ARTICLE IN PRESS

A cooling change-point model of community-aggregate electrical load

Muhammad Tauha Ali*, Marwan Mokhtar, Matteo Chiesa, Peter Armstrong

Mechanical Engineering Program, Masdar Institute of Science and Technology, PO Box 54224, Abu Dhabi, United Arab Emirates

ARTICLE INFO

ABSTRACT

Estimates of daily electrical cooling load for a city of 800,000 are developed based on the relationship between weather variables and daily-average electricity consumption over 1 year. The relationship is found to be nearly linear above a threshold temperature. Temperature and humidity were found to be the largest, at 59%, and second largest, at 21%, contributors to electrical cooling load. Direct normal irradiation intercepted by a vertical cylinder, $\text{DNI}\sin\theta$, was found to be a useful explanatory variable when modeling aggregates of buildings without a known or dominant orientation. The best study case model used $\text{DNI}\sin\theta$ and diffuse horizontal irradiation (DHI) as distinct explanatory variables with annual electrical cooling load contributions of 9% and 11% respectively. Although the seasonal variation in electrical cooling load is large – on peak summer days more than 1.5 times the winter base load – the combined direct and diffuse solar contribution is essentially flat through the year, a condition at odds with the common assumption that solar cooling always provides a good match between supply and demand. The final model gives an electrical cooling load estimate for Abu Dhabi Island that corresponds to 40% of the total annual electrical load and 61% on the peak day.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Models that predict energy use in the built environment have evolved along two largely independent paths—one aimed at utility (supply-side) forecasting applications and the other oriented toward building (demand-side) performance applications. Utility forecasting models typically aggregate demand over a large and diverse set of loads, which complicates the use of physical or engineering based models. Autoregressive models and ARMA models with just one weather variable, temperature, are prevalent.[1] Building performance applications, on the other hand, often try to model response to weather and other factors in order to assign a weather-normalized building energy performance rating or so that annual energy impact of efficiency retrofits can be estimated. Perhaps the oldest supply side application, the heating degree day method has been used for over 100 years to forecast customer fuel requirements [1].

Electric utilities now use short, medium and long-term forecasts routinely for such varied applications as long-term planning (monthly time steps) and load dispatch (fraction of hour to 1 day).

Load control (a form of dispatch) and day-ahead price setting are other important potential applications.

Building level models find application in rating systems, fault detection, savings verification, retrofit and operational decision making, calculation of demand response credits, and model-based control. In fault detection, for example, deviations of model estimates from measured responses may indicate a fault and an associated change in one or more of the model parameters may indicate the type of fault. Models based on daily data are quite common although shorter (hourly) and longer (weekly and monthly) time steps have been used as well. Hourly and shorter time steps are used mainly for component-level fault detection and model-based control.

The present work is motivated by technology and policy questions germane to both supply and demand. For example, on the supply side, the potential for large (urban scale) solar powered cooling in hot, sunny climates is of interest. In order to properly size a solar district cooling plant, the aggregate cooling loads should be based on all of the significant load predictors—solar irradiation, humidity and temperature. Utilizability and optimal sizing are affected by seasonal variations of the solar resource, seasonal variations of the cooling load, and the phase relation of these variations. On the demand-side, policy questions pertain to things such as retrofit programs, energy codes, and equipment efficiency standards. Accurate cooling load models are needed to assess how new demand response technologies, such as smart meters, will impact both customer and energy supplier. These questions can be bet-

* Corresponding author. Tel.: +971 2 698 8139; fax: +971 2 698 8026.
  E-mail addresses: mali@masdar.ac.ae (M.T. Ali), parmstrong@masdar.ac.ae (P. Armstrong).

[1] Except for long-range planning, the opposite extreme, which is usually based on historical trends and demographic and economic forecasts.

ter answered when existing energy end-use intensities are fully understood.

In this paper we explore models based on the predominantly linear and non-interacting influences of four weather variables. The physical basis of such models is described. The idea of a change-point that is a linear combination of the explanatory variables is introduced and the need to model a transition *region*, rather than a discrete change-point, is addressed. The analysis is applied to the electrical system load of a Persian Gulf city, Abu Dhabi, in which cooling load and solar resource are both substantial.

Energy demand forecasting models are typically formulated as multivariate time-series models that may include independent variables representing weather conditions and day-type. For models with a one day or larger time step, the most often used weather variable is mean daily outdoor dry-bulb temperature or a variable related to outdoor temperature such as heating or cooling degree days with respect to a given base temperature [1,4–8].[2] A few electric load forecast models include humidity and day-typing [9,10]. Models that use only past loads [11,12] and models that consider day-typing and past loads without considering weather at all [13,14] can give satisfactory results for one-step ahead forecasts. In [15], the importance of solar irradiation is demonstrated. The most common model forms are linear, linear with change-point, and artificial neural network. A review of utility forecasting methods and their application by different researchers is presented in [16]. Similar techniques are used to infer equivalent thermal parameters (ETP) or energy signatures (ES) of a building when applied on the building level [17].

The ES/ETP models have been used to forecast heating and cooling loads, heating or cooling plant input power, or whole building electric load. When used to forecast heating or cooling load on time steps of under one day, the models typically include transient terms [18,19]. In [15,20–24], regression-based[3] techniques are used while in [12,25,26] fuzzy logic and neural network techniques are used. Most of the efforts have been concerned with the heating loads of buildings. In [22], a model for variable internal gains is presented. In [26] among others, parameters accounting for both external and internal gains are considered for estimating the building energy signature. In [23], a detailed review of different models for heating loads from building-level data is presented. In [27,28], regression analysis is used to identify building ETP or response function models and develop controls for optimum operation of a chiller plant. In [29,30] energy signatures are used to assess energy savings and financial feasibility of retrofit measures. A critical review of regression modeling applied to building-level loads is presented in [17]. An exhaustive review of energy models and their applications in planning, forecasting, emission reduction, etc. – from building level to utility service area – is given in [31].

The technique one chooses for developing an energy signature, ETP, or other forecasting model depends on the amount and type of data available, application, and the simplifying assumptions that can or must be made [25,32]. In [33], comparison of 14 regression models used for forecasting is presented and the need for more data with increasing model complexity is illustrated. In [12], a comparison of fuzzy logic, neural network and linear regression-

based modeling techniques is made and the data required to train a model is shown to be least for linear regression models because load–weather relationships are nearly linear and can therefore be identified with just enough data to properly represent the extreme conditions.

2. Building load data availability and characteristics

Residential and commercial building electrical loads typically exhibit strong seasonal variation arising from operation of the equipment that heats and cools the occupied spaces of these buildings. These are the so-called heating, ventilating and air-conditioning (HVAC) loads.

Non-HVAC loads, on the other hand, do not vary much with weather. The non-HVAC loads include refrigeration, lighting, clothes and dish-washers, water heating,[4] consumer and office electronics, and IT equipment. When considering the impacts of energy supply and demand-side policies, it is imperative to know how energy is used within a given population. The breakdown of electricity use varies from one region to another based on price, climate, economic activity and local customs. However, the energy consumption for heating and/or cooling of buildings is almost always significant, ranging from 10% to over 50% in most developing and developed countries [34]. Policies to reduce HVAC energy consumption may target the thermal loads, by introducing standards for the performance of building envelopes, and minimum energy conversion efficiency of HVAC equipment. Building codes may also target the energy impacts of HVAC systems and controls.

The most reliable way to assess energy end-use is to measure it directly, but end-use metering is expensive and invasive. Because the variability among buildings is large, confidence that a given sample is representative of the underlying population can be achieved only with large samples as in, for example, the End-Use Load and Consumer Assessment Program (ELCAP) [35]. If ETP or ES methods are used to infer HVAC end-uses, the uncertainty is typically much greater than for end-use metering. A similar but little-used approach involves identification of heating and cooling loads from the seasonal load variations observed in aggregates of buildings monitored at the feeder, substation, or higher level [36].

Both of these approaches are much less costly, per building, than end-use metering and they can be combined. A modest sample of building level load observations can provide information about variance, while substation or higher level load observations provide estimates for a very large aggregate sample. The combination of community-level and building level seasonal load analyses can thus address the sample size problem in a cost-effective way.

In this study we explore the community-level by looking at the daily total energy delivered to a city. Abu Dhabi Island, a city of about 800,000 [37] comprised largely of commercial and residential buildings serves as the study case. The level of building type homogeneity, although uncommon at this scale, is quite common at the feeder or substation level. The methodology illustrated here can therefore provide useful assessments of heating and cooling loads at modest cost in many communities and urban areas. A method of change-point identification for aggregates of buildings is developed and a multivariate regression model that uses all the physically meaningful explanatory variables (that can be applied at the building level as well as to large aggregates of buildings) serves to identify a community energy signature. The resulting model provides an estimate of the sensitivity of cooling load to each weather variable. To demonstrate the method, 2008 daily-average electrical

---

[2] A *degree day* is defined in terms of the difference between mean daily temperature and a temperature threshold or *degree day base* [2]. The sum of positive differences during a week, month or year is the cooling degree days for the period in question and the absolute value of the sum of negative differences is the heating degree days. Change point models typically estimate both the degree day base and ratio of heating or cooling load to temperature difference [3].

[3] In the context of building energy use investigators use the term *regression model* to mean a multi-variate linear or linear change-point model as opposed to a fuzzy or neural network model. Change-point, ETP, and energy signature models are similar in that they are all typically identified by regression.

---

[4] Water heating loads vary with water mains temperature, which typically lags (and blurs—i.e. low-pass filters) outdoor air temperature by months. For this study we neglect the seasonal variation in water heating load.
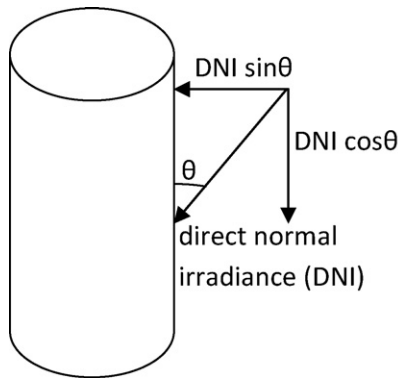
Fig. 1. Components of DNI used to model aggregate load of buildings without a known wall/roof surface ratio and without a known dominant orientation. Daily values are based on sum of hourly DNI components.

Fig. 2. RMSE versus temperature for linear-growth model with DHI and DNI $\sin\theta$ terms (model 7, data set A).

load data for the entire city were obtained from the Abu Dhabi Distribution Company (ADDC) and 2008 daily-average weather data from a coastal weather station near Abu Dhabi served as the independent explanatory variables.

3. Methodology

Our analysis procedure, in outline, follows accepted practice: (1) formulate a set of candidate models, (2) identify transition regions or points, (3) test and refine the model. For aggregates of buildings it is advantageous to model a transition region, rather than a distinct change-point and, because buildings are added during a study period that spans months or years, we must add a growth term. The use of more than two weather variables is also novel for change-point models with 1-day or longer time steps.

The candidate models considered are multivariate linear and bi-linear combinations of weather, day-type and urban growth terms. Candidate solar irradiation terms include global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), direct normal irradiance (DNI) on a vertical surface (DNI $\sin\theta$) and DNI on a horizontal surface (DNI $\cos\theta$). The components of DNI incident on a hypothetical cylindrical building are shown in Fig. 1.

Other weather variables include ambient temperature and specific humidity. To account for growth in building stock during the relatively short period of 1 year, a linear-growth model[5] is considered. The aggregate load model described above may be expressed as follows:

$$W_{\text{electric}} = (1 + C_0 t)(C_1 + C_2 D_s + C_3 D_f + C_4 T + C_5 w + C_6 \text{GHI}$$

$$+ C_7 \text{DHI} + C_8 \text{DNI}\cos\theta + C_9 \text{DNI}\sin\theta) \tag{1}$$

where W_{electric} = average daily electrical load (MW), time = day of year, D_s = Saturday indicator variable, D_f = Friday indicator variable, T = daily mean outdoor dry-bulb temperature, w = specific humidity (mass ratio of kg moisture per kg dry air), DHI = diffuse horizontal irradiation (W/m^2), DNI = direct normal irradiation (W/m^2), θ = solar zenith angle (radian), $\cos\theta$ = ratio of direct irradiation on a horizontal surface to DNI, $\sin\theta$ = ratio of direct irradiation on a vertical surface to DNI, GHI = global horizontal irradiation.

The use of daily time steps means that transient thermal response (lags) will have relatively little effect. Similar models have been used for whole-buildings and aggregates of buildings
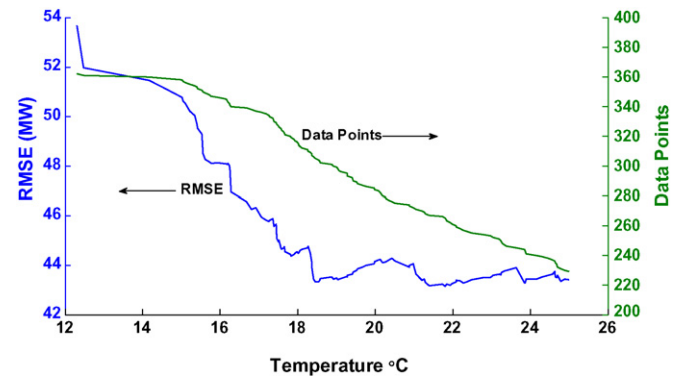
operating in conditions that require heating every day (i.e. depths of heating season) or cooling every day (depths of cooling season). Change-point models with fewer weather terms (usually temperature only) have been widely used to model energy use through transition seasons as well as heating and/or cooling season response. Although coefficient of performance (COP) of cooling equipment is in fact a function of outdoor temperature, part load fraction and indoor conditions, its sensitivity is normally not very large. The result of this constant-COP assumption is a linear load model.

Eq. (1), when applied to a single building, models the daily cooling load on days when conditions result in some amount of cooling and no heating. For aggregates of buildings we may relax the requirement by saying that some amount of cooling is required in *most* buildings equipped for cooling, provided that none are being heated. We refer to days that satisfy this condition as *cooling days* and to the locus of conditions on such days as the *linear region*. Since Eq. (1) applies only to the linear region, it is necessary to identify the transition region shown in Fig. 7, (usually modeled, for single buildings, as a change-point) which is typically observed in swing seasons. We refer to the lower boundary of the linear region as a change-point to be identified by an iterative process that may be described as successive filtering. Successive filtering is used to find change-points defined in terms of temperature only, temperature and humidity, and a linear combination of irradiation terms and temperature.

The simplest and most commonly chosen [5,6,9] threshold for the linear-to-transition region change-point is temperature. This choice is consistent with the fact that most cooling systems are thermostatically controlled. For the present case study we note that identification of the exact change-point is not critical because the number of points in the linear region is ample.

The linear-to-transition region change-point is estimated by plotting RMSE as a function of threshold value. Fig. 2 indicates that for this data set, the linear region ends at 18.5 °C. The average of the specific humidity at this temperature is 0.0085 kg of water/kg of dry air. Similar hot-climate temperature thresholds are reported in [6,9] and also suggested in [5].

To establish the boundary between the base load[6] region and the transition region we could also use a simple temperature threshold. However, it is desirable to identify this change-point as accurately as possible because it corresponds to the base load which, in turn, determines what fraction of the total annual electrical load may be attributed to cooling. Because the base load sample is small (7 days) each day in the sample has a large impact on the base load

---

[5] Incorporation of the growth term results in a model with bilinear terms. The least squares regression may be accomplished by applying a Levenburg–Marquardt or partial least squares algorithm [38].

[6] There is undoubtedly some heating on cold days but not enough to reliably estimate a heating temperature coefficient.

Table 1
Correlation matrix (pair-wise R^2) for the 2008 daily mean electrical load and weather data.

	T	w	GHI	DHI	DNI $\cos\theta$	DNI $\sin\theta$	Load
T	1.000	0.874	0.651	0.570	0.201	−0.497	0.963
W	0.874	1.000	0.452	0.452	0.094	−0.455	0.916
GHI	0.651	0.452	1.000	0.398	0.686	−0.132	0.581
DHI	0.570	0.452	0.398	1.000	−0.394	−0.901	0.508
DNI $\cos\theta$	0.201	0.094	0.686	−0.394	1.000	0.582	0.180
DNI $\sin\theta$	−0.497	−0.455	−0.132	−0.901	0.582	1.000	−0.451
Load	0.963	0.916	0.581	0.508	0.180	−0.451	1.000

estimate. We therefore define an adjusted temperature (similar to sol–air temperature) as:

$$T_{\text{adj}} = T + \frac{C_6\text{GHI} + C_7\text{DHI} + C_8\text{DNI}\cos\theta + C_9\text{DNI}\sin\theta}{C_4} \qquad (2)$$

The adjusted temperature corresponds to outdoor temperature and solar radiation parts of the cooling load model identified for the linear region. This model accounts for the fact that thermostats will call for cooling when there are strong solar gains at a lower outdoor temperature than when there are not strong solar gains.

Weekend and holiday electrical demands are presumed to exhibit behaviors different from the demand behavior of typical weekdays [5]. Therefore, day-typing is implemented to estimate this behavior [9,16]. Possible day-types considered were Fridays, Saturdays, and holidays. UAE Schools and offices are generally closed on both Fridays and Saturdays. Shops are generally open on Saturdays but closed on Fridays. Holidays spanning a period of three days consecutively which are specific to Islam (30th September–2nd October and 8th–10th December in 2008) are also treated as Fridays. By setting $C_0 = 0$, Eq. (1) reverts to a no-growth model and by setting $C_2 = C_3 = 0$ it reverts to a model with no day-typing.

Ten combinations of the four solar terms mentioned above were investigated and the best model for the linear region was chosen based on least Root Mean Squared Error (RMSE), physical plausibility and standard errors of estimated coefficients.

The resulting regression models were used to estimate the daily-average electricity demand, a major component of which is supposed, in the study case, to be cooling demand. The data set was categorized in three data sets. Data set A comprises all the data records. In order to assess the sensitivity of cooling load to weather in the linear region, cold days were excluded in data set B. While, in data set C weekends and holidays were also excluded.

In summary, the methodology as applied to a cooling-dominated community like Abu Dhabi proceeds as follows:

- Develop candidate models based on observed response to all plausible available independent variables, day-types, and time.
- Compute the correlation matrix to identify possible co-linearity problems that may impact the later selection or elimination of variables in the model.
- Select model training and validation sets (not done in this study for lack of data).
- Identify change-points in outdoor temperature (or linear combination of weather variables) by applying successive filtration. The data are sorted by daily electrical load and the regression is performed as each successive record is added. The transition region is identified when the rate of increase in RMSE becomes statistically significant.
- Test all combinations of explanatory variables. Select the best regression model based on RMSE, t-values and Durbin–Watson tests.

- In the case of base load change-point, fit successive trend lines to days with lowest cooling potential and apply Grubbs method [39] on the resulting data set with subsequent point for detection of outliers.

Having identified the change-points, the coefficients of the linear region models and a transition region interpolating function, a further analysis can be performed to estimate the share of cooling load attributed to each weather variable.

4. Application

Least squares regression is used to estimate the coefficients for the linear regions of the different data sets and models considered. The day-type constants are removed from the models of data sets B and C from which weekends were eliminated. Although the models containing two or three irradiation components have lower RMSE than models containing one or two, the former often have negative coefficients for some of the terms. Any model with a negative irradiation term is physically implausible because solar irradiation always contributes to the cooling load positively. The negative coefficients are partly due to multicollinearity present between irradiation terms as can be seen in Table 1. Temperature and specific humidity are also correlated but the coefficients are seen in Appendix A to be much more stable, across models and data sets, than the coefficients of the solar irradiation terms. Therefore, elimination of certain irradiation terms, but not humidity, was undertaken to address collinearity problems [40,41].

Note that in (1) only two of the three terms involving C_6, C_7 and C_8 are admissible in a given model because the associated random variables are related exactly by:

$$\text{DHI} = \text{GHI} - \text{DNI}\cos\theta \qquad (3)$$

Therefore, when DHI is considered, either the GHI or DNI $\cos\theta$ term can be included but not both. The solar terms for models with two and three irradiation terms are shown in Table 2 on the left and the equivalent two- and three-term groups are shown on the right. GHI appears usefully in only two of the Eq. (1) sub-models: one in which GHI is the only solar term and the other with DNI $\sin\theta$.

For the linear region (cooling days) the regression models results for all 11 linearly independent combinations of radiation terms are reported in Table 3. The best physically plausible model was found to have the following form:

$$W_{\text{electric}} = (1 + C_0t)(C_1 + C_2D_s + C_3D_f + C_4T + C_5w$$
$$+ C_7\text{DHI} + C_9\text{DNI}\sin\theta) \qquad (4)$$

Table 2
Choice of solar irradiation terms from two- and three-term equivalent groups.

N	Terms used in the regression	Equivalent groups
2	(DHI, DNI $\cos\theta$)	(GHI, DHI)
		(GHI, DNI $\cos\theta$)
3	(DNI $\sin\theta$, DHI, DNI $\cos\theta$)	(DNI $\sin\theta$, GHI, DHI)
		(DNI $\sin\theta$, GHI, DNI $\cos\theta$)

Table 3

Regression results for the linear region (Data Set B, 305 records) with manually chosen growth coefficient of 0.128/year (0.00035/day). In this table only, specific humidity is in g/kg.

Coefficients:	Model 1 NG	Model 1 LG	Model 2 NG	Model 2 LG	Model 3 NG	Model 3 LG	Model 4 NG	Model 4 LG	Model 5 NG	Model 5 LG	Model 6 NG	Model 6 LG	Model 7 NG	Model 7 LG	Model 8 NG	Model 8 LG	Model 9 NG	Model 9 LG	Model 10 NG	Model 10 LG	Model 11 NG	Model 11 LG
Growth/year	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128	—	0.128
Constant	-612	-478	-600	-537	-605	-479	-704	-501	-635	-507	-597	-537	-803	-646	-681	-548	-597	-537	-718	-490	-850	-621
Saturday	-22.4	-19.9	-22.2	-21.0	-22.2	-19.9	-20.6	-19.5	-22.8	-20.4	-22.1	-21.0	-19.2	-17.3	-20.0	-20.8	-22.1	-21.0	-18.3	-21.4	-16.0	-19.1
Friday	-60.1	-57.4	-59.9	-58.0	-60.4	-57.4	-61.2	-57.7	-60.5	-57.9	-60.3	-58.0	-61.9	-58.5	-61.0	-58.1	-60.3	-58.0	-61.2	-57.8	-62.0	-58.5
T	36.3	35.3	36.8	32.7	37.8	35.2	37.6	35.6	35.8	34.7	38.1	32.8	36.7	34.3	38.9	33.0	38.1	32.8	39.8	33.8	39.0	33.1
w	63.3	47.2	62.3	52.1	61.8	47.3	64.6	47.4	64.8	49.0	61.1	52.1	68.0	52.3	62.3	52.1	61.1	52.1	61.8	49.7	65.8	53.6
GHI	—	—	-0.07	0.320	—	—	—	—	—	—	—	—	—	—	-0.152	0.310	-0.389	0.314	—	—	—	—
DHI	—	—	—	—	-0.35	0.022	—	—	—	—	-0.39	0.314	0.516	0.747	—	—	—	—	—	—	0.672	0.663
DNIcosθ	—	—	—	—	—	—	—	—	0.138	0.173	-0.04	0.320	—	—	—	—	0.345	0.006	-0.33	0.268	0.953	0.211
DNIsinθ	—	—	—	—	—	—	0.303	0.074	—	—	—	—	0.596	0.501	0.321	0.040	—	—	0.530	-0.11	-0.39	0.310
t-statistics:																						
Growth/year	—	0.011	—	0.010	—	0.010	—	0.010	—	0.011	—	0.009	—	0.010	—	0.008	—	0.009	—	0.007	—	0.007
Constant	25.5	23.2	20.3	22.2	25.6	23.2	24.1	17.6	23.9	22.6	20.6	22.1	15.5	14.2	21.2	19.5	20.6	22.1	24.6	14.8	16.2	12.2
Saturday	2.5	2.7	2.5	2.9	2.6	2.7	2.4	2.6	2.6	2.8	2.6	2.9	2.3	2.4	2.4	2.9	2.6	2.9	2.2	2.9	1.9	2.6
Friday	7.0	7.9	7.0	8.3	7.1	7.9	7.4	8.0	7.0	8.1	7.1	8.2	7.5	8.3	7.4	8.3	7.1	8.2	7.5	8.1	7.7	8.4
T	34.9	40.1	29.0	31.0	34.3	37.1	36.4	39.5	33.6	39.0	29.4	28.4	33.3	36.3	30.5	27.3	29.4	28.4	32.3	27.3	31.3	26.9
w	17.0	14.1	15.6	15.7	16.8	14.1	18.0	13.8	17.1	14.7	15.6	15.7	17.6	14.8	16.3	15.5	15.6	15.7	17.0	14.6	17.2	15.1
GHI	—	—	0.72	3.74	—	—	—	—	—	—	—	—	—	—	1.69	3.16	3.04	2.32	—	—	—	—
DHI	—	—	—	—	3.62	0.23	—	—	—	—	3.04	2.32	2.31	3.89	—	—	—	—	—	—	3.01	3.42
DNIcosθ	—	—	—	—	—	—	—	—	1.99	2.99	0.48	3.68	—	—	—	—	3.57	0.06	3.13	2.09	5.68	1.64
DNIsinθ	—	—	—	—	—	—	5.12	1.28	—	—	—	—	4.26	4.14	5.36	0.63	—	—	5.70	0.86	3.69	1.74
Summary																						
RMSE:	53.43	44.89	53.47	43.51	52.38	44.96	51.31	44.80	53.17	44.23	52.45	43.59	50.94	43.66	51.15	43.53	52.45	43.59	50.57	44.16	49.90	43.29
R-Square:	0.961	0.969	0.961	0.966	0.963	0.968	0.964	0.970	0.962	0.969	0.963	0.966	0.965	0.971	0.965	0.967	0.963	0.966	0.965	0.967	0.966	0.968
Adj. R-Square:	0.961	0.968	0.960	0.965	0.962	0.968	0.964	0.970	0.961	0.969	0.962	0.965	0.964	0.970	0.964	0.966	0.962	0.965	0.965	0.966	0.966	0.968
Durbin-Watson:	0.649	0.951	0.668	0.960	0.680	0.951	0.681	0.961	0.646	0.970	0.695	0.961	0.681	1.016	0.715	0.971	0.695	0.961	0.726	0.958	0.726	0.998

Random variables are listed in the left-most column and each model formulation is represented by a pair of columns headed by its numeric designation. Each model has two variants, a no-growth (NG) formulation and a linear-growth (LG) formulation, thus creating the pair of columns for each model number. The coefficients are listed in the top section of the table and the t statistics are reported in the middle section. Four rows at the bottom report the overall performance of each model in terms of RMSE (MW), correlation coefficient R^2, adjusted R^2, and Durbin–Watson statistic. Green cells identify the best models based on RMSE. Because RMSE variation among the candidate models is small, identification of the best model is based on R^2 and Durbin–Watson statistic. Red cells identify shortcomings of the others relative to the best model. Corresponding results for Data Sets A and C can be found on http://web.mit.edu/parmstr/Public/EnergyBuildingsAppendix/.
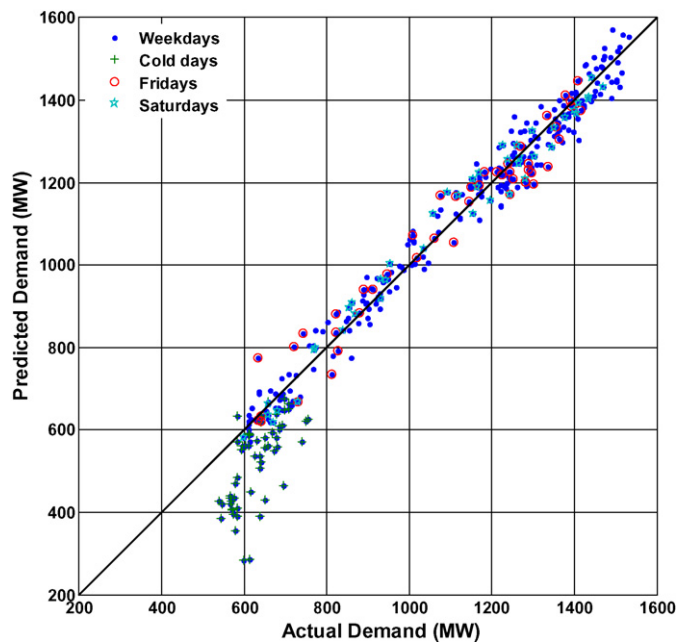
Fig. 3. Predicted electrical load from linear-growth model 7 versus ADWEC data (data set A).

From Fig. 3 and Table 4b, it can be seen that the linear growth model fits the data better than its no-growth counterpart. However, from Table 4a, we observe that the t-value, 0.01, is very low. The growth coefficient, C_0, identified by the model is therefore not reliable. For cooling dominated climates with practically no heating and a data set that begins and ends in the cold season, a manual estimate of C_0 is possible based on matching the base loads at the beginning and end of the year. By this means, a physically reasonable growth constant was found without significant (<1%) increase in RMSE. The manually assigned 0.00035 growth/day corresponds to a 12.8% growth over the year while the regressed value corresponds to 16.2%/year, a large number even by UAE standards. Building stock time-series data could be employed in cases where the base load-matching adjustment used for this Abu Dhabi data set is not an option.

To check for an ill-conditioned data set, the regression analysis was repeated after normalizing the data; the same residuals and term-by-term contributions to electrical load were obtained.

Fig. 4 shows that, except for cold days identified by '+', the residuals are generally without structure. The cold-day residuals are positive. In the "linear region" of $T > 18.5\,°C$ a small systematic variation of residuals with temperature is visible in Fig. 4(a). In Fig. 5 we see some temporal structure in the residuals. This can be quantitatively expressed by the Durbin–Watson test [42]. The calculated value is below the threshold of 1.87 for a 95% confidence interval. This can be due to growth model inadequacy and might possibly be rectified by developing the growth model using building stock data. The Coefficient of Variation (RMSE/W_{avg}) obtained for the best model is better than or similar to the values obtained by other researchers using change-point models on building level data [28,43,44]. Therefore, accuracy in estimation of cooling load from electrical load using the above model is good. However, the savings achieved by a policy decision may be difficult to estimate at the community or regional level if market penetration is slow. On the other hand, a policy with small impact per building but high market penetration might be easier to detect on the community-level than at the individual building level.

From Fig. 6, one can see that the distribution of data set B residuals is nearly normal except in the positive (measured > modeled) tail. This may stem from eliminating (treating as non-cooling days) some days that really belong in set B or from the fact that the tail is artificially truncated at the change-point. These would be days that are in the linear region as far as cooling load is concerned but appear not to be because they have an unusually high non-cooling load—i.e. large deviation from mean base load. It suggests, as does the thermodynamic basis of vapor-compression cooling [45] and as does the slight structure observed in Fig. 3, that the assumption of linear variation of load with temperature is not completely justified. With a stationary data set (loads for parts of the community that are relatively free of demolition and construction) it is possible that a more realistic model with physically meaningful non-linear terms could be identified.

The linear model with temperature, humidity and solar irradiation terms works very well in the region above the $18.5\,°C$ threshold. Change-point model practitioners usually use a temperature threshold to distinguish building level cooling days from non-cooling days [43,46]. We postulate that for aggregates of buildings there may not be a distinct change-point in the cooling load behavior but rather, as illustrated in Fig. 7, a transition region. Below this transition region we find a base load that can be estimated from the days with the lowest cooling potential where cooling potential, W_p, is defined as:

$$W_p = C_4 T + C_5 w + C_7 \mathrm{DHI} + C_9 \mathrm{DNI} \sin\theta \qquad (5)$$

Table 4a
Statistics for best fit no-growth and linear-growth models of the linear region (data set B).

	RMSE	R^2	CV-RMSE	Adjusted R^2	D–W test [52]
No-growth	50.943	0.965	4.53%	0.964	0.681
Linear-growth	43.333	0.975	3.86%	0.975	1.004
Specified linear-growth	43.656	0.971	3.89%	0.970	1.016

Table 4b
Statistics for best fit no-growth and linear-growth models of the linear region (Model 7, data set B).

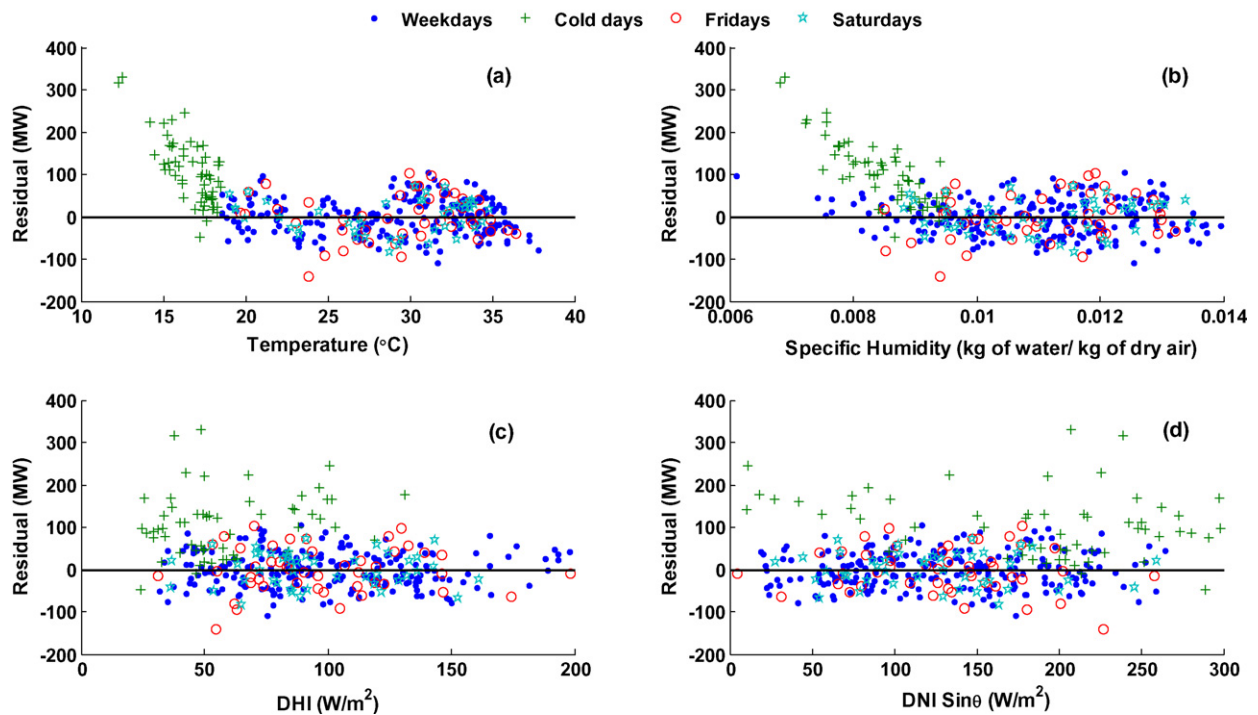|	Coefficient		t-Value [51,41]		t-Value significance	
	No-growth	Linear-growth	No-growth	Linear-growth	No-growth	Linear-growth
Growth/day	–	0.00035	–	0.010	–	0.992
Constant	−803.22	−645.59	−15.484	−14.235	<10^{-3}	<10^{-3}
Saturday constant	−19.186	−17.332	−2.278	−2.401	0.023	0.017
Friday constant	−61.864	−58.475	−7.523	−8.295	<10^{-3}	<10^{-3}
Temperature	36.718	34.333	33.327	36.323	<10^{-3}	<10^{-3}
Specific humidity	68001	52335	17.634	14.783	<10^{-3}	<10^{-3}
DHI	0.516	0.747	2.306	3.885	0.022	<10^{-3}
DNI on vertical surface	0.596	0.501	4.256	4.143	<10^{-3}	<10^{-3}

Fig. 4. Plot illustrating random distribution of residuals with weather parameters of linear-growth model 7 (data set A).

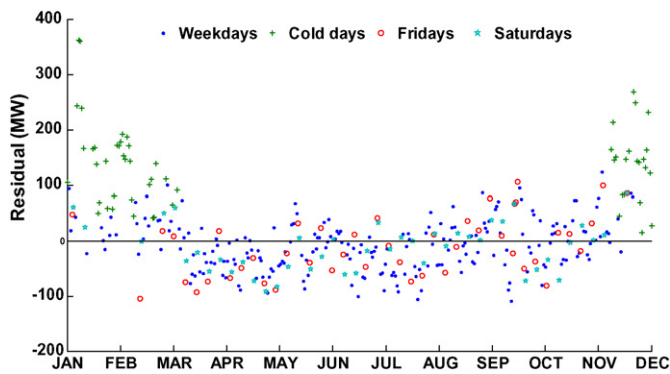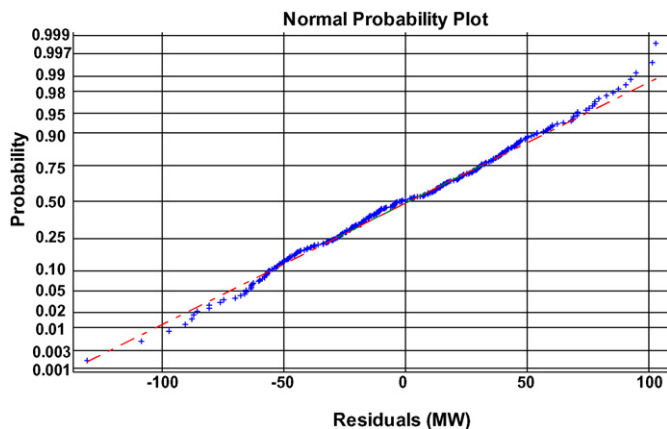**Fig. 5.** Plot illustrating moderate structure in the residuals of linear-growth model 7 (data set A).

Although the cooling load on these days may be so small as to be indistinguishable from random variations in the base load, Eq. (5) can be used to rank the *potential* for total daily cooling load. The most sensitive test is the one that considers each day in order of increasing cooling potential. A linear model was fit to the successive data points representing the lowest cooling potential. The deviation of the subsequent data point from the linear model was computed. Grubbs outlier detection test [39] was applied to the resulting data set at the 95% confidence level. The first outlier was detected at the 8th point.

The *t*-statistic indicates that the trend of first 7 points is not significant. This means that the daily load of the lowest ranked 7 days is essentially invariant with the cooling potential. Therefore, the mean weekday base load is computed from these 7 data points as 596.5 MW with standard deviation 21.73 MW. The foregoing points pertaining to the relation between daily load and cooling potential are illustrated in Fig. 8.

Although a good deal of cooling equipment may be left running, the lack of any relation between load and cooling potential is an
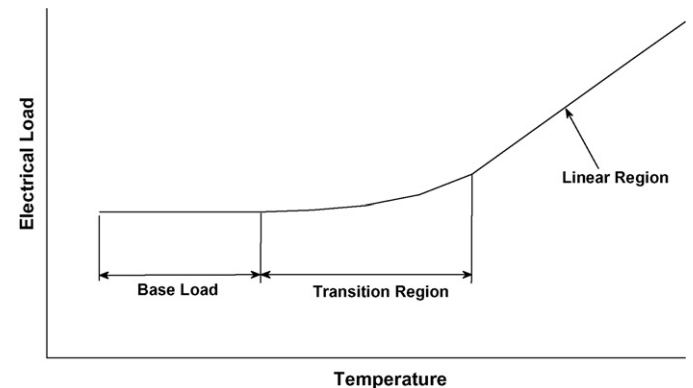

Fig. 6. Plot illustrating verification of normality of errors assumption in regression (data set B).

Fig. 7. Three regions characterized by different behavior of electrical load with temperature (hypothetical non-heating climate).
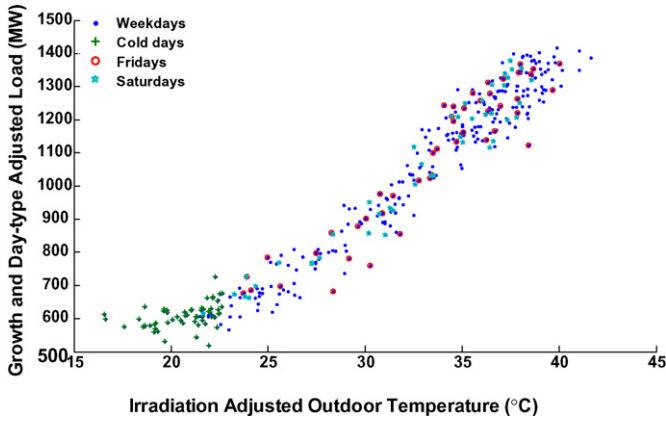
Fig. 8. Urban-growth and day-type-adjusted daily load versus solar-irradiation-adjusted outdoor temperature (data set A).

indication that there is very little *useful* cooling in the lowest ranked 7 days. Many commercial building cooling plants produce chilled water 24 × 7 even when there are no coil loads.

The daily cooling load contributions may be evaluated as follows:

$$
\begin{aligned}
W_T &= C_4(T - T_{BASE})^+; \\
W_w &= C_5(w - w_{@TBASE})^+; \\
W_{DHI} &= C_7 DHI; \\
W_{DNI\_V} &= C_9 DNI \sin\theta
\end{aligned}
\tag{6}
$$

where $(x)^+$ returns only positive values (i.e. x for $x > 0$ and 0 otherwise). To see the weather effects more clearly in Fig. 9 we plot the daytype- and growth-normalized total load (top trajectory) given

by:

$$
W_1 = C_1 + C_4 T + C_5 w + C_7 DHI + C_9 DNI \sin\theta;
\tag{7}
$$

The boundaries of the shaded areas plotted in Fig. 9 are given by successive subtractions:

$$
\begin{aligned}
W_2 &= W_1 - W_T; \\
W_3 &= W_2 - W_w; \\
W_4 &= W_3 - W_{DHI}; \\
W_5 &= W_4 - W_{DNI\_V};
\end{aligned}
\tag{8}
$$

The temperature and humidity at the base of the linear region define a transition load given by:

$$
W_X = C_1 + C_2 D_s + C_3 D_f + C_4 T_{BASE} + C_5 w_{@TBASE}
\tag{9}
$$

When W_1 falls below the base load, W_1 is taken equal to base load. When $W_5 < W_X$, each of the components from W_1 to W_5 are multiplied by an interpolation factor. The interpolation factor must provide a continuous response surface between the linear region (W_X), and the base load (W_{BL}), identified by outlier detection, as the mean of the 7 points that fall below the transition region. One such interpolating factor is defined as:

$$
\text{Interpolation factor} = \frac{W_1 - W_{BL}}{W_1 - W_X}
\tag{10}
$$

Estimates of the annual contributions to the total annual electrical demand of Abu Dhabi can now be evaluated by summing W terms over the year and taking the appropriate differences. The results of these calculations are presented in Table 5. The solar contribution to the electrical load is seen to represent a relatively small fraction, about 8% of total electrical load or 20% of the cooling load. While the combined effect of solar radiation components is remarkably flat over the year, the contribution of the direct solar component, DNI sin θ, is seen to be much higher than the contribution of the diffuse solar component in the winter months. Around
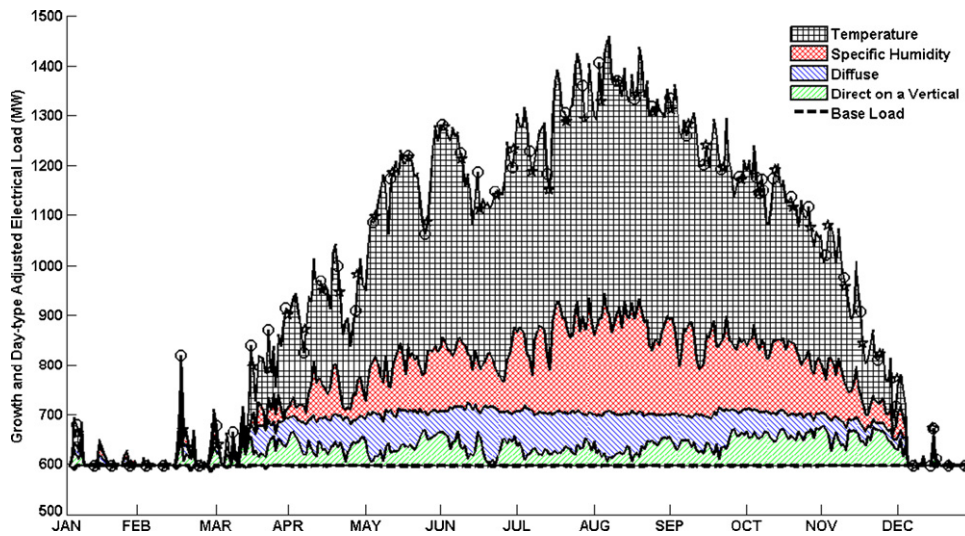

Fig. 9. Daytype and growth-normalized model (Eqs. (6)–(10)) estimates of load attributed to each weather component.

Table 5
Contributions of different model parameters in growth and day-type-adjusted-electrical load for specified-growth model 7 with change-points (data set A).

	Total (TW h)	Average (MW)	% of Total load	% of Cooling load
Temperature	2.0164	231.45	23.4	59.01
Specific humidity	0.7215	82.81	8.4	21.12
DHI	0.3767	43.24	4.4	11.03
DNI on vertical surface	0.3021	34.68	3.5	8.84
Base load	5.1954	596.50	60.3	–
Total load	8.6121	988.68	100	–
Total cooling electrical load	3.4167	392.18	39.7	100

21% of the annual electrical cooling load is associated with humidity which suggests that buildings are either leaky or that the share of fresh air in supply air to the building is much higher than necessary. Temperature accounts for 23% of the total load or 59% of the cooling load. This does not fully explain the relative success of models that use temperature as the only weather variable. Rather it is probably the fact that there is also a strong correlation between humidity and temperature that makes temperature-only based models viable in most building populations.

The average total load, 988.68 MW, minus the estimated average base load, $(366*596.5 − 52*17.332 − 57*58.475)/366 = 584.93$ MW, gives an estimated average cooling load of 404 MW or 40% of the total. The fraction of the peak daily load that can be attributed to cooling is also of interest. The peak load of 1532.5 MW occurs on a Sunday (workday in UAE) and this represents a cooling load that is $(1532.5 − 596.5)/1532.5 = 61\%$ of the peak day average total load. The cooling load fraction is similar to that found in nearby Al-Ain [47].

5. Discussion

As reported by others [48,49], we find that electrical load is strongly correlated with exposure of buildings to solar irradiation. However, in the case of Abu Dhabi, we have found that most of the large seasonal variation in electrical load must be attributed not to solar irradiation but to specific humidity and temperature. The annual solar contribution to cooling load is comparable to the annual humidity contribution but much flatter. Nevertheless, the day-to-day variations in electrical load are well, and to large extent, independently correlated with the two solar terms DHI and DNI $\sin θ$. The coefficients of these two terms may be thought of as corresponding to aggregate equivalent aperture areas for the interception of diffuse and direct solar irradiation. The coefficient of the temperature term may be thought of as a sum of the aggregate envelope thermal conductance and aggregate ventilation thermal capacitance rate. The coefficient of the humidity term can be interpreted as an aggregate ventilation rate times the enthalpy of condensation per unit of ventilation air. Embedded in the foregoing interpretations is an unknown constant of proportionality representing the average COP for community-aggregate cooling systems.

5.1. Limitations of a community-aggregate load model

The impact that non-cooling electrical loads, such as lighting, have on cooling loads cannot be assessed from the models and data investigated here because the base load term includes not only the equipment and lighting inside buildings but also exterior lighting, street lighting, distribution losses, and other urban infrastructure loads such as communications, and water and sewer pumping stations. The interior non-cooling loads contribute to the thermal load which, in turn, increases the electrical cooling load, while the exterior electrical loads do not contribute to cooling load.

In general, each building has its own energy signature comprising a unique change-region and base load intercept: (1) some buildings may follow a linear load-temperature curve to well below the threshold determined for the full sample, (2) after cooling ends, some buildings may begin to call for heating, (3) yet others may call for heating in some spaces while cooling is still called for in others, (4) use of reheat may be substantial in some buildings and (5) others may never call for heating at all. These behaviors have been identified historically by identifying ES/ETP models at the individual building level and there is reason to expect that, to the extent of local homogeneity in building thermal and use characteristics,

feeder-level model identification will result in more clearly defined heating and cooling transition regions.

The effects of COP and saturation with outdoor temperature were not modeled. However with a stationary data set, e.g. load time-series for a part of the community that is relatively free of demolition and construction, it is possible that these second order effects could be reliably identified.

5.2. Limitations of building-level models

It is currently possible for a utility to process and fit a change point or change-region model to every single metered building in its service territory based on monthly billing data [50] and with advanced metering, a daily data feed is economically, as well as technically, feasible so that such a process may be automated. But even with building level data, the allocation of cooling load to temperature, humidity, and solar irradiation, is problematic under light cooling loads. One can begin to estimate that part of the cooling load tied to internal gains only after the base load has been separated into indoor and outdoor loads. These details, if needed, are best (can, perhaps, only) be understood by end-use metering.

6. Conclusion

A change-point model with a transition region interpolating function has been developed. The change points are defined in terms of linear combinations of the weather variables. The linear-region model is found to have different sensitivities to the vertical and horizontal components of DNI as would be expected on physical grounds.

The analysis suggests that the main characteristics of aggregate cooling load can be inferred approximately from electric utility system load time-series data and contemporaneous local weather when the building population in question is sufficiently homogeneous and the aggregate cooling load is a large fraction of the system load. For the case studied, we see that the decrease in electrical load is relatively small on weekends, suggesting that a substantial fraction of cooling equipment and/or other loads are in continuous operation. The significant contribution of humidity to cooling load suggests that ventilation retrofits could have a large impact. These might include weatherization, demand-controlled ventilation, dedicated outdoor air systems, enthalpy recovery and desiccant dehumidification.

Building thermal inertia has not been modeled. For the 1-day sampling time used in this analysis, thermal inertia effects are generally small. Other second order effects like sol-air temperature (effect of wind and irradiation on building exterior surface temperatures), wind- and buoyancy-driven infiltration, and COP variations with ambient dry-bulb and wet-bulb temperatures are not currently modeled. Saturation effects – when loads exceed cooling capacity of some buildings during extreme heat – are likewise not modeled. Further work should attempt to address these effects by using more detailed data sets, e.g. contemporaneous interval meter data from a sample of individual buildings, seasonal water heating estimates, building stock time-series, street lighting schedules and system water use.

References

[1] C. Strock, C.H.B. Hotchkiss, Heating and Ventilating Degree-Day Handbook, Industrial Press, 1937.

[2] H.C.S. Thom, Normal degree days below any base, Monthly Weather Review 82 (1954).

[3] T.A. Reddy, D.E. Claridge, Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption, Energy and Buildings 21 (1994) 35–44.

[4] D.O. Stram, M.F. Fels, The applicability of PRISM to electric heating and cooling, Energy and Buildings 9 (1986) 101–110.

[5] C. Crowley, F.L. Joutz, Weather Effects on Electricity Loads: Modeling and Forecasting, Department of Economics, George Washington University, 2005.

[6] M.A. Al-Iriani, Climate-related electricity demand-side management in oil-exporting countries—the case of the United Arab Emirates, Energy Policy 33 (2005) 2350–2360.

[7] D. Ruch, D.E. Claridge, A four-parameter change-point model for predicting energy consumption in commercial buildings, Journal of Solar Energy Engineering 114 (May) (1992) 77–83.

[8] D. Ruch, L. Chen, J.S. Haberl, D.E. Claridge, A change-point principal component analysis (CP/PCA) method for predicting energy usage in commercial buildings: the PCA model, Journal of Solar Energy Engineering 115 (May) (1993) 77–84.

[9] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D.P. Lalas, M. Moschovits, F. Karagiannis, D. Papakonstantinou, Models for mid-term electricity demand forecasting incorporating weather influences, Energy 31 (2006) 208–227.

[10] A.Z. Al-Garni, Y.N. Al-Nassar, S.M. Zubair, A. Al-Shehri, Model for electric energy consumption in Eastern Saudi Arabia, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 19 (1997) 325–334.

[11] M.A. Badri, A. Al-Mutawa, D. Davis, D. Davis, EDSSF: a decision support system (DSS) for electricity peak-load forecasting, Energy 22 (June) (1997) 579–589.

[12] K. Liu, S. Subbarayan, R.R. Shoults, M.T. Manry, C. Kwan, F.I. Lewis, J. Naccarino, Comparison of very short-term load forecasting techniques, IEEE Transactions on Power Systems 11 (1996) 877–882.

[13] J.R. Forrester, W.J. Wepfer, Formulation of a load prediction algorithm for a large commercial building, ASHRAE Transactions 90 (1984) 536–551.

[14] J.E. Seem, J.E. Braun, Adaptive methods for real-time forecasting of building electrical demand, ASHRAE Transactions 97 (1991) 710–721.

[15] F. Flouquet, Local weather correlations and bias in building parameter estimates from energy-signature models, Energy and Buildings 19 (1992) 113–123.

[16] H.K. Alfares, M. Nazeeruddin, Electric load forecasting: literature survey and classification of methods, International Journal of Systems Science 33 (2002) 23–34.

[17] S. Hammarsten, A critical appraisal of energy-signature models, Applied Energy 26 (1987) 97–110.

[18] T. Kusuda, T. Tsuchiya, F.J. Powell, Prediction of indoor temperature by using equivalent thermal mass response factors, in: Proceedings of the 5th Symposium on Temperature NBS, 1971, p. 1345.

[19] R.C. Sonderegger, Dynamic models of house heating based on equivalent thermal parameters, Ph.D. Thesis, 1978.

[20] G.J. Levermore, Performance lines and energy signatures: analysis with reference to the CIBSE Building Energy Code, Building Services Engineering Research and Technology 16 (1995) 47–50.

[21] T.A. Reddy, J.K. Kissock, S. Katipamula, D.E. Claridge, An energy delivery efficiency index to evaluate simultaneous heating and cooling effects in large commercial buildings, Journal of Solar Energy Engineering 116 (1994) 79–87.

[22] M. Bauer, J.L. Scartezzini, A simplified correlation method accounting for heating and cooling loads in energy-efficient buildings, Energy and Buildings 27 (1998) 147–154.

[23] A.J. Heller, Heat-load modelling for large systems, Applied Energy 72 (2002) 371–387.

[24] C. Ghiaus, Experimental estimation of building energy performance by robust regression, Energy and Buildings 38 (2006) 582–587.

[25] H. Hahn, S. Meyer-Nieberg, S. Pickl, Electric load forecasting methods: tools for decision making, European Journal of Operational Research 199 (2009) 902–907.

[26] S.L. Wong, K.K. Wan, T.N. Lam, Artificial neural networks for energy analysis of office buildings with daylighting, Applied Energy (2009).

[27] F.W. Yu, K.T. Chan, Energy signatures for assessing the energy performance of chillers, Energy and Buildings 37 (2005) 739–746.

[28] J.C. Lam, K.K. Wan, K.L. Cheung, An analysis of climatic influences on chiller plant electricity consumption, Applied Energy 86 (2009) 933–940.

[29] R. Zmeureanu, Assessment of the energy savings due to the building retrofit, Building and Environment 25 (1990) 95–103.

[30] A. Adderly, P. O'Callaghan, S. Probert, Energy-saving options, Applied Energy 30 (1988) 269–279.

[31] S. Jebaraj, S. Iniyan, A review of energy models, Renewable and Sustainable Energy Reviews 10 (2006) 281–311.

[32] A. Al-Shehri, A simple forecasting model for industrial electric energy consumption, International Journal of Energy Research 24 (2000) 719–726.

[33] H.L. Willis, J.E.D. Northcote-Green, Comparison tests of fourteen distribution load forecasting methods, IEEE Transactions on Power Apparatus and Systems (1984) 1190–1197.

[34] P. Waide, Keeping you cool: an overview of trends, International Energy Agency presentation, June (2004).

[35] F.J. Peterson, J.E. Patton, M.E. Miller, R.A. Gillman, W.M. Warwick, W.F. Sandusky, End-Use Load and Consumer Assessment Program: motivation and overview, Energy and Buildings 19 (1993) 159–166.

[36] H. Leong, G. Johnson, Modeling of energy consumption for space heating for a community via GMDH algorithm, in: IEEE International Conference on Cybernetics and Society, 1979.

[37] Welcome to Abu Dhabi—population, http://www.visitabudhabi.ae/en/uae.facts.and.figures/population.aspx.

[38] G.A. Seber, C.J. Wild, Nonlinear Regression, Wiley-IEEE, 2003.

[39] G. Frank, Procedures for detecting outlying observations in samples, Technometrics 11 (1969) 1–21.

[40] S. Katipamula, T.A. Reddy, D.E. Claridge, Bias in predicting annual energy use in commercial buildings with regression models developed from short data sets, in: International Solar Energy Conference Proceedings, Energy Systems Laboratory, Texas A&M University, 1995.

[41] H.M. Wadsworth, Handbook of statistical methods for engineers and scientists, McGraw-Hill Professional (1998).

[42] N.E. Savin, K.J. White, The Durbin–Watson test for serial correlation with extreme sample sizes or many regressors, Econometrica: Journal of the Econometric Society 45 (1977) 1989–1996.

[43] J.K. Kissock, T.A. Reddy, D.E. Claridge, Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings, Journal of Solar Energy Engineering 120 (1998) 168–176.

[44] T.A. Reddy, J.K. Kissock, D.K. Ruch, Uncertainty in baseline regression modeling and in determination of retrofit savings, Journal of Solar Energy Engineering 120 (1998) 185–192.

[45] J.M. Gordon, K.C. Ng, Cool Thermodynamics, Cambridge International Science Publishing Cambridge, 2000.

[46] J. Kelly Kissock, C. Eger, Measuring industrial energy savings, Applied Energy 85 (2008) 347–361.

[47] H. Radhi, Evaluating the potential impact of global warming on the UAE residential buildings—a contribution to reduce the CO_2 emissions, Building and Environment 44 (2009) 2451–2462.

[48] P. Lamp, F. Ziegler, European research on solar-assisted air conditioning, International Journal of Refrigeration 21 (1998) 89–99.

[49] W. Kessling, M. Peltzer, Innovative systems for solar air conditioning of buildings (2004). http://repository.tamu.edu/handle/1969.1/4601.

[50] R.C. Sonderegger, A baseline model for utility bill analysis using both weather and non-weather-related variables, Transactions-American Society of Heating Refrigerating and Air Conditioning Engineers 104 (1998) 859–870.

[51] Hypothesis tests in multiple linear regression, Experiment Design and Analysis Reference, http://www.weibull.com/DOEWeb/experiment_design_and_analysis_reference.htm#hypothesis_tests_in_multiple_linear_regression.htm.

[52] T. Amemiya, Nonlinear regression models, Handbook of Econometrics 1 (1983) 333–389.