2008.

J. Lam, K. Wan, and K. Cheung. An analysis of climatic influences on chiller plant electricity consumption. *Applied Energy*, 86:933–940, 2009.

M. Manera and A. Marzullo. Modelling the load curve of aggregate electricity consumption using principal components. *Environmental Modelling and Software*, 20:1389–1400, 2005.

N. Miller, K. Hayhoe, J. Jin, and M. Auffhammer. Climate, extreme heat, and electricity demand in California. *Journal of Applied Meteorology and Climatology*, 47:1834–1844, 2008.

S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31:208–227, 2006.

P. Molnár. Evaluating the impact of temperature on electricity consumption: Daylight matters. *Working paper*, 2011.

J. Moral-Carcedo and J. Pérez-García. Temperature effects on firms' electricity demand: An analysis of sectorial differences in Spain. *Applied Energy*, 142:407–425, 2015.

J. Moral-Carcedo and J. Vicéns-Otero. Modelling the non-linear response of Spanish electricity demand to temperature variations. *Energy Economics*, 27(3):477–494, 2005.

A. Pardo, V. Meneu, and E. Valor. Temperature and seasonality influences on Spanish electricity load. *Energy Economics*, 24:55–70, 2002.

K. Pilli-Sihvola, P. Aatola, M. Ollikainen, and H. Tuomenvirta. Climate change and electricity consumption - Witnessing increasing or decreasing use and costs? *Energy Policy*, 38:2409–2419, 2010.

B. Psiloglou, C. Giannakopoulos, S. Majithia, and M. Petrakis. Factors affecting electricity demand in Athens, Greece and London, UK: A comparative assessment. *Energy*, 34(1):1855–1863, 2009.

J. Rhodes, W. Cole, C. Upshaw, T. Edgar, and M. Webbe. Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135:461–471, 2014.

D. Sailor. Relating residential and commercial sector electricity loads to climate - Evaluating state level sensitivities and vulnerabilities. *Energy*, 26:645–657, 2001.

D. Sailor and J. Munoz. Sensitivity of electricity and natural gas consumption to climate in the USA - Methodology and results for eight states. *Energy*, 22:987–998, 1997.

C. Sandels, J. Widén, and L. Nordsträm. Forecasting household consumer electricity load profiles with a combined physical and behavioral approach. *Applied Energy*, 131:267–278, 2014.

C. Tung, T. Tseng, A. Huang, T. Liu, and M. Hu. Impact of climate change on Taiwanese power market determined using linear complementarity model. *Applied Energy*, 102:432–439, 2013.

E. Valor, V. Meneu, and V. Caselles. Daily air temperature and electricity load in Spain. *Journal of Applied Meteorology*, 408:1413–1421, 2001.

K. Wangpattarapong, S. Maneewan, N. Ketjoy, and W. Rakwichian. The impacts of climatic and economic factors on residential electricity consumption of Bangkok metropolis. *Energy and Buildings*, 40:1419–1425, 2008.

A. Włodarczyk and M. Zawada. Modeling the impact of the weather factors on the electrical energy consumption in one of the regions in the southern Poland. *Seventh International Conference on European Energy Market, 1-6*, 2010.

S. Wood. *Generalized Additive Models: An Introduction with R.* Chapman & Hall/CRC Monographs on Statistics and Applied Probability, 2006.

N. Xiao, J. Zarnikau, and P. Damien. Testing functional forms in energy modeling: An application of the Bayesian approach to the US electricity demand. *Energy Economics*, 29:158–166, 2007.

T. Zachariadis and N. Pashourtidou. An empirical analysis of electricity consumption in Cyprus. *Energy Economics*, 29(2):183–198, 2007.

Y. Zhou, L. Clarke, J. Eom, P. Kyle, P. Patel, S. Kim, J. Dirks, E. Jensen, Y. Liu, J. Rice, L. Schmidt, and T. Seiple. Modeling the effect of climate change on US state-level buildings energy demands in an integrated assessment framework. *Applied Energy*, 113:1077–1088, 2014.

A Physics-Based Intra-Day Temperature Model

This model of expected intra-day temperatures uses a physics-based expected diurnal temperature variation.[16] Intra-day daytime temperatures are dependent on incoming solar radiation (sinusoidual) and nighttime cooling is dependent on heat radiation (exponential decline).

A.1 Solar Time

Solar time is based on the apparent angular motion of the sun across the sky with solar noon being the time when the sun crosses the meridian of the observer. The time of the solar noon T_{solar} is calculated by adjusting the longitude on which the local standard time (T_{std}) is based using the standard meridian for the local time zone (L_{st}) and the actual (location) longitude L_{loc} plus the time adjustment factor $T_e(D)$. Thus, T_{solar} is determined by the so called *equation of time*, which adjusts for eccentricities of the Earths orbit throughout the year:

$$T_{solar} = T_{std} + 4\big(L_{st} - L_{loc}\big) + T_e(D). \tag{A.1}$$

For Sydney, the local standard time T_{std} corresponds to GMT + 10 hours; the standard meridian for the local time zone $L_{st} = 150$ degrees and the longitude of Sydney's location $L_{loc} = 151.2094$ degrees. $T_e(D)$ representing the approximate equation of time adjustment (in minutes) is given by (see Duffie and Beckman (2013)):

$$\begin{aligned} T_e(D) = 229.2\,\big(&0.000075 \\ &+ 0.001868\cos(B) - 0.032077\sin(B) \\ &- 0.014615\cos(2B) - 0.04089\sin(2B)\big). \end{aligned} \tag{A.2}$$

Here, D represents the day of the year such that $D = 1$ for the January 1 and $D = 365$ for the December 31; and B (degrees) is given by

$$B = (D-1)\,\frac{360}{365}, \qquad 1 \le D \le 365. \tag{A.3}$$

The adjustment for solar time in terms of standard time T_{std} (in minutes) given in Equation (A.1) can then be calculated as

$$T_{solar} = T_{std} + 4\big(150 - 151.2094\big) + T_e(D). \tag{A.4}$$

[16] Diurnal temperature variation is a meteorological term related to the difference between the daily maximum and minimum temperature.

Thus, Sydney time adjusted for the actual longitude is GMT + (10 hours, 4 Minutes, 50 seconds), if not taking the equation of time into consideration. If we also adjust for the equation of time on the 15th October 2014; $T_e(D) = 14.41$ minutes. So that the solar noon on the 15th October is at 11am, 40 minutes and 45 seconds standard time.

The length of the solar day is obviously dependent on the day of the year, D. This is calculated using the result in Duffie and Beckman (2013) and assuming a 365 day year:

$$\omega(D) \;=\; \frac{2}{15}\,\arccos(-\tan(\alpha)\tan(\beta(D)), \tag{A.5}$$

where α is the latitude of Sydney corresponding to $\alpha = 33.8650$ (degrees South).

The day of year change in inclination of the sun at the equator $\beta(D)$ is given by an approximation in Duffie and Beckman (2013) (adjusted for the Southern hemisphere):

$$\beta(D) \;=\; -23.45\,\sin\left(\frac{360}{365}\,(284+D)\right). \tag{A.6}$$

A.2 Modelling Diurnal (Intra-day) Temperature Variation

Modelling intra-day temperature variation across the chosen sample period is relatively complex due to the seasonal variation in daylight hours and inclination of the sun. We choose to use a slightly simplified form of the theoretical model of diurnal temperature derived by Göttsche et al. (2001), which is based on the physics of daytime solar radiation (sinusoidual) and night-time cooling (exponential decline). The intra-day (diurnal) temperature $T_{day}(t)$ between the time of the minimum daily temperature t_{min} (before sunrise) and the time of sunset t_{sunset} is given by:

$$T_{day}(t) \;=\; T_{min} + \big(T_{max} - T_{min}\big)\,\sin\left(\frac{\pi}{2}\Big(\frac{t-t_{min}}{t_{max}-t_{min}}\Big)\right), \quad t_{min} \le t \le t_{sunset}. \tag{A.7}$$

This is a sinusoid, which is theoretically modelled by Göttsche et al. (2001) using the physics of daytime solar radiation absorption by the surface of the Earth. We also define

$$T_{sunset} \;= T_{day}(t_{sunset}). \tag{A.8}$$

The intra-day (diurnal) temperature $T_{night}(t)$ between the time of sunset (cessation of solar radiation) t_{sunset} and the time of the minimum temperature $t_{min(nextday)}$ on the next day is given by

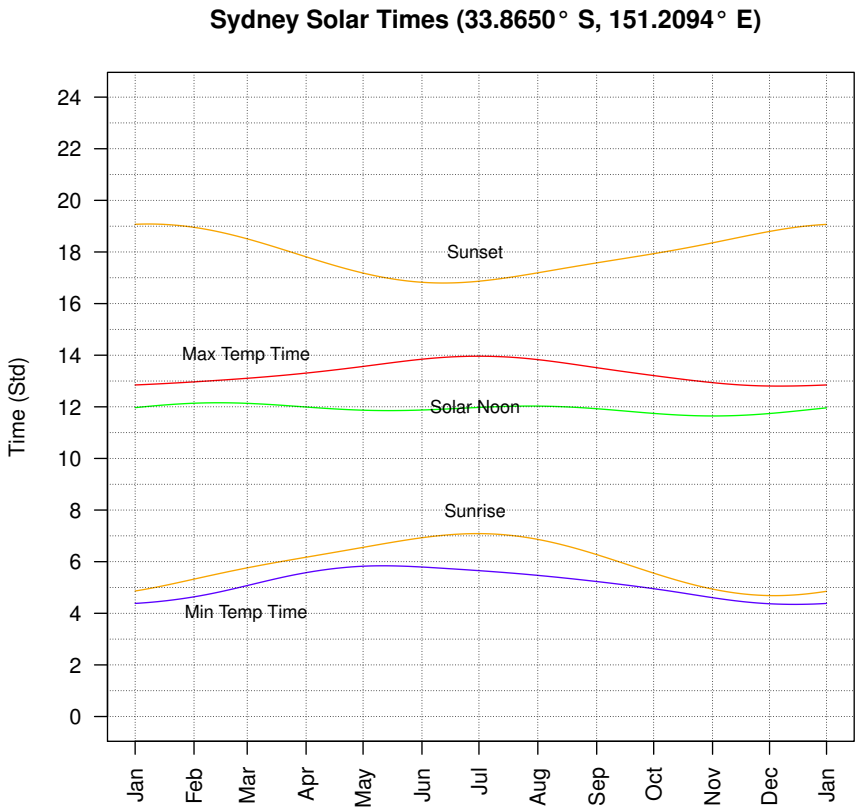
$$T_{night}(t) \;= T_{sunset}\;\exp\left(\log(T_{min(nextday)}/T_{sunset})\,\frac{t-t_{sunset}}{t_{min(nextday)}-t_{sunset}}\right), \quad t_{sunset} < t < t_{min(nextday)}. \tag{A.9}$$

This is an exponential, which is also theoretically modelled by Göttsche et al. (2001) using the physics of the night daytime solar radiation from the surface of the Earth into space.[17]

Figure 10 shows the modelled time of the daily temperature maximum (t_{max}) and minimum (t_{min}) in Sydney throughout the year. The expected time of the minimum daily temperature (blue line in Figure 10) occurs before sunrise and begins rising earlier before dawn in winter. The expected time of the maximum temperature (red line in Figure 10) occurs after the solar noon and is later after the solar noon in winter.

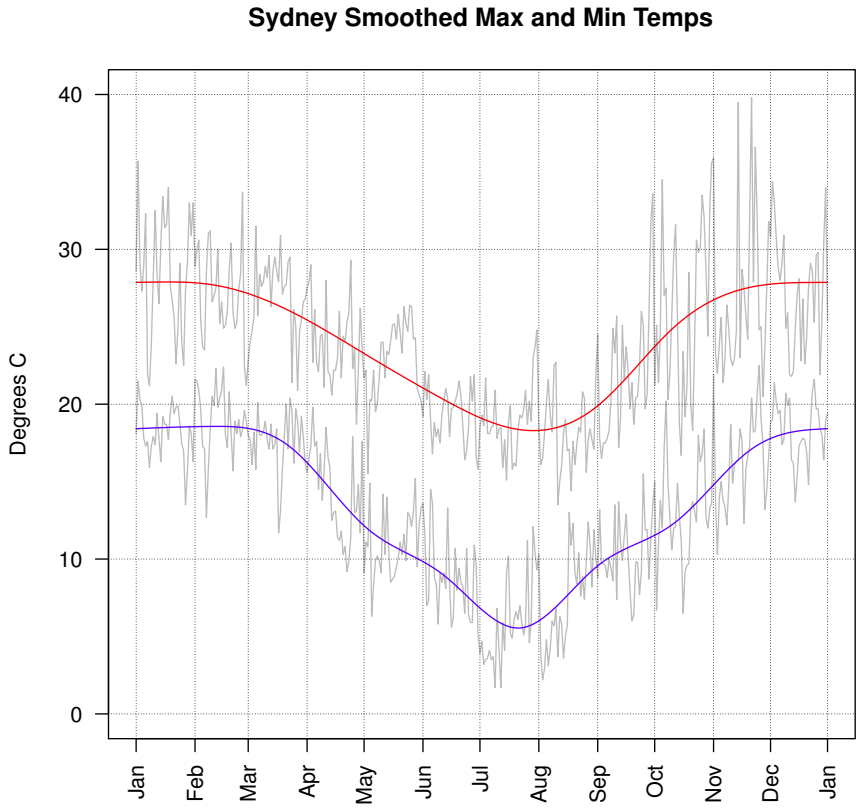
[17] The equations above are slightly simplified from the Göttsche et al. (2001)'s equations and do not have equal first derivatives at sunset (t_{sunset}) and the minimum temperature time (t_{min}).

Figure 10: Sydney sunset (t_{sunset}), solar noon and sunrise in standard time



Note. Sydney sunset (t_{sunset}), solar noon and sunrise in standard time, shown together with the expected time (red line) of the daily maximum temperature (t_{max}) fitted using a smoothed period spline model and the fitted expected time (blue line) of the daily minimum (t_{min}) temperature.

Figure 11: Actual maximum and minimum temperatures and its approximations



Note. The grey lines show actual maximum and minimum temperatures between January 1, 2014 to December 31, 2014. The red and blue lines are the same data smoothed periodic splines, which are approximations to the expected annual daily maximum and minimum temperatures.