Low stratus as a driver of electrical demand variability independent of temperature in the greater Los Angeles region

David W. Pierce
Daniel R. Cayan

Scripps Institution of Oceanography

Version 2 22 May 2014

Abstract

The relationship between low cloud in the in the greater Los Angeles region (Southern California Edison service area) and discrepancies from the electrical demand expected for each day based solely on observed temperatures is examined using historical meteorological station temperature observations, GOES satellite cloud albedos, and daily peak electrical demand as recorded by the California Independent System Operator (CalISO). Electrical demand from the Western Electricity Coordinating Council (WECC) was obtained as well, but not used due to errors in the recorded times in the data set and disagreement between CalISO and WECC in reported loads for 2005. The historical meteorological data exhibit problems as well, including blocks of time when uncorrected shifts in the time of observation cause daily maximum temperatures to be mis-registered, as indicated by disagreements with nearby stations reporting hourly observations. A daily peak electrical demand model was formed by a regression between daily minimum and maximum temperature, weekend/holiday status, and observed electrical load. Observed departures from the load calculated by this model were analyzed to determine their relationship to cloud cover. Days with more electrical demand than anticipated (based upon the temperature) were associated with a weak or absent marine layer, while days with less electrical demand showed a stronger than usual cloud cover in the region. The relationship between

cloud cover and load discrepancies could be further explored with more spatially complete temperature measurements, which might uncover whether the discrepancies are due to lack of temperature measurements in the critical region at the mean inland edge of the marine layer, or to additional solar heating of structures.

1. Introduction

The electrical demand for a day is strongly influenced by the day's peak temperature, especially in relatively warm regions where annual electrical load is dominated by summer air conditioning, such as California. Temperatures from the two previous days also have some effect (Garcia-Cerrutti et al. 2011). Beyond temperature, other meteorological conditions may affect energy demand as well, such as humidity, wind speed, and solar insolation (e.g., Robinson 1997; Mirasgedis et al. 2006). However non-temperature meteorological factors are small by comparison to temperature, and may be hard to uncover given the pronounced serial auto-correlation in meteorological time series (Mirasgedis et al. 2006).

Although energy utilities consider their demand forecast methods proprietary and so are not generally available for study, the California Energy Commission (CEC) evaluates energy demand using a weighted average of the current day and two previous days' temperature, along with information on the day of the week (weekends generally have lower electricity demand for a given temperature) and civic holidays (Garcia-Cerrutti et al. 2011). We adopt this methodology to estimate electrical demand in the greater Los Angeles region (specifically, the Southern California Edison service area) during the summer of 2010 on the basis of temperatures observed on that day, and compare estimated to actual peak electrical demand recorded on that day. Using observed temperatures in our electrical demand estimate excludes as a source of load forecast errors any errors in forecasting daily minimum/maximum temperature, which are not of interest here, although of course they are an operational concern. Our purpose is to examine discrepancies between the actual electrical demand and that anticipated on the basis of temperature, day-of-week, and holiday status, and determine if those discrepancies are associated with varying cloud cover in the region. In other words, we construct a surrogate electrical demand "hindcast" (similar to a forecast, but based on historical observations) using daily temperature and then investigate the errors in the hindcast to determine if there are systematic influences by other meteorological factors, such as cloud cover, which explain why the anticipated demand was not realized.

This study brings together meteorological data with electrical load data to investigate how they are related. Section 2 outlines the sources of our data. We found notable problems with the data from both sources that make this analysis challenging. Results from the data analysis are shown in section 3 and discussed in section 4. Summary comments along with suggestions for future work are given in section 5.

2. Data

2.1 Meteorological data

2.1.1 Station temperature data

The California Energy Commission (CEC) bases the utility-wide average temperature used for forecasting electrical demand on a limited number of stations that have a high quality, long term record. The station locations are shown in Figure 1, and the weighting applied to each station when computing the average temperature for the Pacific Gas & Electric (PG&E), Southern California Edison (SCE), and San Diego Gas & Electric (SDG&E) utilities are shown in Table 1.

Station temperature data was obtained from the Global Surface Summary of Day (GSOD, version 7; ftp://ftp.ncdc.noaa.gov/pub/data/gsod/) archive, and includes daily minimum (Tmin) and maximum

(Tmax) temperature. For two stations (KSEE and KNKX) with substantially incomplete records in GSOD this was augmented with data from the daily Global Historical Climatology Network (GHCN; http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/). Data gaps in individual station records were filled by linear regressions based upon historical relationships with the nearest stations with valid data. After this procedure the utility-averaged, weighted Tmax time series was typically left with ~20 unfilled gaps over the data period with an average length of ~1.5 days (0.04% of the data).

One problem with the station temperature observations is that some stations report a daily reading (as opposed to hourly), and the time of observation has shifted over the years at some stations. At one time many stations that report once-daily temperature took the observation in the late afternoon. In this event, the day's maximum temperature is likely to be reported for the day of observation. More modern practice is to take a single observation in the morning. This means that the maximum temperature reported belongs to the previous day.

Although some meteorological station records are supposed to be corrected for this effect, evidence often suggests otherwise. This can be determined using reference meteorological stations that report hourly data, in which case there is no ambiguity about the day that the measurement applies to. For example, Figure 2 shows the lag (in days) between annual time series of daily Tmax at KBUR (the Burbank airport) from the GSOD once-daily data set, and daily Tmax derived from 4 nearby hourly-reporting reference stations (shown in the 4 panels). In all 4 hourly reference stations a clear pattern emerges that the hourly and GSOD data agree on the reported day between 1975 and 1992, and again after 2005 (i.e., the correlation between the two time series peaks at a lag of 0 days), but disagree between 1993 and 2002. During the latter period there is a one-day lag between the two time series, which suggests that during this period the GSOD data set may be using a morning reporting time without being corrected. In this work, which is meant to be exploratory and not comprehensive, we only use data from 2010, and checked that the lagged correlations suggest that the reported times are correct in the stations we use, but an incorrectly reported day in the temperature data could be a significant issue during some time periods.

The size of the red dots in Figure 2 are proportional to the correlation value (in the one-year window) between the GSOD Tmax time series and the Tmax time series derived from the hourly reference station. The top left panel is particularly interesting because it shows the GSOD and hourly data for the same station, KBUR. Correlation values are not uniformly close to 1.0, as might be expected. To see whether this is a quirk of KBUR or a more general problem, Figure 3 shows time series of hourly temperature (red line) and daily Tmax (horizontal blue lines) for KSAN (San Diego Lindbergh Field) for the period 7 April 2006 to 20 April 2006. Although many days show the same maximum temperature in the two records, there are also some sizable differences. On 14 April 2006, the hourly temperature readings never exceeded 18°C, while the daily maximum temperature reported for this day in the GSOD data set exceeds 31°C, a difference of 13°C.

Lacking complete metadata on the instruments we can only speculate why such a discrepancy might occur. An instrumental problem is always possible. On the other hand, the sudden drop in daily maximum temperature from April 13 to 14 suggests that a strong marine layer may have formed on the 14th, which is not unusual at this location. Such marine layers can have sharp transitions between clear and cloudy conditions, so perhaps the discrepancy reflects a different location for a daily maximum recording thermometer and an hourly reading temperature instrument. Or, perhaps a maximum recording thermometer is used for the GSOD data and it was not reset correctly, and so is reporting the previous day's late afternoon high temperature as the value for the 14th. The one thing we can say with certainty is that the various temperature records are not self-consistent, even for a single reported location.

2.1.2 Cloud data

Cloud data in the form of albedo estimated using the visible wavelength sensor on the GOES satellite (http://www.goes.noaa.gov/) was provided by S. Iacobellis of the Scripps Institution of Oceanography (Iacobellis and Cayan 2013). The data were processed as documented at http://www.star.nesdis.noaa.gov/smcd/spb/fwu/homepage/GOES_Imager.php, including corrections for degradation of the sensor (post-launch calibration), and has been interpolated to a fixed 1 km grid (S. Iacobellis, *pers. comm.*). Data for the period we analyze here (summer 2010) was obtained by GOES11.

2.2 Electrical load data

Electrical load data were obtained from two sources: 1) Hourly load data from the Western Electricity Coordinating Council (WECC) as archived by the California Energy Commission over the period 2005-2010 (described in detail in section 2.2.2). These data are not publically available. 2) Hourly load data from the California Independent System Operator (CalISO), which operates California's electrical power grid, via its Oasis web site, both the newer (http://oasis.caiso.com) and older (http://oasishis.caiso.com/) versions (section 2.2.3). These data are freely available.

2.2.1 Hour beginning versus hour ending load data

The electrical load data are reported using one of two different time standards: hour beginning or hour ending. The meaning of these two standards was determined by comparing reported hour beginning and hour ending values to the CalISO's real-time updating load graph (http://www.caiso.com/Pages/Today's-Outlook-Details.aspx), which was assumed to provide a correct reference load versus time relationship. Based on this comparison, hour beginning values are (as the name suggests) the instantaneous demand values at the start of the hour. So, for example, a reported electrical demand value for "hour beginning 07" shows the instantaneous electrical demand that occurred at 7:00 AM.

A similar comparison shows that "hour ending" values are arithmetic means over the hour previous to the reported time. So, for example, a value for "hour ending 9" is approximately the average from 8:01 AM to 9:00 AM, i.e., centered at a time of 8:30 AM. The comparison of recorded values to the real-time load graph is not sensitive enough to determine if the exact time averaging interval is 8:00 AM to 8:59 AM, 8:01 AM to 9:00 AM, 8:00 AM to 9:00 AM, or something else. However, for our purposes, this one-minute uncertainty is not relevant, since the meteorological observations are only available with a resolution of 1 hour or longer in any event.

Since hour starting data is instantaneous at the start of the indicated hour but hour ending data is the average over the previous hour, time series of the same actual loads but reported in the two different ways will be shifted by half an hour, with the hour ending data lagging the hour starting data. E.g., an 8 AM reported value is centered at 8 AM in the hour starting data but 7:30 AM in the hour ending data.

2.2.2 WECC load data

The WECC data is available with hourly time resolution over the period 2005-2010. Values are spatially aggregated by electrical utility.

In California during the summer, electrical load peaks in the late afternoon due to air conditioning and is smallest in the early morning hours (~4 AM). The mean daily cycle found in the WECC data in July for each available year (2005-2010) is shown in Figure 4 for the Los Angeles Department of Water and Power (LADWP). To facilitate comparing different years the data have been normalized so that the difference between the minimum and maximum value is 1.0, which reduces discrepancies between the cycles that arise from changes in population and economic activity over the period. (In particular, the recession of 2008 had a notable effect on electrical demand.)

For convenience the horizontal grey line and annotations in Figure 4 show the local time of day at which the average July electrical demand reached 63% of maximum. It is immediately evident that the reported times in the WECC data set cannot be taken at face value, since it is implausible that electrical demand as a function of time of day would shift in nearly exact half- or full-hour increments between different years, as can be seen by the time when each year's curve reaches 63% of maximum (annotated on the figure). More likely is that different standards of reporting have been used in different years, resulting in a varying (and undocumented) time standard used for each year. Some of the factors that could explain this varying time standard include: 1) Switching between hour starting and hour ending reporting, which, as noted above, causes a half hour shift in the time of reported data; 2) Switching between reporting times that include daylight savings time and times ignoring daylight savings, which would cause a 1 hour shift in the time of the reported data; 3) Switching between reporting times based on time at WECC headquarters (located in Salt Lake City UT, which is in the mountain time zone) and time at the local utility (the Pacific time zone, for LADWP data). This would also cause a 1 hour shift in the time of the reported data.

Figure 4 suggests that 2010 was reported with a different hour starting/ending standard than the other years, since otherwise there is no plausible way to generate a half-hour shift in the data. Data for year 2005/06 has a one-hour shift relative to data for years 2007-9, which suggests that one of the one-hour shifts noted above are at play, i.e., a change between reporting times that include or exclude daylight savings time, or a change between reporting times on the basis of time at the WECC headquarters versus time at the local utility. Given the absence of metadata recording the process by which the times of the electrical loads were recorded this cannot be determined with certainty, however the large, integer multiple of 30-minutes shifts between groups of consecutive years (2005/6, 2007/8/9, and 2010) that themselves have relatively minor variation between them is a clear sign that these problems exist in the reported times. Errors in reported time on sub-year timescales were not examined.

2.2.3 CalISO OASIS load data

The OASIS data is available from before 2002 to present time, with near-real time updating. In this work we do not examine data prior to 2002 since the so-called California electricity crisis affected loads in 2000 and 2001 (e.g., Sweeney 2002).

The daily cycle of electrical demand in July over the period 2005-2010 is shown in Figure 5 for Southern California Edison (SCE). As indicated in Figure 1 and Table 1, SCE demand is dominated by temperatures in the greater Los Angeles region (represented by stations in Riverside, Long Beach, and Burbank). The actual service area of SCE covers a large part of both southern and central California, which is why Fresno temperatures contribute to SCE electrical demand (Table 1), but the weighting is only 6%. Accordingly, Figure 5 (SCE) can be compared to Figure 4 (LADWP), at least in a qualitative way.

It is clear from Figure 5 that the OASIS data does not suffer from the time reporting problems seen in the WECC data. In addition, the OASIS data is specifically indicated as being reported on an hour ending basis, and that is consistent with the fact that OASIS-reported SCE load reaches 63% of peak load at 10:30 AM (i.e., on the half hour; Figure 5).

2.2.4 Comparing WECC and OASIS data

Comparing Figure 5 to Figure 4, we speculate that WECC year 2010 was reported on an hour ending basis but neglecting daylight savings time (10:30 PDT is 9:30 PST), which would account for a 1-hour shift earlier in the WECC data, while 2005/6 were reported on an hour beginning basis, which would account for a half-hour shift later in the WECC data. Similarly, 2007/8/9 may have been reported on an hour beginning basis with respect to the WECC headquarters time zone (since the reported times appear to be 1.5 hours later than found in OASIS).

The problems with the changing time standard in the WECC data can be avoided if the daily maximum load is examined, since that is the same no matter what time standard is used. Since we have data for both WECC and OASIS, we can compare their daily maximum loads to see if they agree. Results are shown in Figure 6, presented as both time series (upper panel) and scatterplots (lower panel).

Although there are a few random days when the peak loads disagree (generally with WECC reporting higher loads than OASIS), by and large the WECC and OASIS peak loads are consistent in years 2006-2010 (nearly all the green circles fall on the black one-to-one regression line). However in year 2005 the two data sets notably disagree, as can be seen both in the time series and in the regression, with the WECC data reporting uniformly lower daily peak loads than OASIS (brown circles in lower panel). The fact that the WECC values are much lower than other years in 2005 than 2006, while the OASIS data is about the same in 2005 and 2006 as other years, suggests that it is the WECC data that has the error, barring a reasonable explanation for why electrical demand should have been anomalously low in 2005. The differing slopes of the green and brown lines in the lower panel of Figure 6 suggest that the error is proportional to the total load, i.e., as if there is a missing load in the WECC data that is responding to the same environmental/meteorological conditions as the rest of the electrical system, rather than being a fixed constant load error.

Because of these errors in the WECC data, we used the CalISO OASIS load data in this work.

3. Results

3.1 Load-temperature relationships

Our purpose is to examine discrepancies between the actual electrical load on summer days and the anticipated load based on temperature alone, so we must first develop a credible model for demand as a function of temperature. We do this using regressions, as will now be described.

The relationship between utility region-averaged Tmax (using the stations and weights in Table 1) and electrical demand is shown in Figure 7 for PG&E, SCE, and SDG&E. Different years are indicated by colors. Red shows results from 2002, the first year in the data set, which has the highest cool-temperature energy use of any of the years in PG&E. In the other utilities 2002 is one of the lowest consuming years at the cooler temperatures, as might be expected given the general increase in population and economic activity over the time period (notwithstanding the recession of 2008/9). The reason for this discrepancy in the 2002 PG&E data is unknown.

In all the utilities electrical demand is minimized when daily maximum temperature is just under 70°F. Additionally, the increase in electrical demand with temperature is approximately linear for warm daily temperatures, which we take advantage of by exploring regressions between electrical demand and Tmax only during a three-month summer period (15 June to 15 September, taken to be consistent with Garcia-Cerrutti et al. 2011). Since it is anticipating the year's highest peak electrical demands that is of primary interest for avoiding the potential of brownouts or rolling blackouts, it is sensible to restrict our attention to load forecast errors found in the summer.

The results of the regression model predicting observed load from temperature are shown in Figure 8. A generalized least squares fitting routine was used, with a method based on maximizing the restricted log-likelihood. Additionally, because of the strong serial autocorrelation in meteorological data, a first-order autoregressive (AR1) model was used to describe the correlation structure of the data being fitted. The estimated AR1 parameter is 0.44, 0.50, and 0.55 for PG&E, SCE, and SG&E, respectively, values that are large enough to justify taking the serial autocorrelation into account in the regression. The regression uses current day's Tmax rather than a weighted combination of the current and previous two days since a test with the latter formulation did not return a superior fit (as measured by fraction of variance explained) even though it is more complicated. Additionally, Tmin is included along with a flag indicating whether or not the day is a weekend or civic holiday.

Because of curvature in the load vs. temperature relationship for SDG&E even in summer (unlike the other two utilities), the CEC perform the fit for three temperature bands for SDG&E (Garcia-Cerrutti et al. 2011). Here we try to reduce the number of parameters somewhat by fitting SDG&E using both Tmax and Tmax^2, unlike the other two utilities. This increases the fraction of variance explained from 0.92 to 0.95, but the strong correlation between the Tmax and Tmax^2 (> 0.99) must be kept in mind.

Values for the regression parameters are given in Table 2 (including a regression for SDG&E that does not include $Tmax^2$, for comparison). Explained variance is lowest for PG&E, and highest for SCE, the utility we examine in more detail below.

Although the previous load-temperature regression was done for daily peak load, the regressions can be computed using hourly load as well (i.e., daily Tmin and Tmax related to each hour's electrical load). This addresses the question of how well each hour's load can be forecast given meteorological predictions of daily Tmin and Tmax. The fraction of explained variance in daily load for the summer of 2010 in SCE as a function of hour-of-day is shown in Figure 9. Daily temperatures are most predictive of loads in the afternoon hours, peaking at 2 PM. So even though Tmin is more statistically significant and explains a proportionally greater share of the variance than Tmax in the 5 AM regression (the reverse of the situation in the late afternoon), the Tmin-dominated early morning regression still does not forecast early morning electrical demand as well as Tmax forecasts afternoon electrical demand.

3.2 Clouds and load forecast errors in Southern California Edison

Using the regression described in Table 2 to provide an estimate of daily electrical load based on utility-averaged Tmax in SCE, and actual observed loads from OASIS, Figure 10 shows the difference (actual – predicted load) in SCE during the summer of 2010. The RMS error is ~550 MW. The brown areas show where actual demand exceeded that anticipated based upon temperature, which is potentially a difficult situation for the electrical system to handle, and can in extreme cases lead to brownouts or rolling blackouts if the unanticipated demand is high enough and regionally widespread. Green areas show where actual demand was less than anticipated. The time series of errors does not appear to be white noise, instead exhibiting a modest degree of serial auto-correlation (the actual value is 0.37), which is consistent with a meteorological influence on the load forecast discrepancies.

We examine the relationship between load forecast discrepancies and cloud cover by calculating the mean cloud albedo on days when the load forecast discrepancy was in the top tercile (most positive, i.e., the load higher than anticipated based on temperature), the bottom tercile (load lower than anticipated), and taking the difference, top tercile minus bottom tercile. The difference in load discrepancy between the top and bottom terciles is approximately 1180 MW, which represents about 5% of the peak SCE load of 24,000 MW seen over the recorded period (Figure 7). This approach shows if the cloud albedo is systematically different on days with positive and negative load forecast error. In the event that cloud albedo (and therefore cloud cover) is unrelated to the load forecast errors, then there will be no systematic difference in albedo between days with positive and negative load forecast errors. We also examined specific humidity (Abatzoglou 2012), but found no significant difference linked to load forecast errors.

Results from the tercile analysis of cloud albedos are shown in Figure 11. Since the strength of the marine layer (and therefore degree of cloudiness) is a strong function of time-of-day, with a tendency towards low stratus clouds in the morning that burn off by afternoon, the results are shown at 7 AM, 9 AM, 11 AM, and 1 PM local time. The top row shows the mean cloud albedo across all days in summer of 2010 (15 June to 15 September) at the indicated local time. There is a strong tendency for cloud cover to be present along the coastal regions of Los Angeles, Ventura, and Orange counties in the morning hours. The coastal cloud cover persists until at least 9 AM, but in the mean tends to retreat by 11 AM and is nearly gone by 1 PM.

The second row of Figure 11 shows the difference in cloud albedo (percentage points) between days with the highest and lowest load forecast discrepancies. There is a clear pattern of about −10 percentage points difference across areas of Los Angeles and Orange counties that are slightly inland from the coast, extending into the southwestern tip of San Bernardino county, as well as the coastal areas of Ventura county, that peaks between 9 and 11 AM local time. Since the inland edge of the mean cloud cover is very nearly coincident with the center of this negative pattern, we interpret this as a shift in the cloud cover's edge either inland or towards the coast. The sense of the difference is that during the highest load forecast discrepancies (demand higher than anticipated based on temperature), the albedo is lower (less clouds) than during the days when demand is lower than anticipated, when the albedo is higher and clouds extend inland further.

The difference in albedo shown in the middle row of Figure 11 is given in percentage points, but to really understand the significance of the difference it is necessary to compare the differences to the mean cloud albedo (top row of Figure 11). The bottom row of Figure 11 shows the percentage change in cloud albedo (with respect to mean climatological conditions over all included days) that is associated with the difference between the top and bottom terciles of load forecast errors. There is up to a 40% difference in albedo associated with the load forecast discrepancies.

In sum, these results show that electrical demand tends to be higher than anticipated (based solely on temperature) on days with a weak or missing marine layer along the coast, and vice versa. A point worth re-emphasizing is that the anticipated load already takes into account the actual day's observed temperature. It is not surprising that a day with no clouds is warmer than a day with clouds, and that since temperatures are warmer, electrical demand will be higher on days with a weak or missing marine layer. However the load model already uses the actual observed temperature for the day, so these results show that even when the higher temperatures experienced on days with no marine layer are taken into account, there is still a residual error such that lack of a marine layer leads to higher than anticipated electrical load.

The results of the regression and tercile analysis are worth illustrating with a specific case study. Figure 12 shows the forecast and actual loads for all days in the summer of 2010 for SCE. The two red dots show August 17 and 27, which are notable for having almost the exact same Tmax (differing by only 0.02°F) and similar Tmin (differing by ~1.7°F). Both were a weekday and not a holiday, and are separated by less than 2 weeks. Despite the similar temperatures and time of year, electrical demand was substantially higher on the 17th (20,700 MW) than on the 27th (18,900 MW), a difference of about 9.5%. Based on the regressions developed in section 3.1 the anticipated difference in electrical load (given the observed temperatures) was only about 0.3 MW (1.7%), so the actual difference was more than five times larger than forecast.

The cloud albedo fields for these two days are shown Figure 13. There is a considerable difference between the marine layer in the greater Los Angeles region between the two days, with 17 August (the day with high electrical demand) showing substantially less coastal clouds than 27 August (the day with low electrical demand). While any particular example using a pair of similar days does not constitute proof of the relationship suggesting that anomalously high electrical loads in SCE are associated with days lacking a marine layer, it is at least consistent.

4. Discussion

The analysis presented here suggests that electrical demand in the greater Los Angeles basin area is higher than the demand expected based on temperature alone when the marine layer is absent. What are some possible causes of this?

One possibility is that the limited number of stations used to estimate the utility-wide Tmax does not fully capture all the important temperature variability that drives demand variations. In the CEC scheme to forecast SCE electrical demand, which we use here, temperature is calculated from four stations (Riverside, Long Beach, Burbank, and Fresno), only one of which (Long Beach) is near the coast (see Table 1 and Figure 1). Our analysis suggests that the load forecast errors are associated with cloud variability on the inland edge of the marine layer (blue areas in the middle row of Figure 11). It is possible that additional temperature information taken specifically from that area could improve the load demand forecasts. The temperature gradient perpendicular to the coast could be estimated with better station coverage, and may be useful in the demand forecast model.

Another possibility is that the additional solar insolation on clear days (weak marine layer) increases the demand above and beyond what would be expected due to the ambient air temperature, perhaps by more efficiently heating structures through windows. This possibility could be ruled out if additional temperature measurements from regions underlying the inland edge of the marine layer showed that temperature alone was enough to anticipate the extra electrical demand, although a more elaborate analysis would be required to demonstrate a connection of load forecast errors to an increase in direct solar heating if the explanation based on temperature was not supported by additional data and analysis. In either event, the most straightforward way to explore this issue further would be to obtain temperature data from additional locations in the relevant parts of Los Angeles, Ventura, and Orange counties, and see if a higher-resolution depiction of daily maximum temperature were enough to account for the errors in the load forecast that we find in this region.

The possible role of early morning satellite observations of low stratus cover could also be explored, perhaps providing some limited same-day predictability of the marine layer's effect. The relationship between duration of the marine layer and the effect on load also has not been explored in this work, but could have some influence.

5. Summary

We have analyzed meteorological station temperature data, cloud fields from the GOES satellite, and historical electrical load data to examine the relationship between clouds and electrical load forecast errors in the greater Los Angeles region. The electrical load anticipated for a day is based on that day's observed minimum and maximum temperature and weekend/holiday status. Performing such an analysis is considerably hindered by problems with data quality in both the meteorological and electrical load data.

Comparing two sources of electrical load data, from the Western Electricity Coordinating Council (WECC) and from the California Independent System Operator (CalISO) OASIS web site, we found notable inconsistencies and discrepancies between them, particularly in the reported times associated with the WECC data. Implausible shifts in the daily cycle of energy use seen in different calendar years that occur at integer multiples of 30 minutes strongly suggest undocumented changes over the years in the reporting methodology of the times associated with the measured load. Additionally, an overall discrepancy between loads reported for 2005 in the WECC and OASIS data is evident, with no obvious explanation except that the WECC data set is missing about ~8% of California's electrical load in that year.

The meteorological data also has some distinctive problems. Based on lag correlations with hourly reporting reference stations , some of the temperature records clearly indicate that changing observation times in the daily temperature time series have not been adjusted for correctly (or at all). Additionally the observations occasionally exhibit substantial differences between daily maximum temperature values recorded in the hourly and daily time series even for the same station, far larger differences than can be accounted for by the hourly stations having a limited sampling rate. Such problems might be caused by physically separated instruments being affected differently by the (sometimes) strong cloud edge associated with the marine layer, but it seems more likely that they could be due to instrument or procedural errors. More investigation using other station records would help to better understand these differences. These problems underscore the importance of monitoring high quality hourly observations from reference stations that can be used to verify the quality of the traditional max/min temperature data from the network of stations that are employed.

Using Southern California Edison (SCE) electrical loads in the summer of 2010 as a test case, we constructed a regression between observed daily maximum and minimum temperatures in the utility service area and recorded electrical load (from OASIS) to serve as our "forecast" of electrical demand. This approach using observed historical temperatures is useful because it removes the contribution of weather forecast errors to the load forecast error, allowing us to focus on electrical demand variations that do not arise from unanticipated temperature variations.

We find that differences between actual peak load experienced in a day and that expected based on the observed temperatures are associated with changes in the inland penetration of the marine layer. The difference can be up to ~5% of the utility's peak load, which is smaller than 15% reserve margin utilities are required to maintain, but still might affect operations or reliability in some cases.

On days where the marine layer is strong and penetrates farther inland, electrical demand tends to be lower than would be expected given the day's actual temperatures. On days with a weak or missing marine layer, demand tends to be higher than expected based simply on the day's temperatures. We speculate that this could be due to the limited number of meteorological stations used to forecast electrical demand for SCE missing the effects of shifts in the inland edge of the marine layer, or possibly due to an extra increment of solar insolation that heats the interiors of structures more during clear days. An expanded analysis that attempted to find and use more meteorological data in the greater Los Angeles area would be able to discriminate between these two possibilities. Additionally, early morning satellite observations of the stratus layer might be useable for a limited, same-day prediction of the layer's effect on local electrical demand.

The present analysis was exploratory and is based only upon 2010 data—other years should be examined to determine to what extent occurs the cloud influence on Load forecast error. Besides the LA Basin region which was the domain considered in the present analysis, another avenue that could be useful would be to explore the PGE and SDGE region to determine if there is a cloud effect there. Also, the particular meteorological characteristics of yearly peak demand days could be examined differently from normal days. For example, perhaps yearly peak demand days are associated with breaks in the marine layer of multiple consecutive days. Finally, this work has considered low stratus effects on electrical load in the Los Angeles area but not other meteorological influences. For example, there are periods during which the monsoonal circulation can affect cloud cover and humidity in the region. Such flows might have different impacts from the cooler air and cloud cover associated with the marine layer, and would be useful to explore in future work.

Acknowledgements

This work was made possible by support from the California Energy Commission through agreement #500-10-041. Additional support was provided by the California Nevada Applications Program (CNAP) RISA, NOAA award NA11OAR4310150. We thank Tom Gorin and Guido Franco of the California Energy Commission for valuable comments on this work.

References

Abatzoglou J. T., 2012: Development of gridded surface meteorological data for ecological applications and modelling. Internat. J. Climatology, doi: 10.1002/joc.3413.

Garcia-Cerrutti, M., T. Gorin, C. Kavalec, and L. Marshall, 2011: Revised short-term (2011-2012) peak demand forecast, Committee final report. California Energy Commission, Electrical Supply analysis Division, report CEC-200-2011-002-CDF. 32 pp.

Iacobellis, S. F., and D. R. Cayan (2013), The variability of California summertime marine stratus: Impacts on surface air temperatures, J. Geophys. Res. Atmos., 118, 9105–9122, doi:10.1002/jgrd.50652

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

Robinson, P. J., 1997: Modeling utililty load and temperature relationships for use with long-lead forecasts. J. Appl. Meteorol., 36, 591-598.

Sweeney, J. L., 2002: The California electricity crisis. Hoover Institution Press Publication No. 513, 291 pp.

Abbreviation	Name	Pacific Gas & Electric weight	Southern California Edison weight	San Diego Gas & Electric weight
KUKI	Ukiah Municipal Airport	0.067		
KSFO	San Francisco International Airport	0.069		
KSJC	San Jose International Airport	0.282		
KSAC	Sacramento Executive Airport	0.169		
KFAT	Fresno Yosemite International Airport	0.413	0.062	
KRIV	March Air Reserve Base (Riverside)		0.371	
KBUR	Bob Hope Airport (Burbank)		0.243	
KLGB	Long Beach Airport		0.324	
KSAN	San Diego Lindbergh Field			0.333
KNKX	Miramar Marine Corps Air Station			0.333
KSEE	Gillespie Field, El Cajon			0.333

Table 1. Meteorological stations used by the California Energy Commission for forecasting electrical load for the Pacific Gas & Electric, Southern California Edison, and San Diego Gas and Electric utilities.

Utility	Tmax coeff, (std err), MW/°F	$Tmax^2$ coeff, (std err) , MW/°F^2	Tmin coeff (std err) , MW/°F	Weekend/holiday coeff (std err), MW	Percent of variance explained (R^2)
PG&E	219.5 (22.4)	N/A	71.4 (36.8)	-1141.3 (137.7)	85.6
SCE	209.6 (13.8)	N/A	203.3 (24.2)	-1860.1 (116.4)	95.4
SDG&E	-274.9 (43.4)	1.992 (0.27)	17.1 (4.7)	-206.9 (18.7)	94.5
SDG&E (no T^2)	41.5 (3.1)	N/A	30.3 (5.3)	-219.5 (23.6)	91.9

Table 2. Regression parameters for relating daily maximum load (MW) to daily maximum temperature (°F), minimum temperature (°F), and weekend/holiday status, for the summer of 2010.

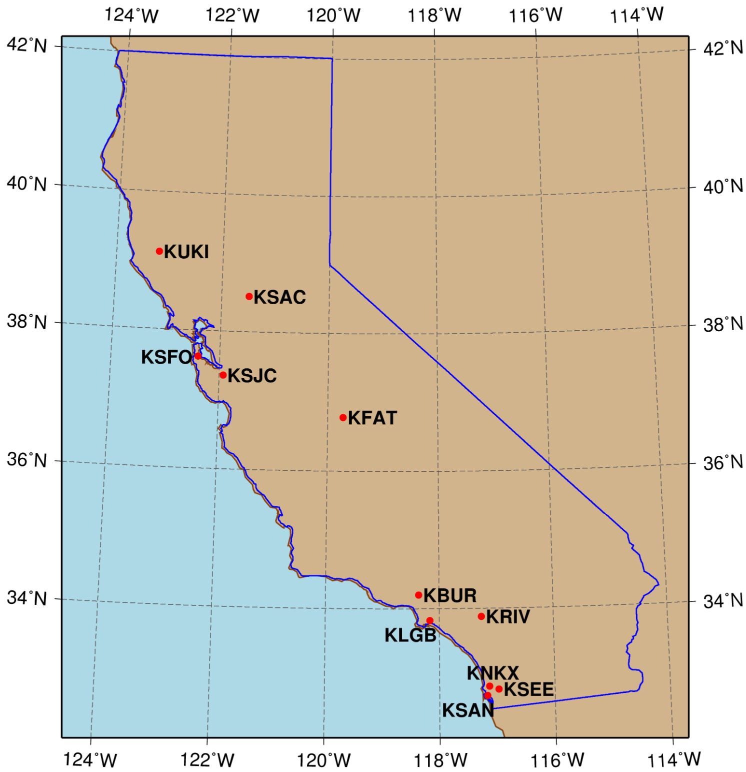


Figure 1. Meteorological stations used for utility load forecasting. See Table 1 for more information. KUKI: Ukiah. KSAC: Sacramento. KSFO: San Francisco. KSJC: San Jose. KFAT: Fresno. KBUR: Burbank. KLGB: Long Beach. KRIV: Riverside. KNKX: Miramar. KSEE: El Cajon. KSAN: San Diego.

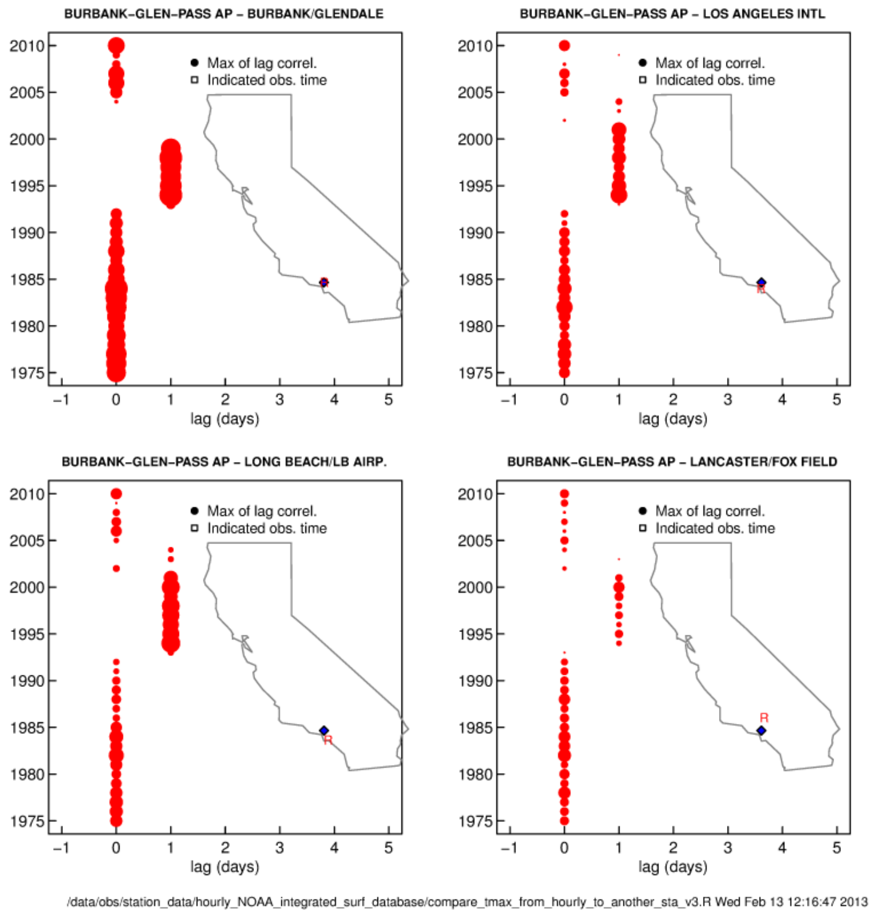


Figure 2. Red circles show, for calendar years from 1975 to 2010 (vertical axis), the lag (days, on the horizontal axis) the maximizes the correlation between two time series: a) daily maximum temperature at KBUR (Burbank airport) derived from global summary of day (GSOD) reports; and b) daily maximum temperature derived as the highest daily value observed by the indicated hourly-reporting reference station. Each panel shows a different hourly reference station used for the comparison. Top left: KBUR (the same station as used in GSOD). Top right: Los Angeles International Airport. Bottom left: KLGB (Long Beach airport). Bottom right: Lancaster/Fox Field. The inset map shows the location of KBUR (black diamond) and the reference station being used for the comparison (red “R”). The size of the red circle is proportional to the correlation between the GSOD and hourly-reporting station’s time series.

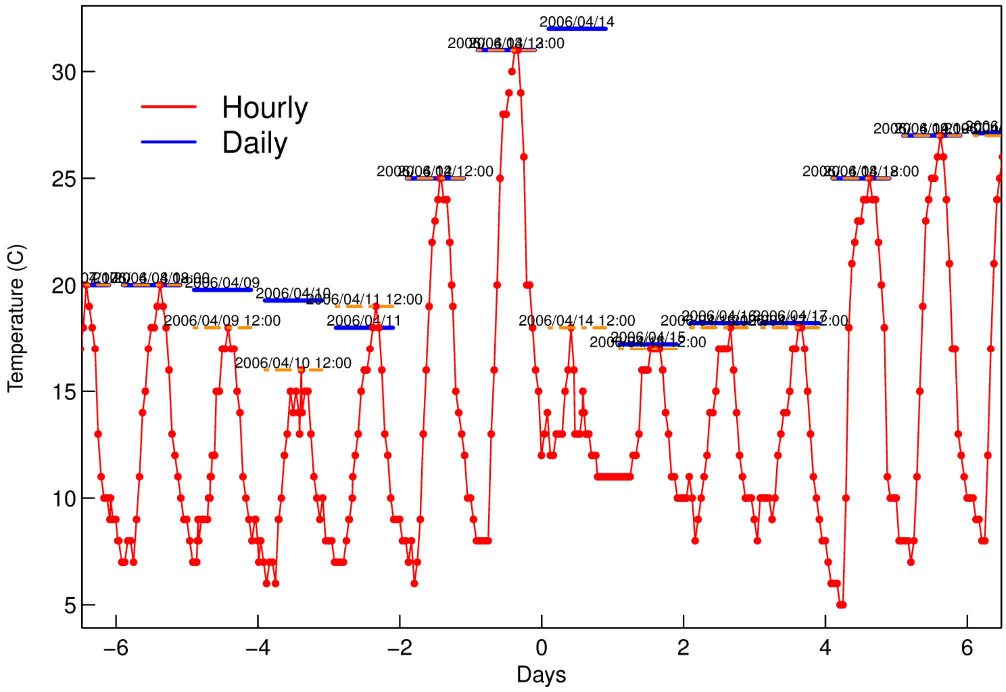


Figure 3. Time series of hourly (red line) temperature and daily maximum temperature (blue horizontal lines with the reporting date indicated) from station KSAN (San Diego Lindbergh Field), as reported in the hourly temperature data set and daily maximum temperature data set, respectively.

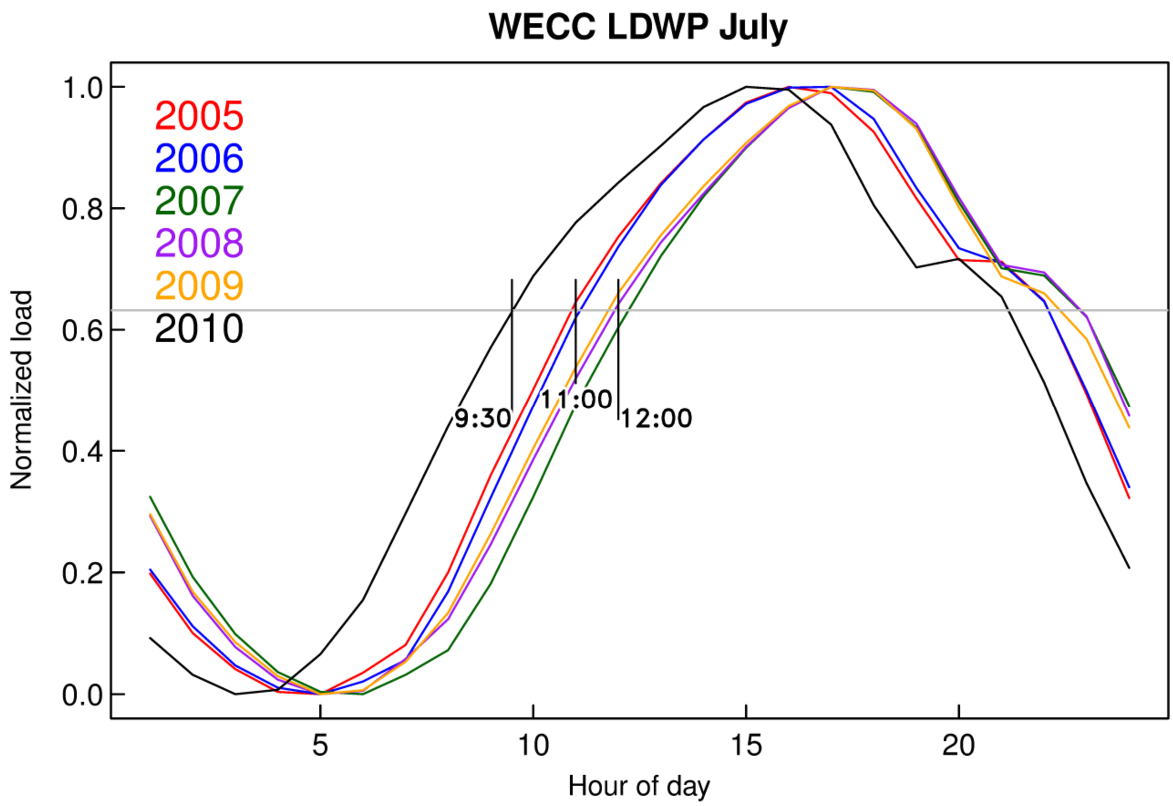


Figure 4. Normalized mean daily cycle of electrical load in the Los Angeles Department of Water and Power service region for July 2005-2010 (indicated by different colors) in the WECC data set. Also shown is the time of day at which the normalized load reaches 63% of the maximum (horizontal grey line with annotations).

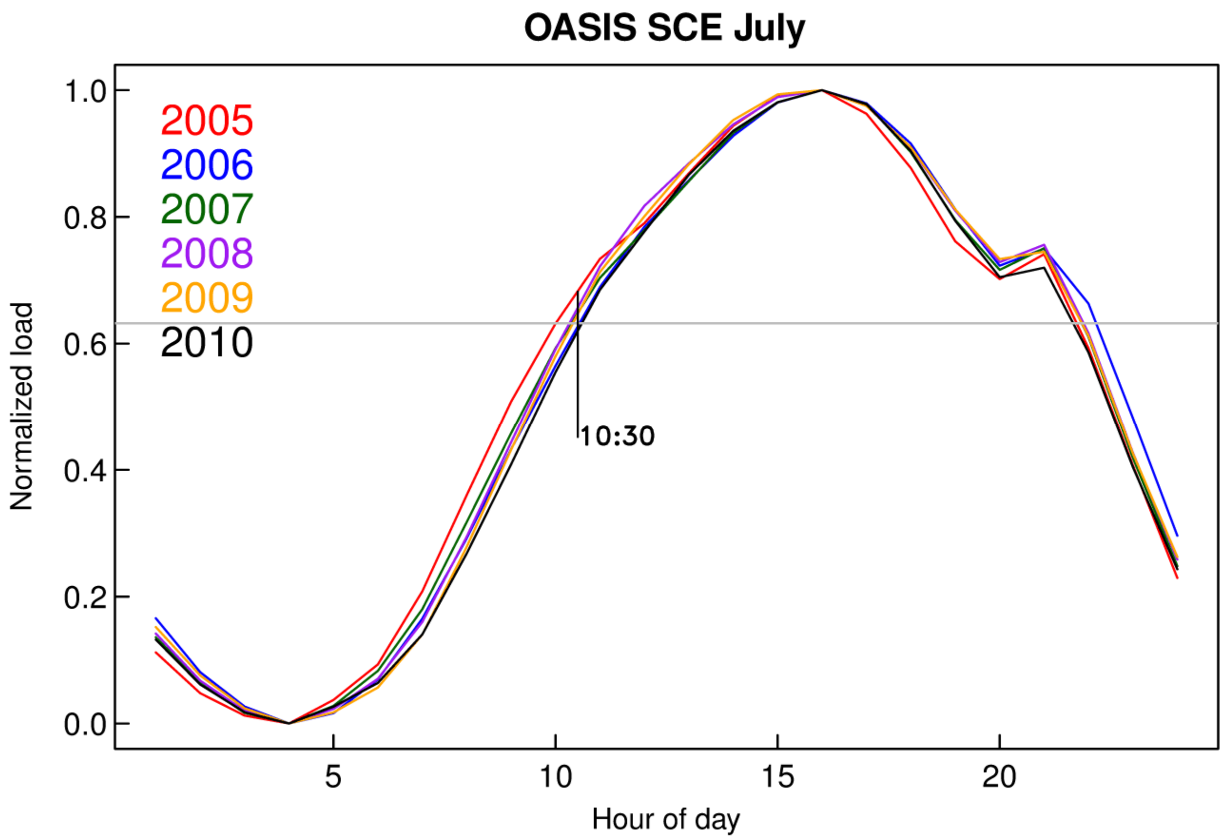
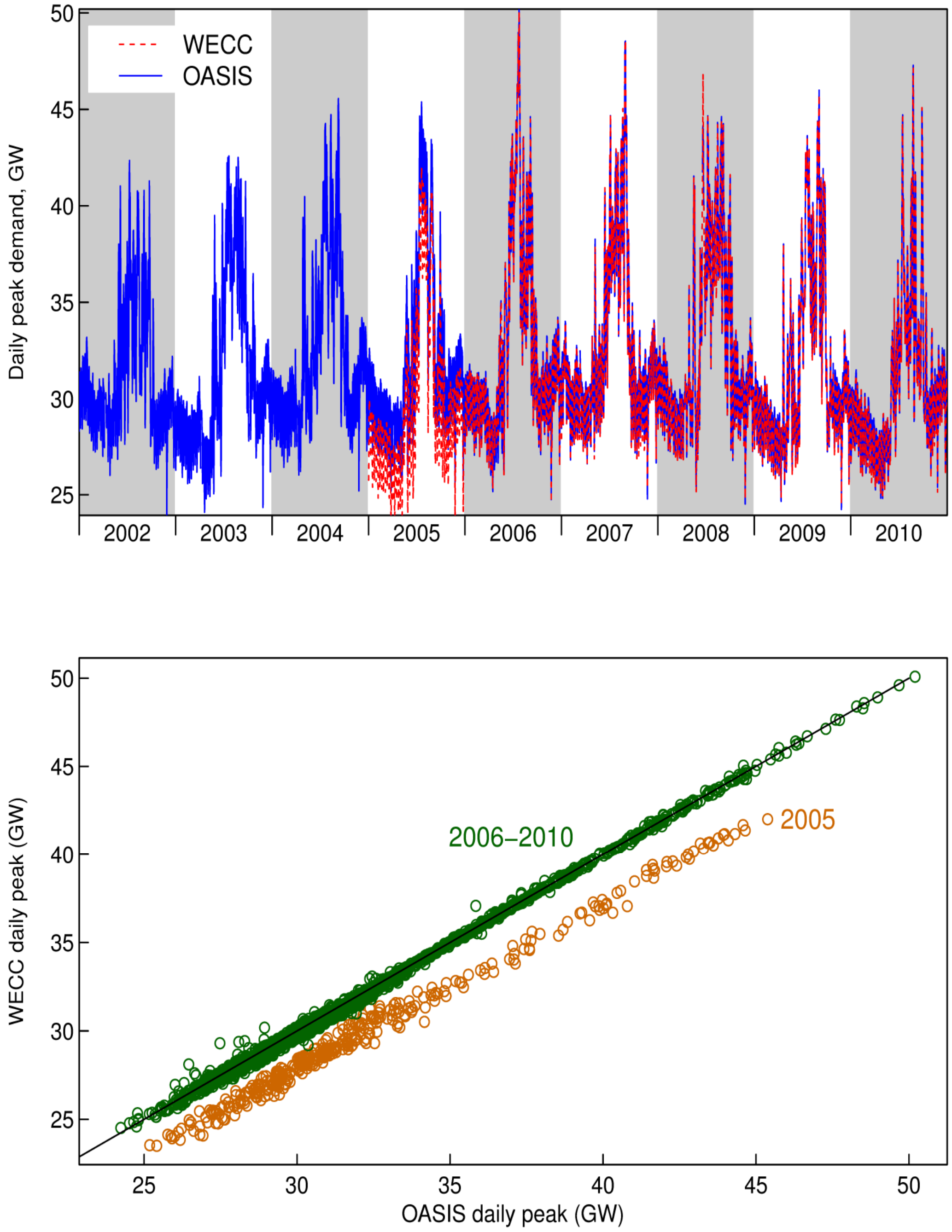


Figure 5. Normalized mean daily cycle of electrical load in the Southern California Edison service region for July 2005-2010 (indicated by different colors) in the CalISO OASIS data set. Also shown is the time of day at which the normalized load reaches 63% of the maximum (horizontal grey line with annotation).



/home/pierce/projects/cec_heatwaves/load_data_from_CalISO_OASIS/compare_daily_peak_load_OASIS_vs_WECC.R Fri Jan 18 15:35:54 2013

Figure 6. Upper: time series of CalISO service area daily peak electrical load from WECC (dashed red line) and OASIS (solid blue line). Lower: Regression between OASIS and WECC daily peak values over the period 2006-2010 (green circles) and for year 2005 (brown circles). The black line in the lower panel shows a one-to-one relationship.

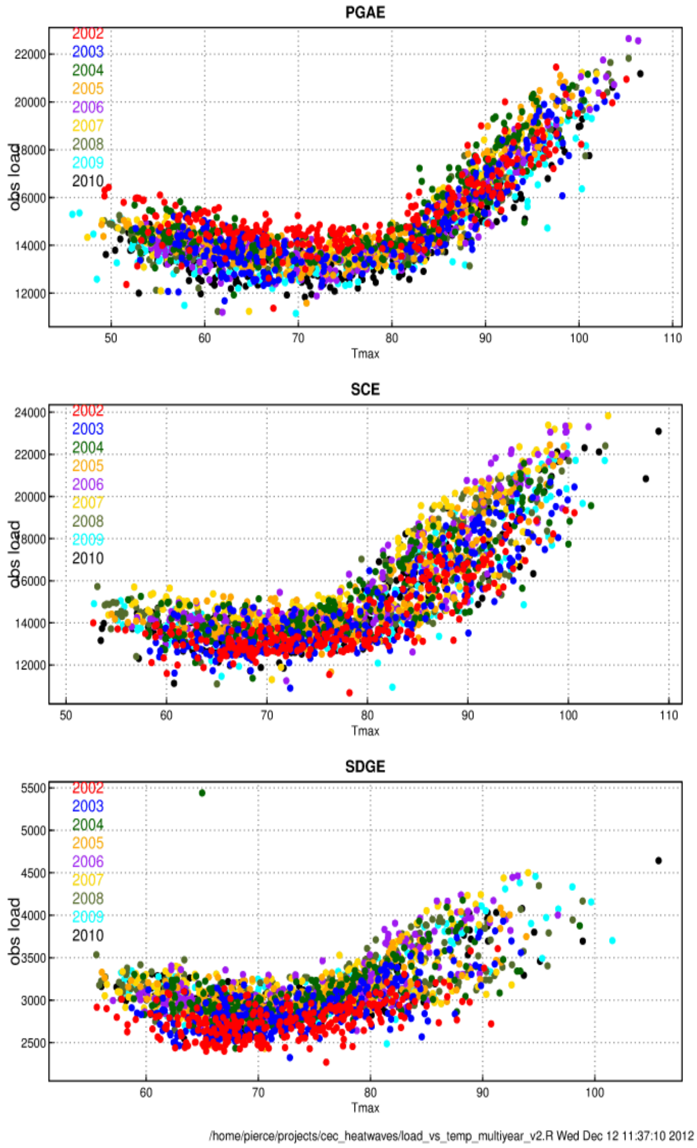


Figure 7. Relationship between utility-averaged daily maximum temperature (Tmax, horizontal axis, in degrees F) and peak daily load (vertical axis, megawatts) from CalISO OASIS for every day in the period 2002-2010 for Pacific Gas & Electric (top), Southern California Edison (middle), and San Diego Gas & Electric (bottom).

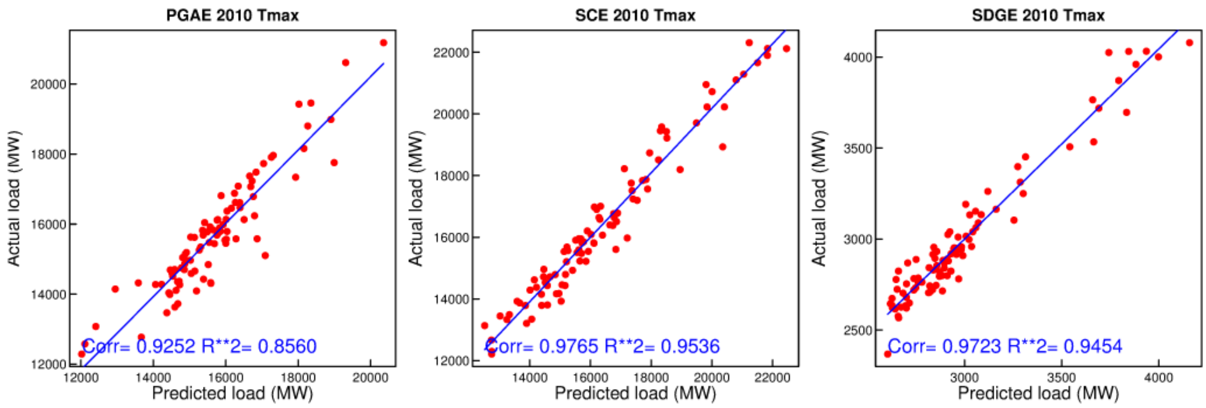


Figure 8. Regressions between actual electrical demand for summer (15 June to 15 September) of 2010 and demand predicted on the basis of Tmax, Tmin, and day-of-week status.

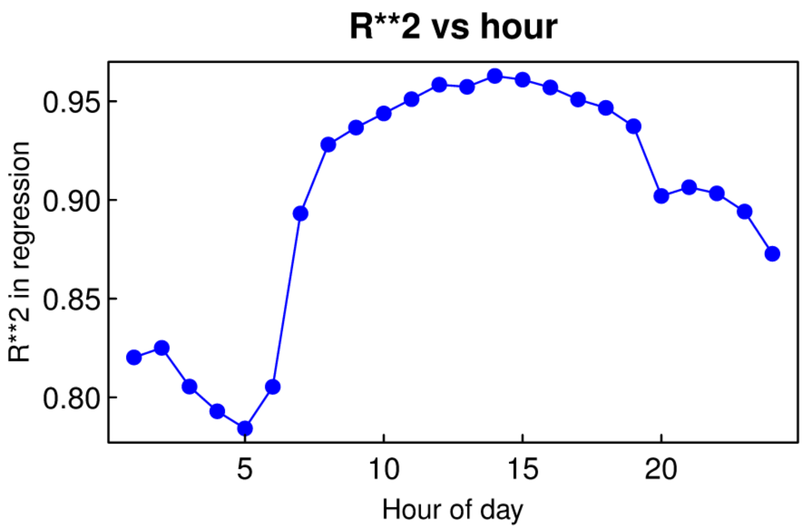


Figure 9. Fraction of variance in the electrical demand vs. temperature relationship explained when electrical demand is taken by the indicated hour (but still using daily Tmin and Tmax), for SCE over the summer of 2010.

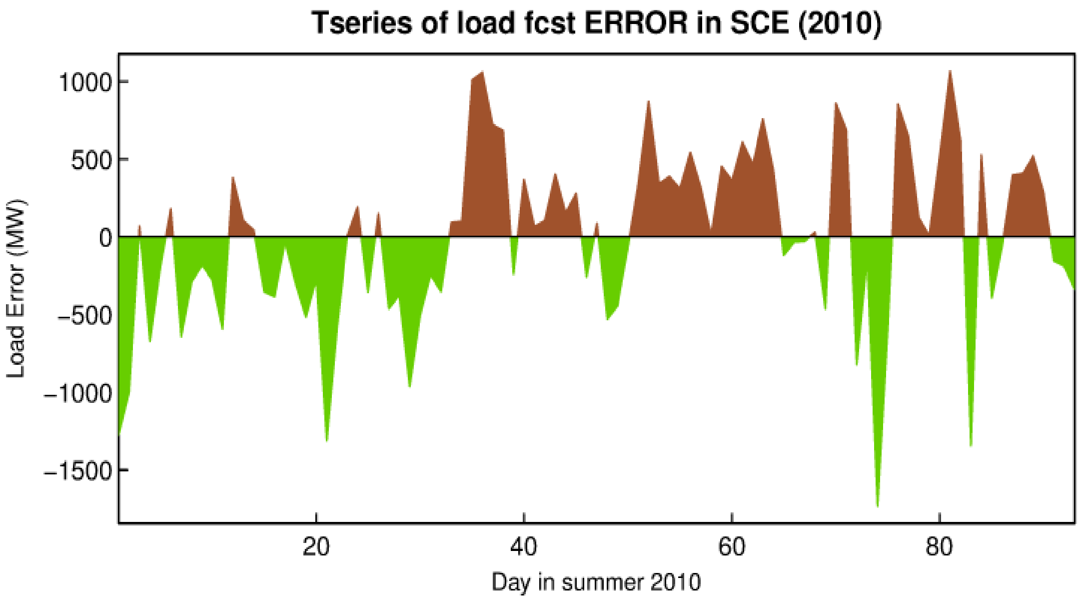


Figure 10. Load forecast error, MW (actual daily peak load – peak load predicted based on temperature and day-of-week) for SCE in the summer (15 June to 15 September) of 2010.

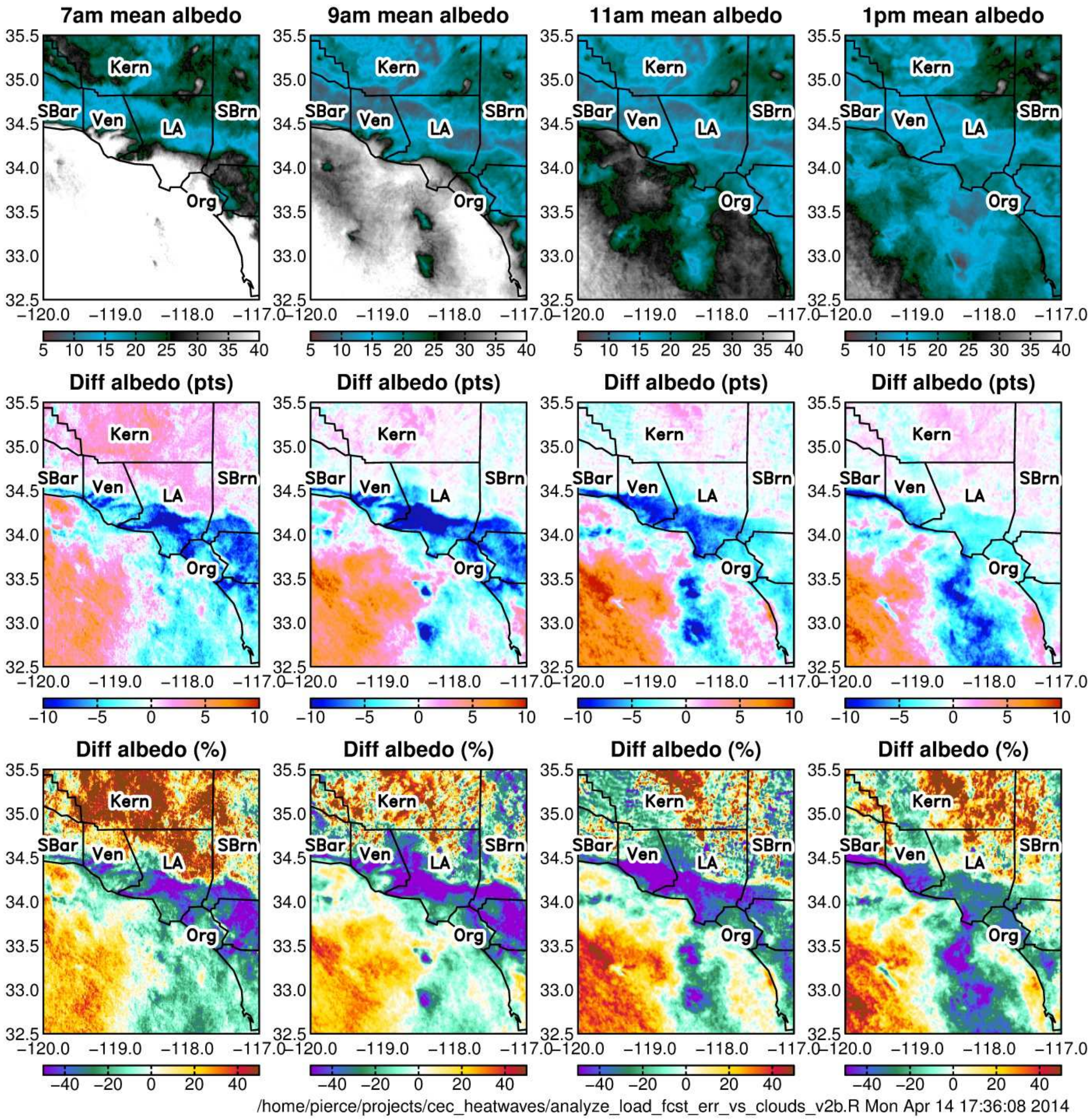


Figure 11. Top row: mean albedo (percent) in the greater Los Angeles basin area during the summer of 2010, from GOES satellite observations, for the indicated local time of day (columns). Outlines and annotations show county boundaries. Middle row: the difference (percentage points) in albedo between days in the top and bottom terciles of load forecast errors. Bottom: same as the middle row, but in percent instead of percentage points.

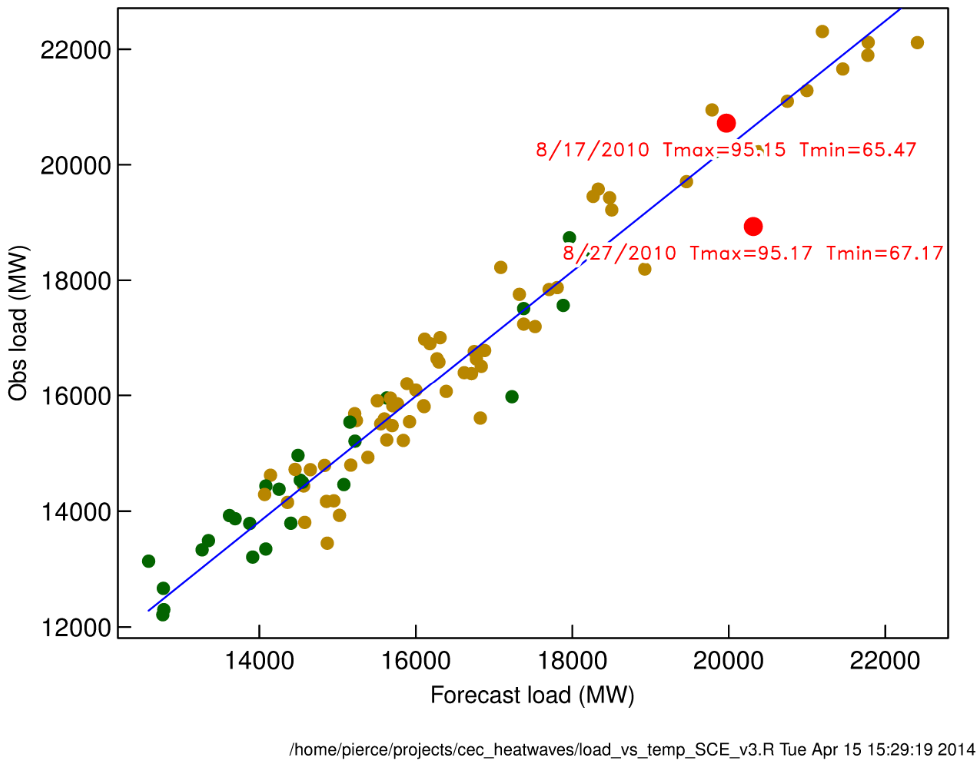


Figure 12. Scatterplot of forecast versus observed load on individual days (dots) in the summer of 2010 for SCE. Green dots show weekends or holidays; gold dots show weekdays. The two red dots show the days examined in more detail in the text and Figure 13.

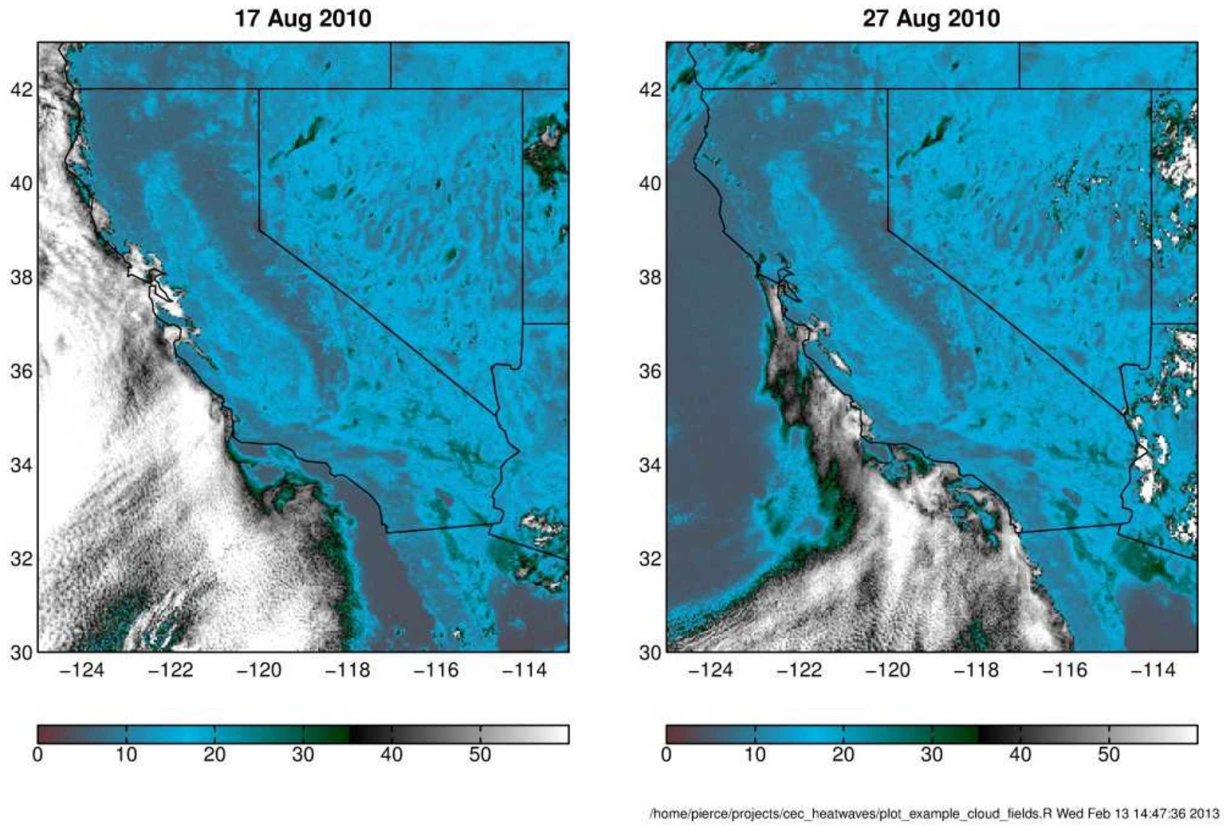


Figure 13. Cloud albedo field on August 17, 2010 (left) and August 27, 2010 (right) at 9 AM local time.