Future projections of hourly surface temperatures in California

David W. Pierce

Daniel R. Cayan

Division of Climate, Atmospheric Sciences, and Physical Oceanography
Scripps Institution of Oceanography
La Jolla, CA 92093-0230

dpierce@ucsd.edu / dcayan@ucsd.edu

24 May 2019

A report prepared in partial fulfillment of a grant by the California Energy Commission

Abstract

An analog matching method is developed to produce projections to the end of this century of hourly temperatures at selected stations in California using daily minimum and maximum temperature (Tmin and Tmax). The historical record is first examined for the observed 3-day sequence of Tmin and Tmax that best matches the 3-day sequence of model Tmin and Tmax centered on the model day being processed. Hourly temperature values from the best-matching analog day are used to construct the final hourly model values by scaling the analog days' values to match Tmin and Tmax of the model day being processed. Results from 10 general circulation models that do well simulating the regional climate show an extensive decrease in the number of chill hours in the state, which could have significant impacts on agriculture. Most locations also show a strong increase in the number of hours above thresholds of 100 and 110 °F, sometimes by more than an order of magnitude. Such increases would drive a strong increase in cooling electricity demand. Future changes in the shape of the diurnal cycle are seen, particularly in autumn along the Southern California coast, but the magnitude of the changes are small (~2%). An analysis of the cross-validated historical observations suggests that the method might underestimate the

size of projected changes in the diurnal cycle, which would require more frequent saving of temperature than the twice-daily Tmin and Tmax to resolve.

1. Introduction

Future changes in hourly near-surface air temperatures due to human-caused climate change will affect important aspects of our society and economy, such as agriculture, energy demand, and human health. Despite their importance, hourly temperature changes have not been examined as much as other less complete measures, such as daily minimum and maximum temperature (Tmin and Tmax). A key reason for this neglect is that the majority of global climate models (GCMs) participating in the Climate Model Intercomparison Project version 5 (CMIP5), the primary source of GCM projections, did not save hourly temperatures over an appreciable time span. Examining future changes in hourly temperatures therefore requires a method to infer hourly information from data saved by the GCMs.

The purpose of this project is to generate future projections of hourly temperature over California at key selected meteorological stations that are used by the energy sector for understanding and anticipating the state's energy demand. The basis for the temperature projections is the Localized Constructed Analog (LOCA) statistically downscaled dataset (Pierce et al. 2014), a gridded product that provides daily Tmin and Tmax from 32 global climate models (GCMs) at a spatial resolution of 6 km. The LOCA data cover the CMIP5 historical period from 1950 to 2005, and include medium and high future emissions scenarios over the period 2006 to 2100. We focus here on the 10 LOCA-downscaled GCMs identified by the California Department of Water Resources climate change technical advisory group (DWR-CCTAG, 2015) as being well suited for climate analysis in the region, based on an evaluation of their simulated local climate and variability as well as teleconnections the tropical Pacific Ocean region.

The fundamental issue addressed in this work is how to transform the twice-daily (Tmin, Tmax) LOCA temperature projections into hourly values. This process is referred to as "disaggregating" Tmin and Tmax into hourly values.

A typical approach for hourly temperature disaggregation is to use historical hourly data to create a climatological curve of hourly diurnal temperature variation, often broken out by month or season, and adjust the amplitude of the climatological daily curve to match the provided LOCA daily Tmin and Tmax values. This climatological curve approach will be used as here as a benchmark for comparison and evaluation. The drawback to the climatological curve approach is that it reduces variability in the hourly results, since a climatological sequence of hourly values is fitted at each day.

In this work we use an analog matching approach inspired by the Localized Constructed Analogs (LOCA; Pierce et al. 2014) technique to better depict variability in the final hourly disaggregated product. The process is described in detail in section 3. Briefly, to obtain the hourly time sequence for a particular day, we examine the historical record for the observed sequence of three days (termed analog days) that best matches the 3-day sequence of model days centered on the model day being disaggregated. We then use hourly information from the matching analog days to construct the hourly disaggregation, scaling the analog days' values to match the model day being disaggregated. Full details are given below.

This report is organized as follows. In section 2, the observed hourly data and its quality control procedures is described. In section 3, the hourly disaggregation process is given, and results compared to the climatological curve approach. In section 4, some analysis of future changes in hourly temperature data and relevant applications are outlined. Results are discussed in section 5, and conclusions are given in section 6.

2. Observed hourly temperature data and quality control process

Observed hourly near-surface air temperature data was obtained from the California Energy Commission (CEC) for 29 meteorological stations, and will be referred to as the CEC data. Station locations are listed in Table 1 and plotted in Figure 1. Geographical coverage is patchy, with better sampling of Southern California, the Central Valley, and the San Francisco Bay Area, and little to no coverage in far Northern California, Central California, the inland desert regions, and east of the Sierra Nevada. All of the stations have data covering the period 1 Jan 2000 through 31 Dec 2018, with the exception of KSEE, which starts 17 Feb 2010. To extend the coverage of KSEE to match the other stations, it was filled with ASOS data, which is described in more detail below.

Table 1. Hourly temperature stations used in this analysis

Call sign	Latitude	Longitude	State	Name
KBFL	35.4344	-119.0542	CA	Bakersfield Meadows Field
KBUR	34.2006	-118.3575	CA	Burbank-Glendale-Pasadena Airport
KFAT	36.78	-119.7194	CA	Fresno Yosemite International Airport
KIPL	32.8342	-115.5786	CA	Imperial County Airport
KLGB	33.8117	-118.1464	CA	Long Beach Daugherty Field
KLAX	33.9381	-118.3889	CA	Los Angeles International Airport
KMCE	37.2847	-120.5128	CA	Merced Municipal Airport
KMOD	37.6242	-120.9506	CA	Modesto City-county Airport
KOAK	37.7214	-122.2208	CA	Oakland Metro International Airport
KRBL	40.1519	-122.2536	CA	Red Bluff Municipal Airport
KRAL	33.9519	-117.4386	CA	Riverside Municipal Airport
KSAC	38.5069	-121.495	CA	Sacramento Executive Airport
KSAN	32.7336	-117.1831	CA	San Diego Lindbergh Field
KNKX	32.8667	-117.1333	CA	San Diego Miramar Wscmo
KSFO	37.6581	-122.4378	CA	San Francisco International Airport
KSJC	37.3592	-121.9239	CA	San Jose International Airport
KSBA	34.4258	-119.8425	CA	Santa Barbara Municipal Airport
KWJF	34.7411	-118.2117	CA	Lancaster William J Fox Field
KPSP	33.8281	-116.5053	CA	Palm Springs Regional Airport
KCQT	34.0217	-118.2914	CA	Los Angeles Downtown Usc Campus
KBLH	33.6186	-114.7142	CA	Blythe Asos
KEED	34.7675	-114.6189	CA	Needles Airport
KOXR	34.2008	-119.2069	CA	Oxnard Ventura County Airport
KSBP	35.2372	-120.6414	CA	San Luis Obispo Airport

KUKI	39.1258	-123.2008	CA	Ukiah Municipal Airport
KTRM	33.6333	-116.1667	CA	Thermal / Palm Springs
KSNA	33.68	-117.8664	CA	Santa Ana John Wayne Airport
KSEE	32.8167	-116.9667	CA	Gillespie Field / El Cajon
KLAS	36.0719	-115.1633	NV	Las Vegas McCarran +International Ap

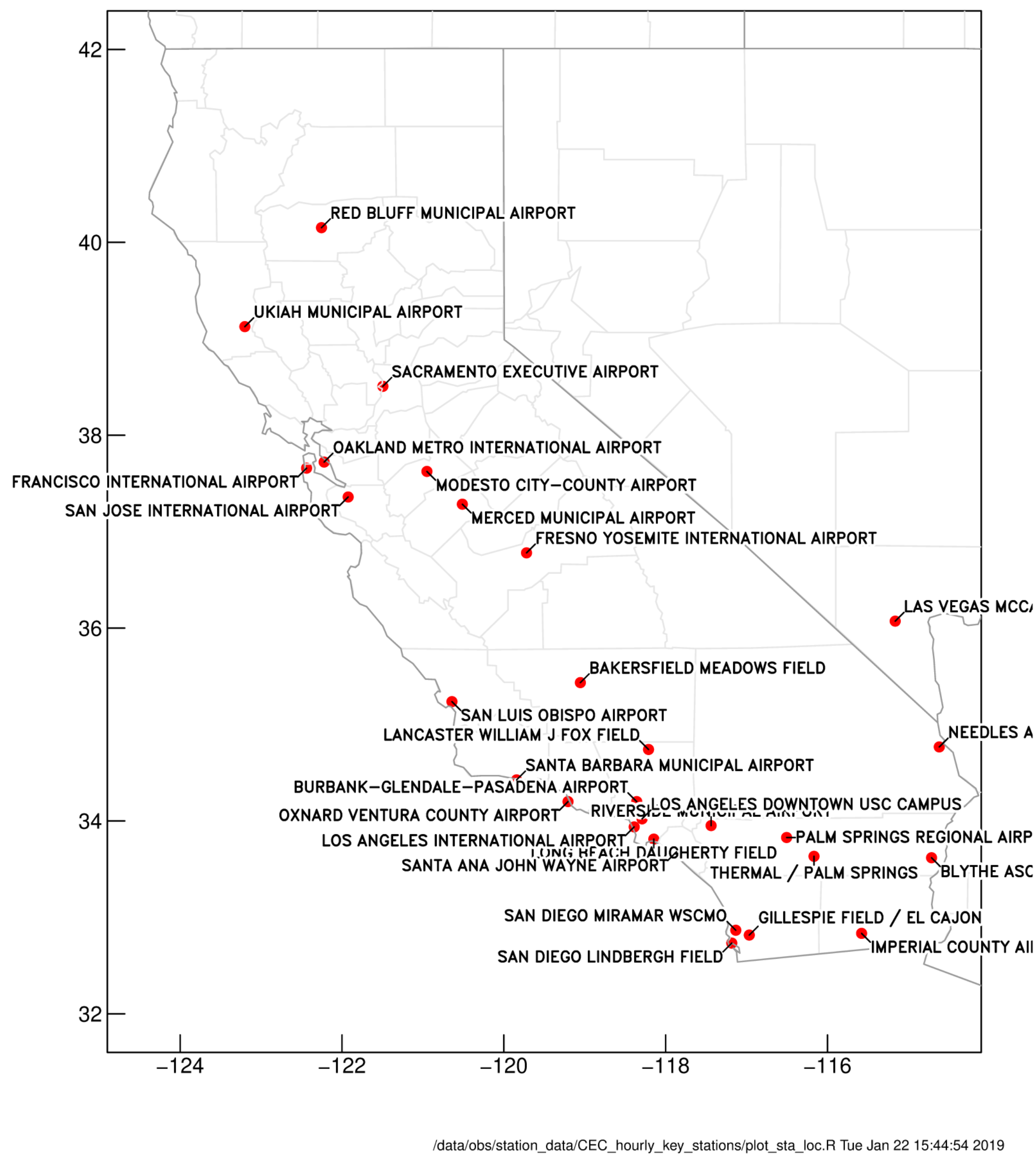

Figure 1. Station locations providing hourly near-surface air temperature data for this study.

2.1 Quality control

Although the data are serially complete, inspection showed data quality problems, with behavior such as unphysical extreme outliers, sequences of repeated bad values, and evidence of prior infilling and interpolation contaminated by interpolating across bad values. While only a small fraction of the data requires quality control (on the order of 0.01%), the LOCA historical analog matching approach needs as clean a historical data set as possible to find matching analogs. Questionable values in the training data set could be conveyed by the historical analog downscaling process into the final result. This motivates spending time on the quality control even though bad values are a small fraction of the data.

Quality control consisted of eliminating extreme outliers, comparing the CEC data to independently downloaded ASOS data, identifying strings of bad values, and eliminating hourly values that represented an unrealistically large deviation from hourly values immediately before or after. These issues will now be discussed in turn.

2.2 Extreme outliers

Hourly temperature data has both an annual and diurnal cycle. So, for example, an extreme warm temperature for midnight in December is a very different value than for 3 PM in July. Accordingly, for this step of the quality control process the data were first converted to anomalies by removing the best-fit annual and semi-annual harmonics. This was done separately for each hour of the day. The data for each hour were then standardized by dividing by the standard deviation (sigma) of that hours' anomalous values. This yields a time series of hourly standardized anomalies.

To account for meteorological phenomena such as heat waves or cold spells, which can produce extreme values but are characterized by extending over several days, the data were high pass filtered using a boxcar filter with a width of 3 days. This prevents, for example, flagging extreme high temperatures that occur during a heat wave, even though the same temperature value would be flagged if it was found during a cold spell.

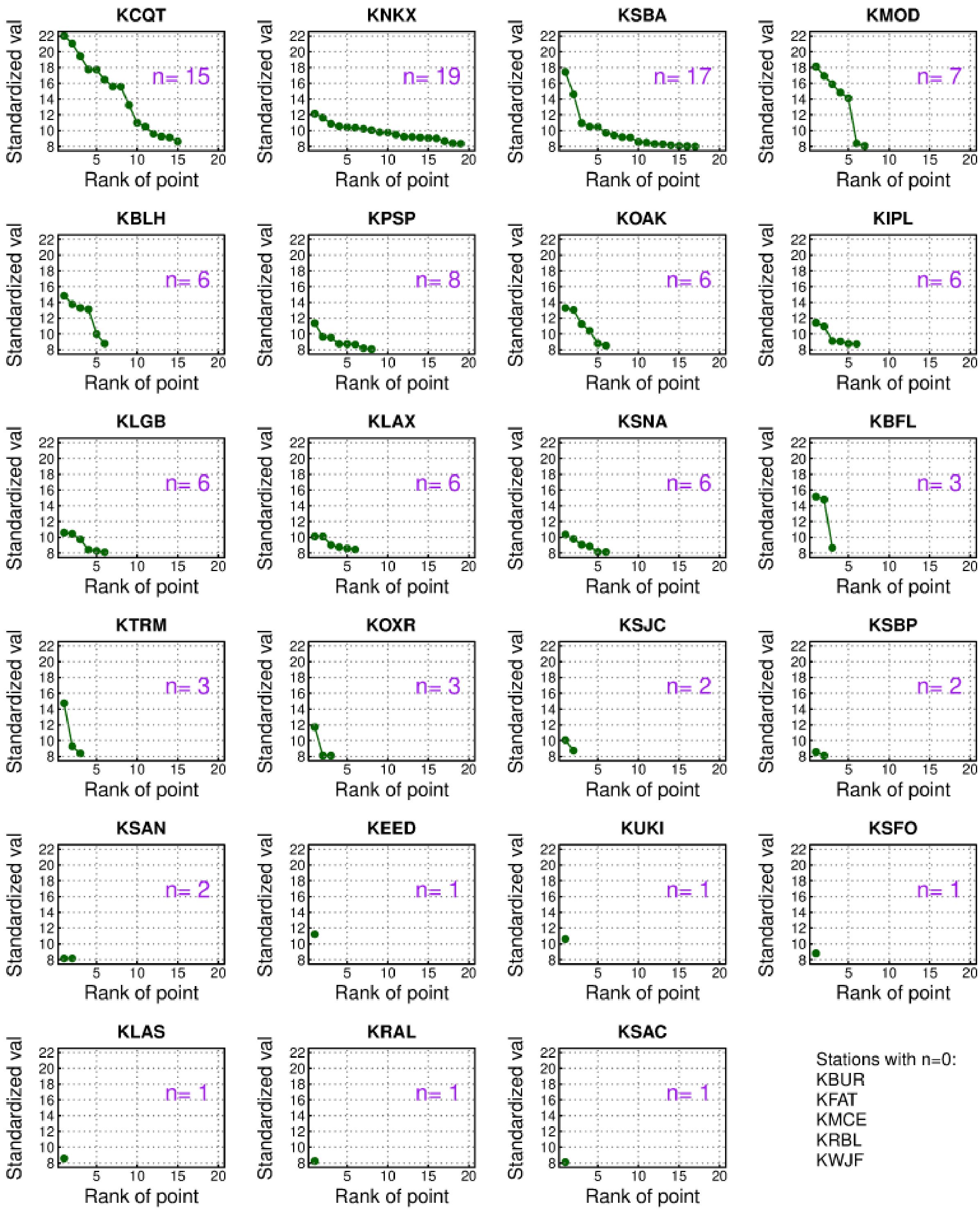

Figure 2. Census of extreme (>= 8 sigma) temperature values by station. The X axis shows the rank of the plotted point (1=highest ranking point), and the Y axis shows the standardized anomaly for that point. Values are calculated on the basis of standardized hourly anomalies that have been high pass filtered as described in the text. The total number of points with sigma >= 8 is shown in purple. Panels are ordered by the sum of the plotted points.

A census of extreme values is shown in Figure 2. The number of extreme values varies considerably by station, to such a degree that the variation is likely more indicative of station quality than of meteorology. So, for example, KCQT, KNKX, and KSBA all have a large amount of bad data, with at least 15 points with sigma >= 8, while many of the major airport stations, such as KLAS, KSFO, and KSAN have 2 or fewer.

Selecting the threshold value for flagging erroneous points is complicated by the fact that a few truly extreme outliers exist in the historical record. For example, on 6 July 2018, Santa Barbara experienced evening (7-11 PM) temperatures close to 100 F, which represent a sigma value of nearly 11, considerably more extreme than many obvious errors in the data. To address this, the extreme threshold was set at 8 and the individual flagged values from each station were manually examined, along with consulting of historical meteorological and news records for the period in question to see if extreme temperatures were prevalent near the station. Data values associated with verifiable extreme temperature periods, such as the Santa Barbara warm event noted above, were manually added back in to the data set. The stations that had points manually added back this way are: KIPL (1 point); KLAX (3); KSAN (1); KSBA (6); KCQT (2); KOXR (2); KSNA (2); and KSFO (1).

Examples of extreme outliers eliminated by this process are shown in Figure 3. Note that numerous unphysical values do not reach the 8-sigma threshold, but those are eliminated later in the quality control process, as described below.

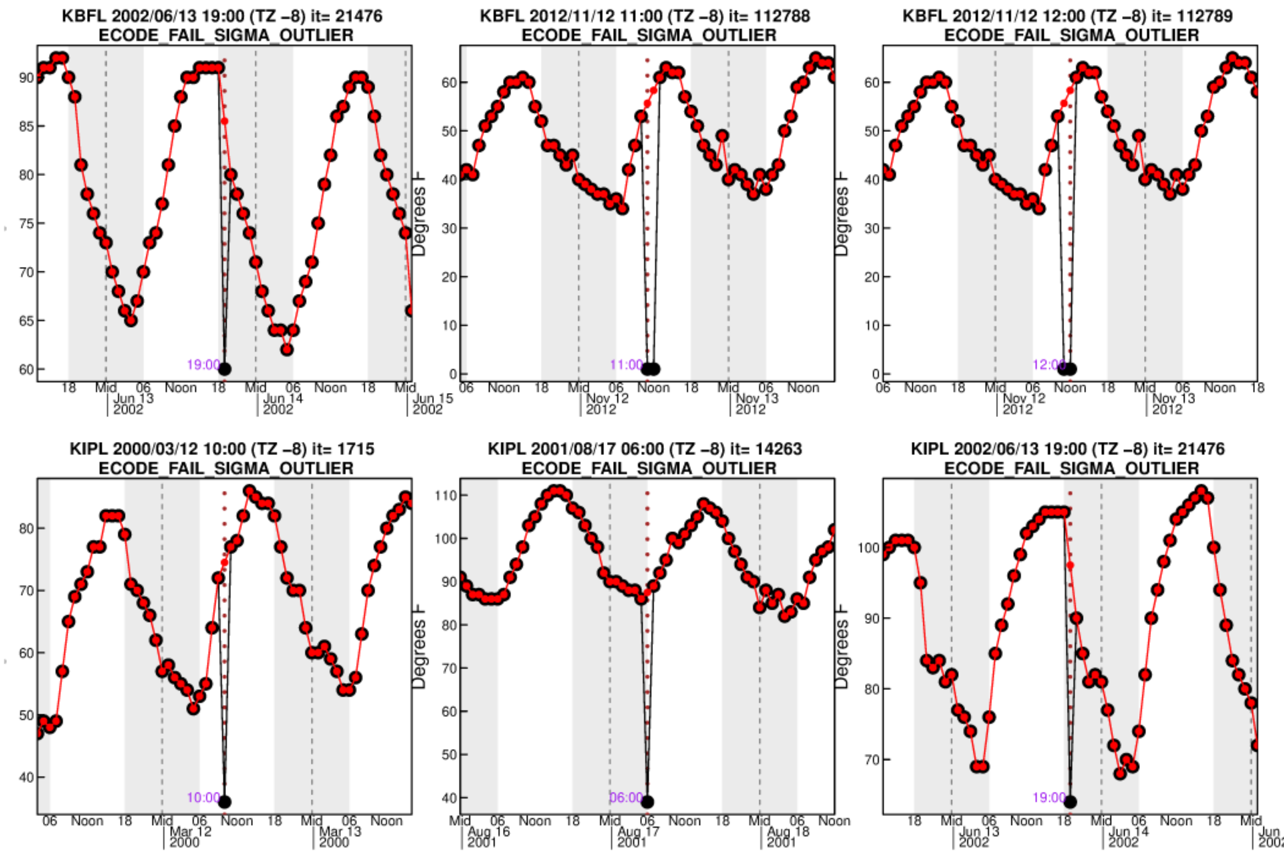

Figure 3. Examples of extreme temperature outliers (degrees F), defined as greater than or equal to 8 standard deviations of the hourly anomalous temperature values. Black dots show the original CEC hourly data, while red dots show the final quality-controlled values. In each panel, the identified outlier is in the middle of the panel and indicated by the vertical brown dotted line. The purple text near the outlier shows the hour of day (PST) that the outlier was experienced. The station is indicted in the panel title.

2.3 ASOS data

Automated Surface Observing System (ASOS) meteorological data is available for nearly 1000 sites across the U.S. Temperature is generally reported at 1-hour intervals, but stations can have gaps in their time series, and sometimes (particularly when conditions are changing rapidly) the data may be reported more frequently than hourly. ASOS data for each CEC station was downloaded from the Iowa State University Iowa Environmental Mesonet (https://mesonet.agron.iastate.edu/request/download.phtml), which, despite its name, makes available ASOS data from all states. Since the ASOS data reporting does not necessarily occur on the hour, the

ASOS data were interpolated to values on the hour by using the mean of a linear and cubic spline interpolation.

In the large majority of points (typically > 96% of the time at the worst stations, and > 99.7% of the time at the best stations) the ASOS and CEC data agree to within 2 F. However, disagreements can in some circumstances be large, and in at least a few instances the records systematically differ, with the ASOS and CEC time series showing different temperature values over a period of a few days or more.

To merge the ASOS and CEC data, all sequences of 1 or more hourly values where the two data sources differ by 2 F or more were identified. Statistics of how often this occurred are shown in Table 2. The station with the most disagreements was KCQT, where over 4% of the CEC hourly values differed from the ASOS hourly values. By contrast, at many of the largest airports included in this study (including KSFO, KLAX, KLAS, KSAN, and KSJC), which are likely to have the highest quality data, less than 0.2% of the hourly data points differ between the two sources.

Table 2. Statistics of the number of hourly points which differed by more than 2 F between the ASOS and CEC data, the percent of total number of points that differed, the number of differing points where the CEC value was retained, and the number of differing points where the ASOS value was retained. Stations are sorted from greatest to least number of points that differ between CEC and ASOS values.

STATION	# DIFF PTS	% DIFF PTS	# KEPT CEC	# KEPT ASOS
KCQT	6972	4.19	3871	3101
KTRM	6026	3.62	3292	2734
KIPL	4867	2.92	2731	2136
KMCE	4018	2.41	2653	1365
KOXR	2590	1.55	532	2058
KRAL	2349	1.41	1403	946
KPSP	1837	1.10	169	1668
KUKI	1388	0.83	287	1101
KEED	1361	0.82	619	742
KWJF	1253	0.75	572	681
KBLH	804	0.48	273	531
KBFL	767	0.46	200	567
KRBL	757	0.45	272	485
KNKX	733	0.44	356	377
KSBA	689	0.41	260	429
KSBP	669	0.40	207	462
KMOD	621	0.37	224	397
KBUR	561	0.34	209	352
KSAC	479	0.29	143	336

KSNA	413	0.25	138	275
KLGB	379	0.23	181	198
KOAK	331	0.20	142	189
KFAT	318	0.19	80	238
KSJC	314	0.19	86	228
KSAN	293	0.18	137	156
KLAS	268	0.16	96	172
KLAX	266	0.16	91	175
KSFO	260	0.16	85	175

In order to determine which values to retain (ASOS or CEC) when they differed, we used an a-priori expectation that the more plausible data sequence should have the least jumps or discontinuities over the period starting 1 hour before the ASOS-CEC difference period begins and ending 1 hour after the ASOS-CEC difference ends. I.e., if the ASOS-CEC difference exceeds 2 F over *N* consecutive hourly points, we consider a segment of *N+2* points. For this segment a simple measurement of "jumpiness" was adopted as the sum of the squared difference between each point and its successor, where both the point and its successor must be in the interval. Because the measure uses the square of the consecutive differences, the sign of each pair's difference does not matter, and large discontinuities are highly penalized. The sequence with the lowest value of this metric – either the CEC or ASOS sequence – was used as the final data for the interval. Examples of periods when the CEC and ASOS data differ, and the final selected values to use based on the jumpiness index, are shown in Figure 4. Panel e is an instance where the two differ over an extended period and there is no obvious way to choose which data set is correct, but the algorithm chooses the one with less variability, which is the CEC sequence in this case. It is interesting that 5 out of 6 of the worst quality stations (measured by the number of ASOS-CEC disagreements) ended up keeping more CEC data than ASOS data, while the rest of the stations kept more ASOS data than CEC data.

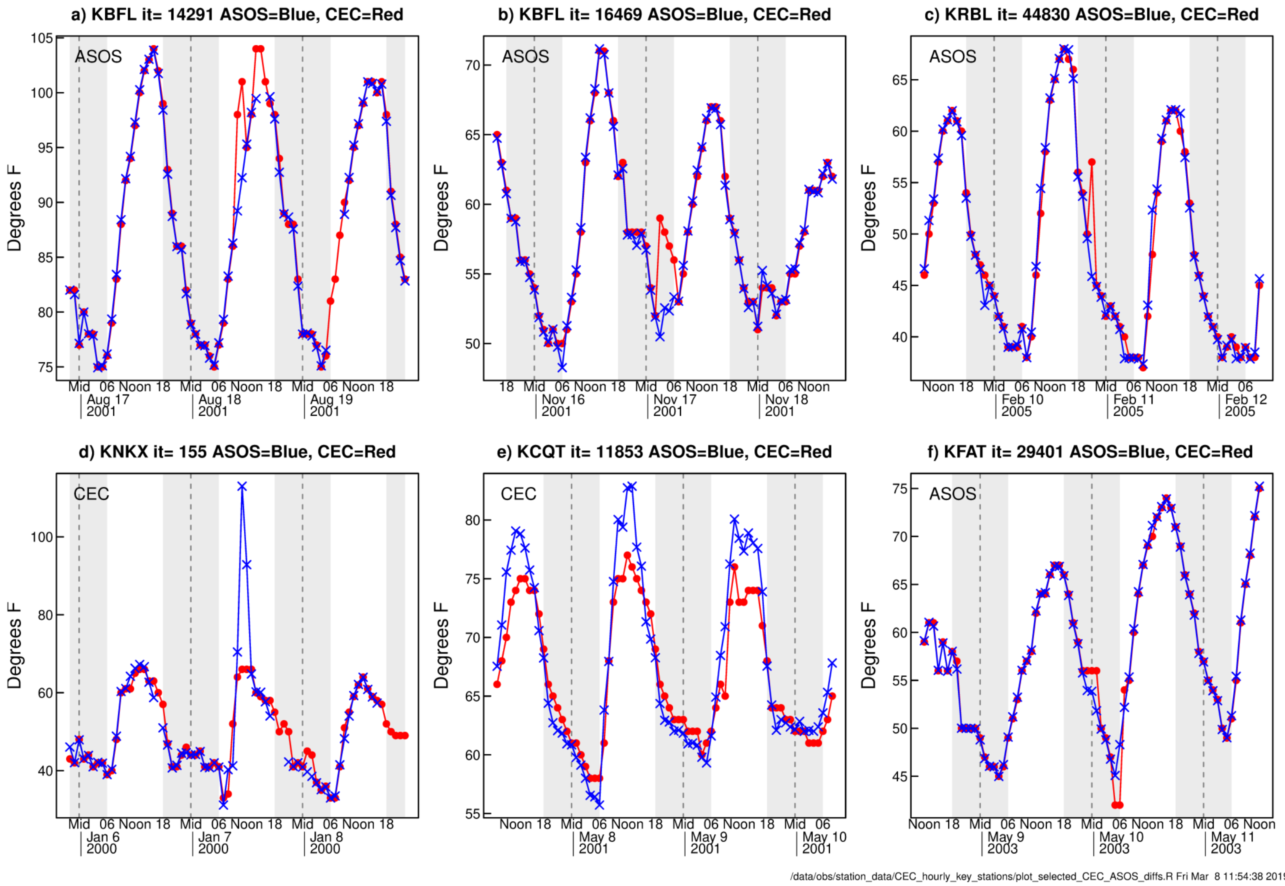

Figure 4. Examples showing periods when the CEC data (red) differs by more than 2 F from the ASOS data (blue). The time span when the two differ is centered in each panel. The notation in the upper left of each panel, either “ASOS” or “CEC”, shows which data set was ultimately used over the period where the two differed.

2.4 Sequences of bad values

There are various periods in the data that show sequences of unrealistic values. For example, in the original CEC data set, KMOD on 2013/11/28 from 10 AM to 3 PM, and KCQT on 2005/10/18 from 1 AM to 10 AM (Figure 5). These intervals are important for two reasons. First, they demonstrate unambiguously that the original CEC data has quality control issues that need to be addressed. Second, the sequence at KMOD (Figure 5, upper panel; point circled in purple) shows that a linear interpolation has been previously performed on the data, and that it included the bad data as one of the endpoints. This can be seen by the point at a value of 34 F, halfway between the bad value and the subsequent correct value. This unfortunately makes the quality control procedure more difficult, as such interpolated values are less extreme outliers than the original erroneous data and so harder to identify.

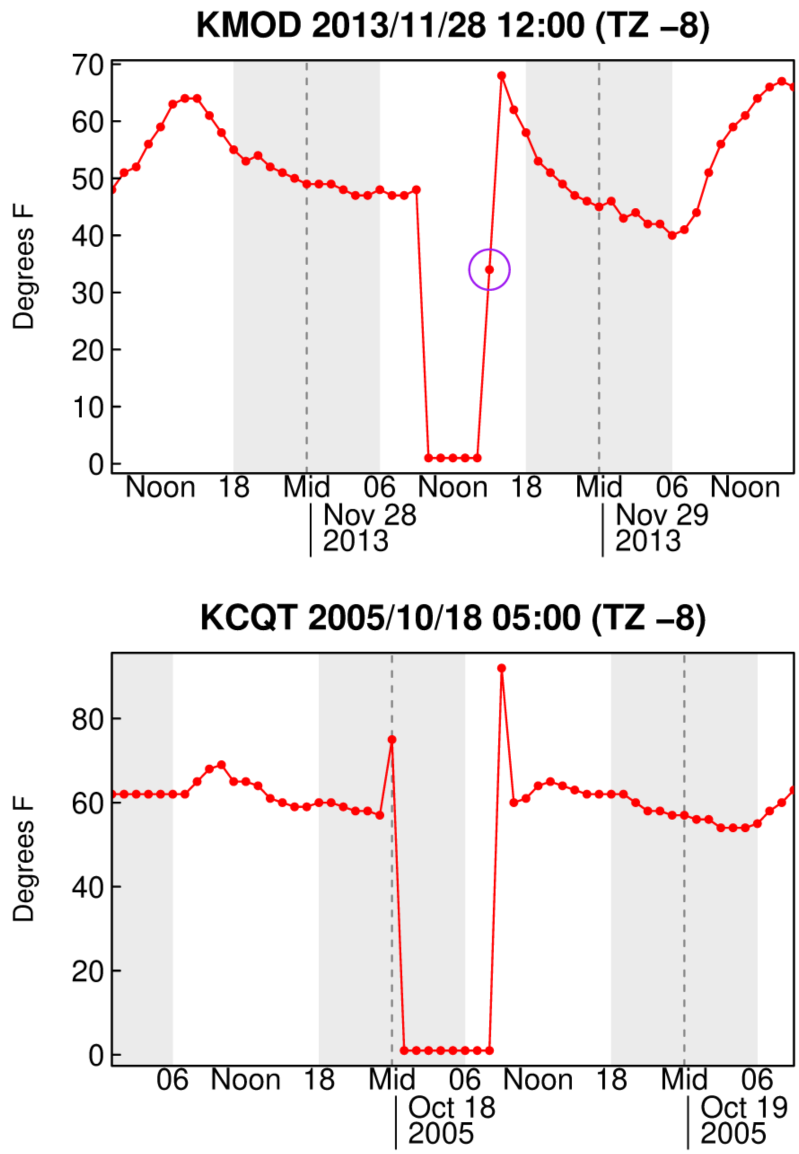

Figure 5. Sequences of bad hourly temperature values found at the indicated stations and dates/times. The point circled in purple in the upper panel strongly suggests that the data have been linearly interpolated between the bad zero values and the subsequent valid value.

Most sequences of bad values are eliminated by infilling with the ASOS data. Examples of other sequences that are eliminated in the quality control process are shown in Figure 6.

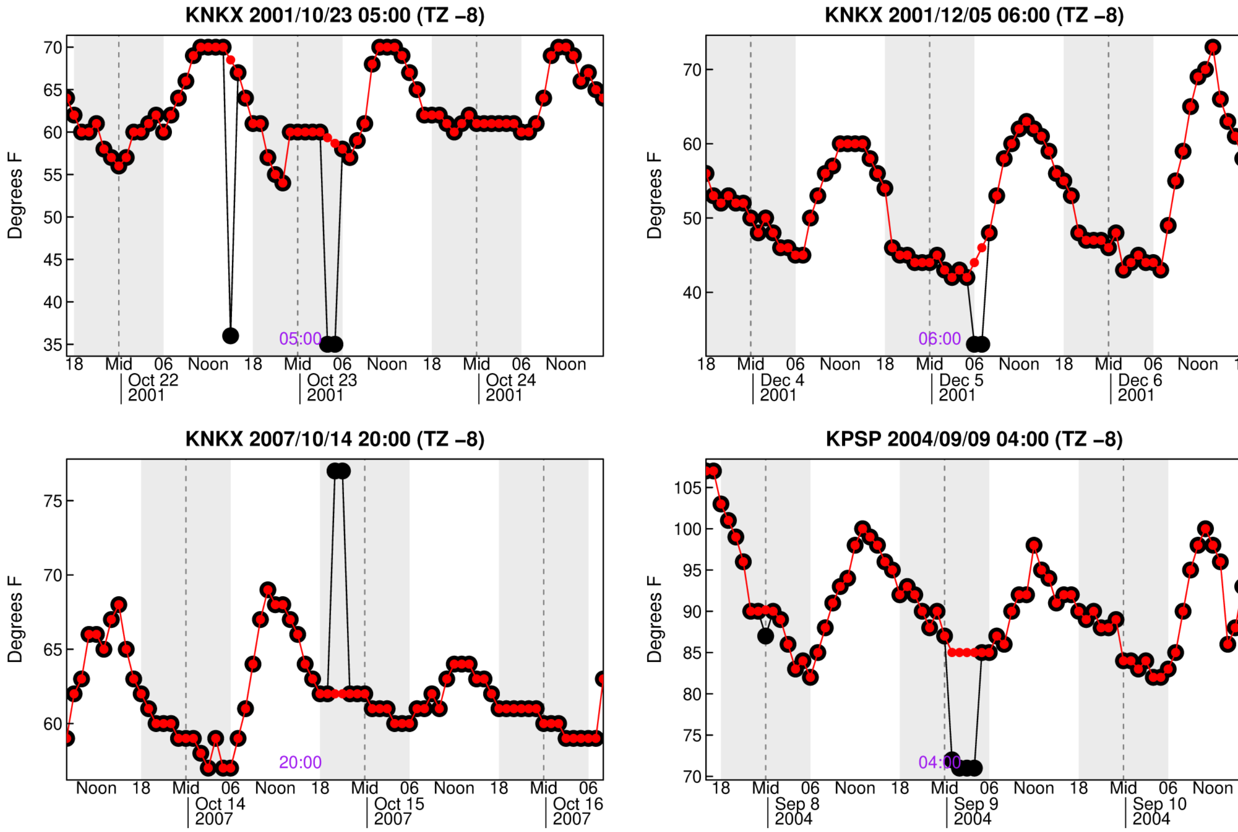

Figure 6. Illustration of sequences (n >= 2) of bad hourly temperature values. In each panel, the black dots are the original CEC data, and the red dots are the data after errors have been identified and filled by linear interpolation. The beginning hour of the sequence is plotted in the center of each panel. The hour the sequence starts is shown in purple (PST).

2.5 Large deviations ("jumps")

The data show points with substantial deviations from immediately preceding and subsequent hourly values (i.e, jumps). The four largest jumps in the hourly data set are shown in Figure 7, and have deflections of roughly 50 degrees F from the preceding and subsequent data values. Jumps can be in the positive or negative sense. Jumps differ from extreme outliers (section 2.2) in that they need not be extreme outlier values themselves; rather, it is their departures from the immediately preceding and subsequent points that are extreme.

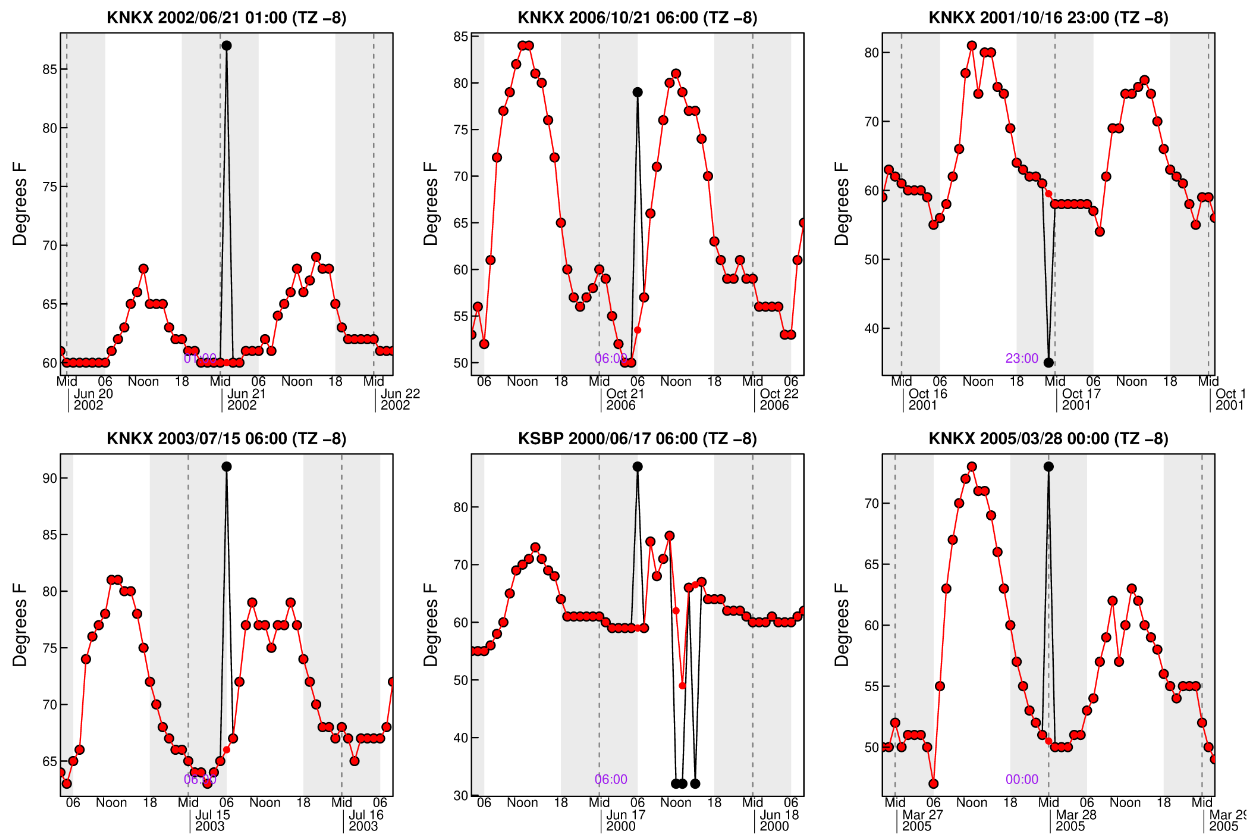

Figure 7. The six largest single-point jumps in temperature in the hourly record data set. As previously, the black dots show the original data values, and the red dots show the data after the quality control and infilling process.

To qualify as a jump that is removed and infilled by the quality control process, a point must have large, opposite signed departures from the preceding and subsequent point. So, for example, the top left panel of Figure 7 shows a point with a large positive jump from the preceding value, and a large negative jump to the next value. The top right panel shows the opposite, with a large negative jump followed by a large positive jump. Points with very large jumps on only one side are termed “lopsided” and indicated via a special error code, but not eliminated or infilled. This might be seen, for example, when a cold front moves over a station, resulting in a sudden drop in temperature, or when a stratus deck recedes from a coastal location in the summer, resulting in a rapid increase in temperature. .

To be specific, given a sequence of hourly temperature values X^1, X^2, X^3, …, where the superscript indicates the hour index, we define the *previous* jump j^p of a point at timestep t as $X^{t-1} – X^t$ and the *next* jump j^n as $X^t - X^{t+1}$. The final jump value for the point is then abs($j^p - j^n$), which is also the 1-D discrete Laplacian operator, as is often used in edge detection algorithms. A jump is considered "lopsided" if 1/3 <= abs(j^p/j^n) <= 3.

Although Figure 7 shows actual data values, these are not used for calculating jump values. This is because the climatological diurnal cycle at most stations has stronger temperature changes at sunrise and sunset, and smaller rates of change in the middle of the day and night. Therefore, an excursion from the immediately preceding and subsequent values of a fixed amplitude is more unusual at some hours of the day than others (e.g., a 4 degrees F excursion is more unusual at midnight than at 10 AM). The diurnal variation of large jump values is illustrated for 4 stations in Figure 8, which shows the 99th percentile value of jump values as a function of time of year and hour of day. Systematically larger jump values in the morning and evening are particularly clear at KSAC and KIPL, while large jumps are more likely to be seen in the early afternoon at KSFO, especially during the warm season, likely due to fog burning off around this time.

To account for the larger typical jump values at certain parts of the day, the hourly temperature data are first converted to anomalies by taking out the annual and semi-annual harmonics calculated separately for each hour of the day, and then standardized separately for each hour of the day. Although excursions are sensitive to the time of year as well, this factor is not taken into account when standardizing the data.

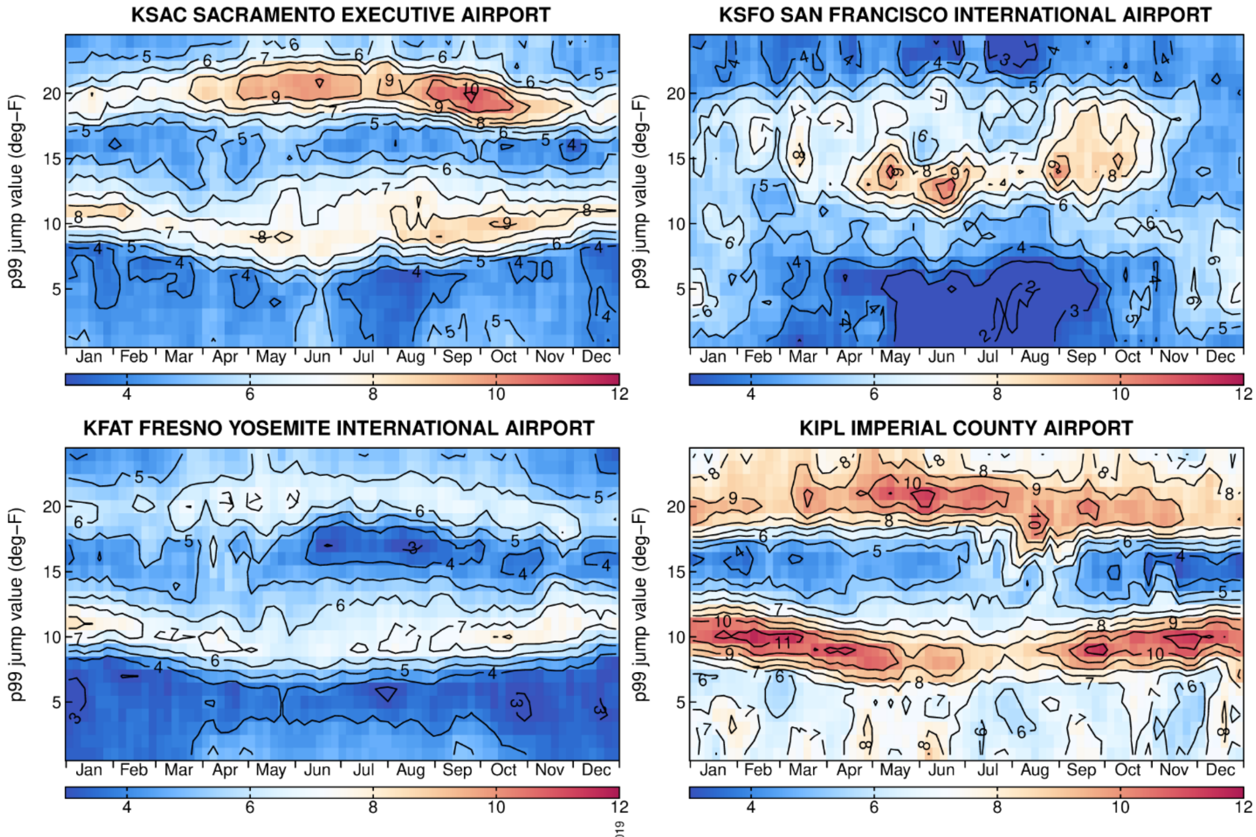

Figure 8. 99th percentile jump value (degrees F) as a function of time of year and hour of day illustrated for 4 stations. The data have been moderately smoothed to reduce noise before plotting.

Which jump values to identify as indicative of bad data and subsequently eliminate are based on a percentile threshold value of the standardized anomalies, such that the largest jump values are eliminated. The selection of this threshold percentile value is important since it dictates how much of the data is rejected. In this work, the chosen threshold percentiles are based on a log-log plot of the sorted size of the jump values, which we term a "jump value spectrum" by way of analogy with the conceptually similar eigenvalue spectrum used in principal component analysis. In an eigenvalue spectrum, physically significant (non-noise) modes are distinguished from background noise by identifying the first few largest eigenvalues that lie above the distribution of values fitted to the exponential noise tail (e.g., Wilks 1995). Likewise, we seek to identify large jumps that rise systematically above the distribution of the noise tail and likely indicate bad values and other non-physical problems with the data. Our methodology is therefore to sort the jumps from largest to smallest, fit an exponential distribution to the middle of the distribution of jump sizes (we use points 20 to 1000 in the fitting), and examine the largest jump values that fall above this fitted curve. It should be kept in mind that sampling variability means the occurrence of extreme jump values can depart from the fitted distribution due to chance as well as due to the data errors we seek to identify and mark as bad.

Examples of jump spectra from 6 representative stations are shown in Figure 9. The left column shows spectra for two large airport locations, KLAX and KSFO, with the red line showing the jump spectrum of the original CEC data and the green line showing the jump spectrum obtained after the quality control procedures described up to this point (removing extreme outliers and sequences of bad values, and merging the CEC data with the ASOS data). At KLAX and KSFO the jump spectra at the highest jump values (first few points along the left hand edge of the plot) are consistent with the best-fit line to the values over the center of the distribution from points 20 to 1000 (dashed black line). The quality control processing performed so far has not altered the original jump spectrum appreciably (i.e., there are only minor differences between the red line and the green line), as the data at these primary airport stations is high quality to begin with. Stations where the original data was reasonably well fitted even at large jump values, and the quality control procedures to this point have resulted in modest alterations to the jump spectra, include KBUR, KLAX, KRAL, KSFO, and KSNA.

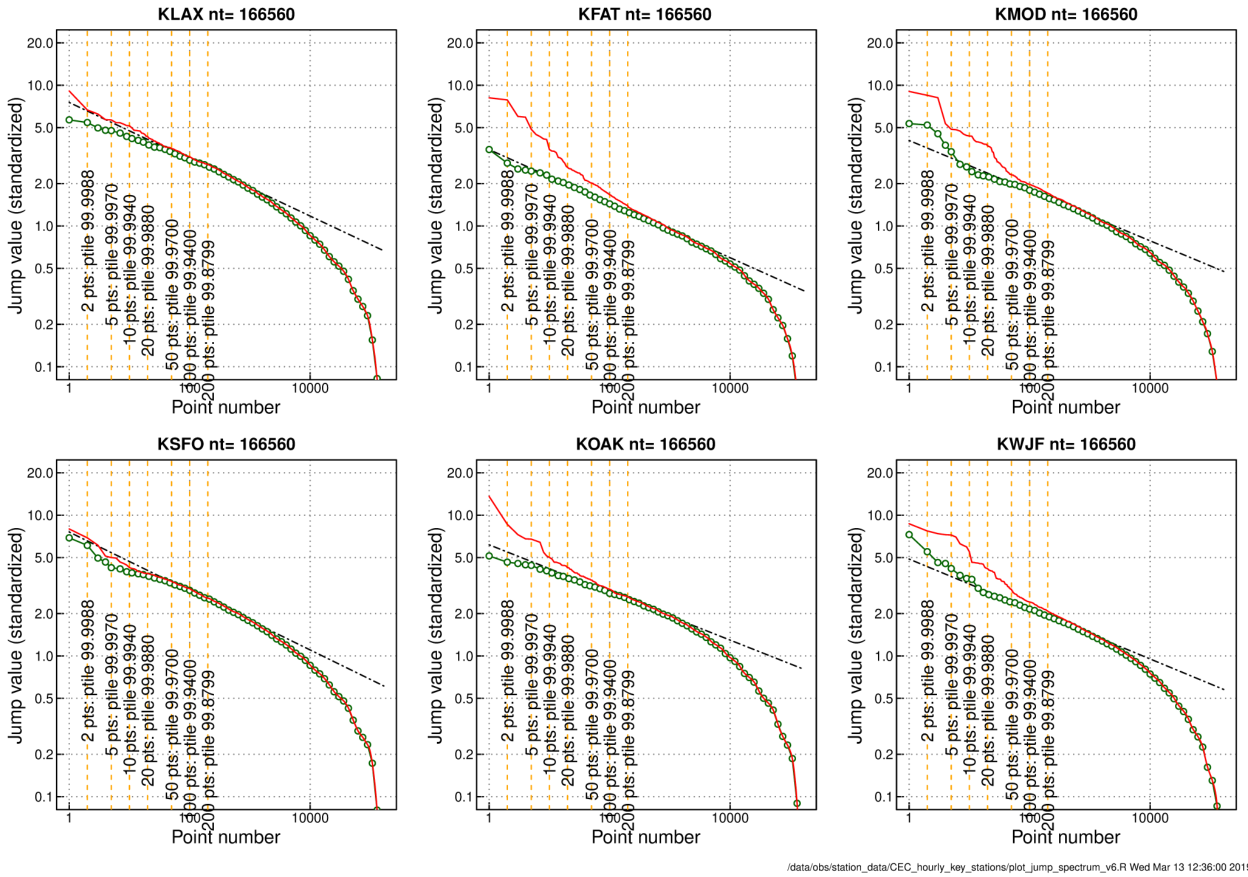

Figure 9. Example jump value spectra for 6 stations, as indicated in the plot titles. Red line: the original CEC data's jump spectrum before any quality control. Green line with circles: the jump spectrum after extreme outliers and sequences of bad values have been removed, and merging the CEC data with the ASOS data, but before any jumps have been removed. The dashed black line shows the best-fit line (in log-log space) to the values between point number 20 and point number 1000. The vertical labels show percentiles for different numbers of points; for instance, 5 points represents a percentile value of 99.997 (given the 166,560 hourly points in the original time series).

The middle column of Figure 9 shows cases where the original jump spectrum (red) has a pronounced enhancement relative to the best-fit line at the highest jump values (left edge of plot), indicative of exaggerated jumps associated with bad data values, but the spectrum after quality control to this point has resulted in a jump spectrum (green line) consistent with the best-fit line (black dashed line). At such stations the quality control procedures to this point have eliminated the majority of bad values. Stations with this kind of behavior include KBFL, KEED, KFAT, KLAS, KLGB, KOAK, KOXR, KRBL, KSAC, and KSAN.

The right hand column of Figure 9 shows cases where the quality control to this point has still left points with appreciable jumps that do not fit the distribution of noise values, and so need further processing. The number of points in excess of the fitted noise distribution can be read off the jump spectra using the vertical labels and are marked as bad data and infilled by linear interpolation. The stations processed this way and the number of points removed are shown in Table 3. KNKX is a clear outlier, with many more residual bad data values than the other stations.

Table 3. The number of extreme jump value points eliminated from each station. The number eliminated is obtained from jump spectra of the type illustrated in Figure 9.

STATION	# PTS REMOVED
KBLH	9
KCQT	1
KIPL	20
KMCE	1
KMOD	5
KNKX	50
KPSP	9
KSBP	10
KSJC	18
KSBA	4
KTRM	4
KUKI	20
KWJF	6

3. Hourly disaggregation

With the quality control of the data addressed, we turn our attention to the hourly disaggregation process. We compare two approaches to disaggregating the twice-daily (Tmin and Tmax) data to hourly values. The first is based on climatological diurnal cycles of temperature, which we call the climatological curve approach. The second identifies historically observed analog days with similar 3-day sequences of daily Tmin and Tmax and uses the hourly values from the analog days to generate the hourly values. We describe these approaches in detail below.

To quantitatively evaluate the characteristics of the hourly disaggregation methods, we use a three-way cross-validated approach by splitting the observed hourly temperature record at each station into three equal parts. We then train the disaggregation on the hourly data from two of the parts. The remaining part is used as a validation period. We calculate the *daily* Tmin and Tmax from the hourly data over the validation period and use the disaggregation technique to produce hourly values from the computed Tmin and Tmax. The disaggregated hourly values can then be compared to the known hourly values from the validation period using a variety of measures, as shown in the next sections.

3.1. Hourly disaggregation with climatological curves

For the climatological curve approach, we first compute mean hourly sequences of diurnal temperature based on the quality controlled historical hourly data, and then scale and offset those climatological curves to match the actual Tmin and Tmax values for the model day being disaggregated. Finally, the multi-day sequence of hourly curves is spliced smoothly together across midnights to form a time sequence of hourly values covering the entire period. In this work we use separate climatological curves for each station, month, and quartile of the diurnal temperature range (Tmax – Tmin). This yields $12 \times 4 = 48$ curves for each of the 29 stations.

There is a surprising variety of climatological curves found at the different stations (Figure 10). For example, at some stations variations in the diurnal temperature range (DTR) in January (as indicated by different colors in the figure) are associated with variability in the early morning hours (KBFL), while at other stations DTR variations are associated with variability in the late night hours (KBUR). In some stations stratifying by DTR makes very little difference to the shape and value of the curve in July (KBFL, KMCE), while at other stations it makes a very large difference (KLAX, KSAN). At some stations the curve associated with the highest DTR quartile (red line) is well separated in value from the other curves throughout the year (KLAX and KSAN), while at other stations the curve for the highest quartile is has a separation that is part of an even progression from the smallest to largest DTR quartiles (KRAL, KBUR).

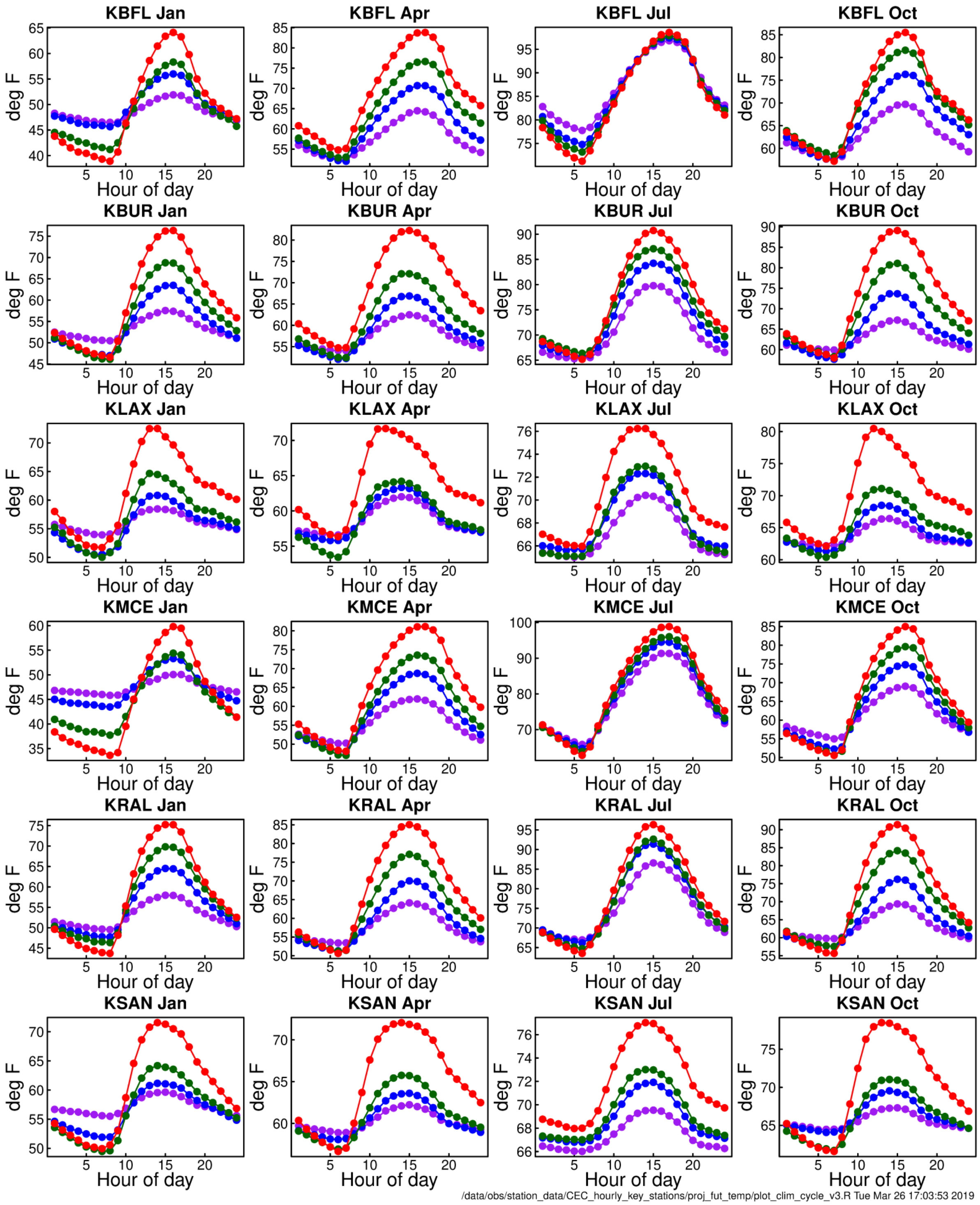

Figure 10. Climatological daily curves of near-surface air temperature for selected months (columns) and stations (rows). The color indicates the quartile of the diurnal temperature range used in constructing the curve. The progression in colors from purple to blue to green to red maps to the progression from lowest to highest quartile of diurnal temperature range.

The actual values of the climatological curves are shown in Figure 10. The variability illustrated in that figure includes changes in the amplitude of the diurnal cycle as well as its mean value. A version of the same figure but with the diurnal curves normalized to the range of 0 to 1, so that the shapes of the curves can be more easily seen, is shown in Figure 11. When considering the shapes of the curves, without regard to their mean value or amplitude, it is evident that stratifying by the quantile of DTR influences the shapes of the curves chiefly in the hours leading up to midnight. The shapes of the curves in the early morning hours are less strongly related to the DTR.

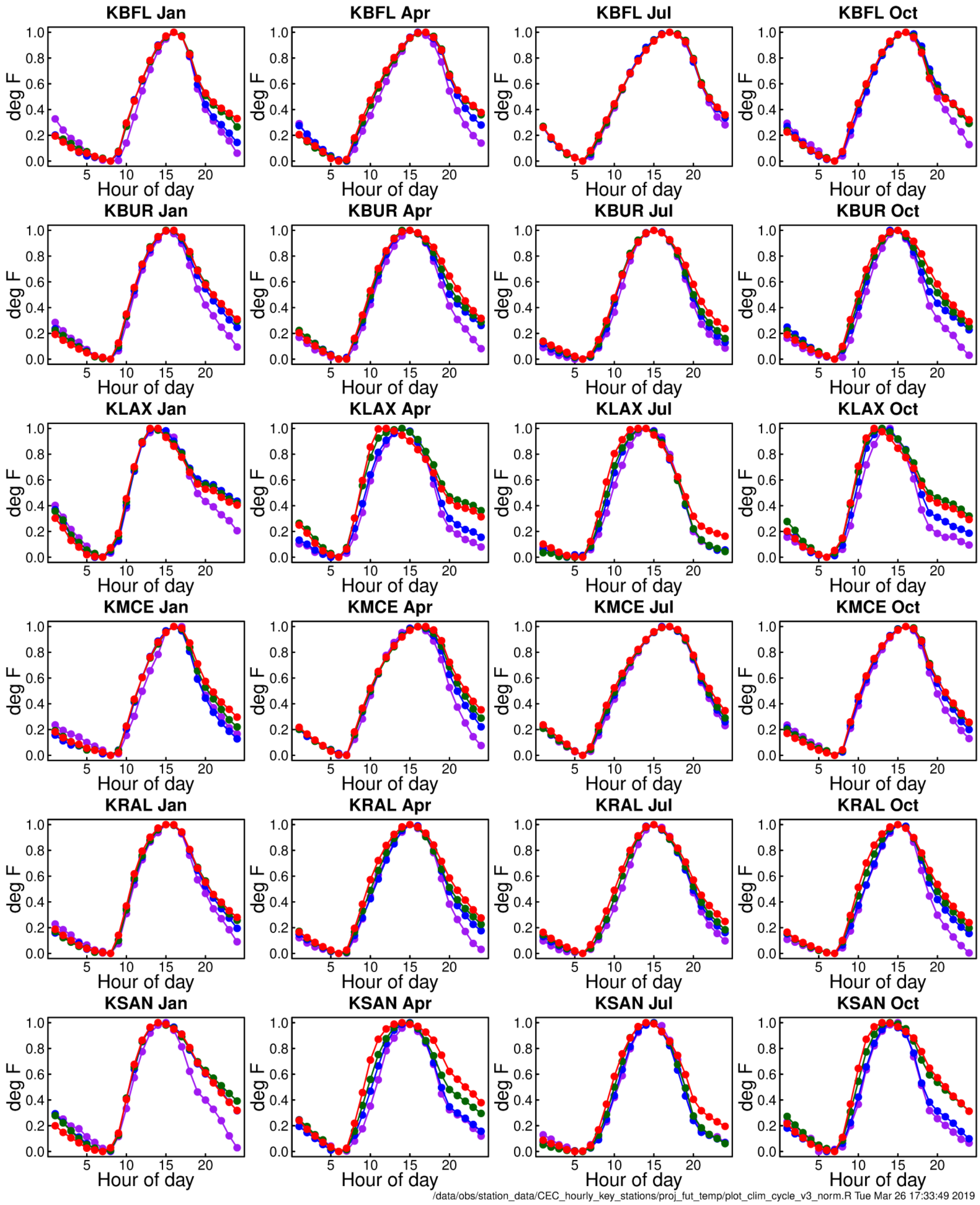

Figure 11. As in Figure 10, but with the curves normalized to the range 0 to 1.

When using the climatological curve approach, values just before and just after midnight can be discontinuous, since each day is matched to its own climatological curve without regard to the day before or after. To smooth the transition across midnight, we splice the days together: values leading up to and just after midnight are damped towards the mean temperature between midnight of one day and 1 AM of the next day.

3.2 Hourly disaggregation using analog days

The idea behind hourly disaggregation using analog days is straightforward. Imagine we are disaggregating the Tmin and Tmax from a model day to hourly values. We find the historically observed day whose Tmin and Tmax best matches Tmin and Tmax from the model day being disaggregated and use the hourly temperatures from that observed day as the solution.

Although that is the basic idea, we elaborate it slightly here. When disaggregating model day number *N,* we consider the 3-day sequence of Tmin and Tmax *anomalies* in model days *N*-1, *N*, and *N*+1. We use the 3-day sequences of model Tmin and Tmax anomalies as the basis for matching model day *N* because a) we want to improve the realism of the transitions between days, and b) we want to better constrain the match by using more data (3-day sequences yield 6 Tmin/Tmax values for the matching, vs. 2 Tmin/Tmax values that would be used if only one day were used). The result of this matching is the three-day sequence of observed days, *M*-1, *M*, and *M*+1, that has the smallest RMS difference between the sequence of 6 model Tmin/Tmax anomaly values and the sequence of 6 observed Tmin/Tmax anomaly values. We additionally require that the day-of-year of *M* must be within 45 days of the day-of-year of *N*, so that, for example, a winter sequence in the model cannot be matched to a summer sequence in the observations. This use of the 45 day window is why the matching is done on anomalies with respect to a centered 30-year climatology rather than the actual Tmin/Tmax values; using anomalies prevents future climate warming from systematically matching model days to observed days closer to summer, which tend to be warmer. Hourly temperature values observed on day *M* then become the basis for the hourly disaggregation of model day *N* (the preceding and subsequent days obtained in the matching are discarded).

The above matching results in a sequence of discontinuous observed analog days to use as the basis of the hourly disaggregation. To make smooth transitions across midnights, we use the same splicing technique as applied to the climatological curve approach. Values leading up to and just after midnight are damped towards the mean temperature between midnight of one day and 1 AM of the next day.

Figure 12 shows example time series of hourly temperatures from the original observations, and the results of disaggregating daily Tmin and Tmax to hourly values using the analog day and climatological curve methods, both applied in a cross-validation framework. The most obvious difference in character between the curves is that the climatological curve solution (dotted red line) lacks the hourly variability seen in the observations or analog day disaggregation. This is expected, but emphasizes that the advantage of the analog matching approach is that it better captures the statistics of hourly variability than does the climatological curve approach, which by construction has little hourly variability.

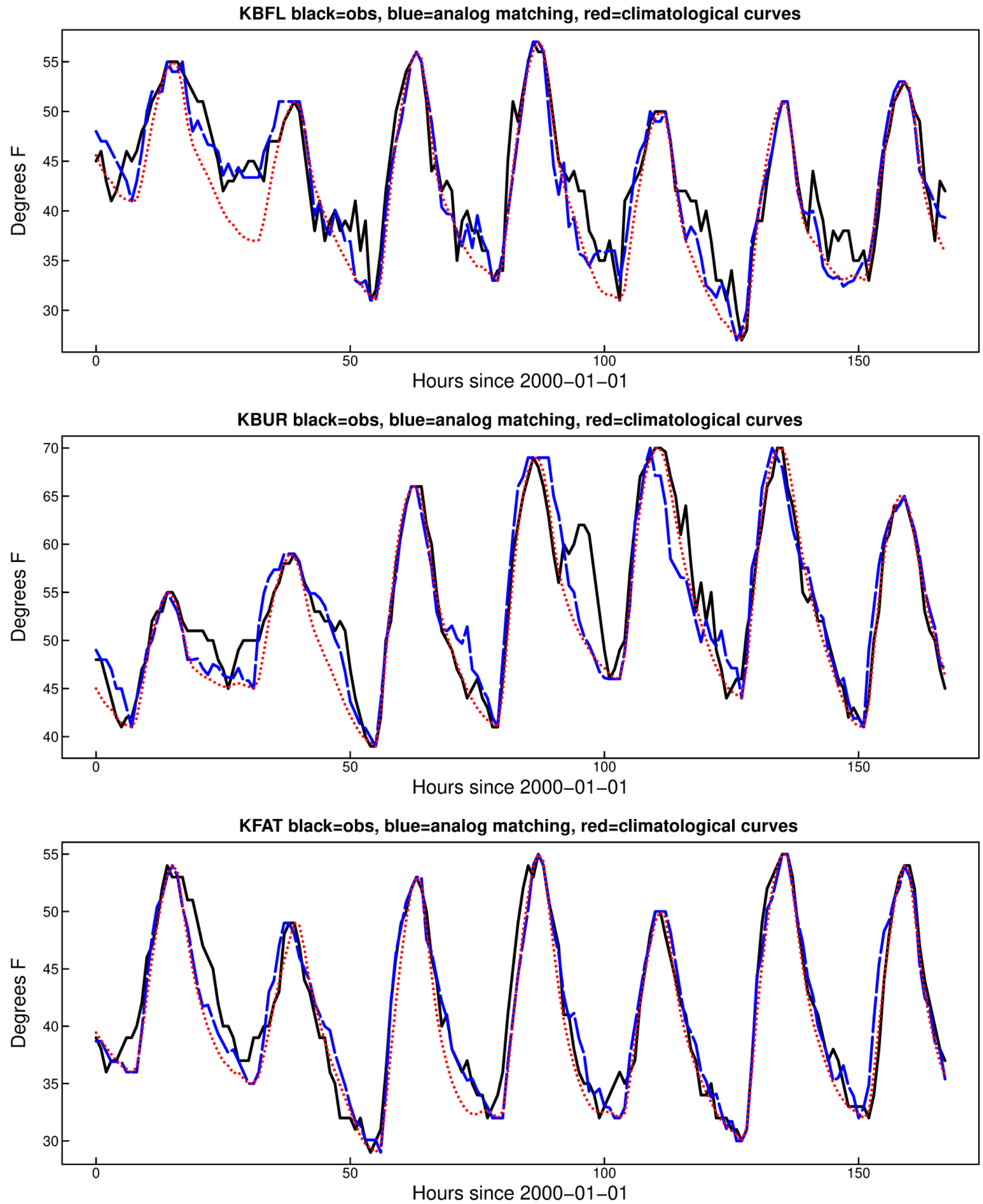

Figure 12. Example time series of hourly temperature values at three stations (one per panel), covering a period of one week. In each panel the black line is the original hourly observations. The broken blue line is the matching analog day solution, and the dotted red line is the climatological curve solution. The LOCA and climatological curve solutions are obtained using cross-validation.

The difference in the variability characteristics of the analog day and climatological curve approaches can be seen more quantitatively in the spectra (Figure 13). The spectrum obtained using the analog day approach matches the observed spectrum very well at all timescales from 2 hours to 19 years.

However the spectrum obtained from the climatological cycle approach differs substantially from the observed spectrum at time scales between 2 hours and 1 day, with spectral power in the climatological curves approach being about two orders of magnitude low at the highest resolved frequencies.

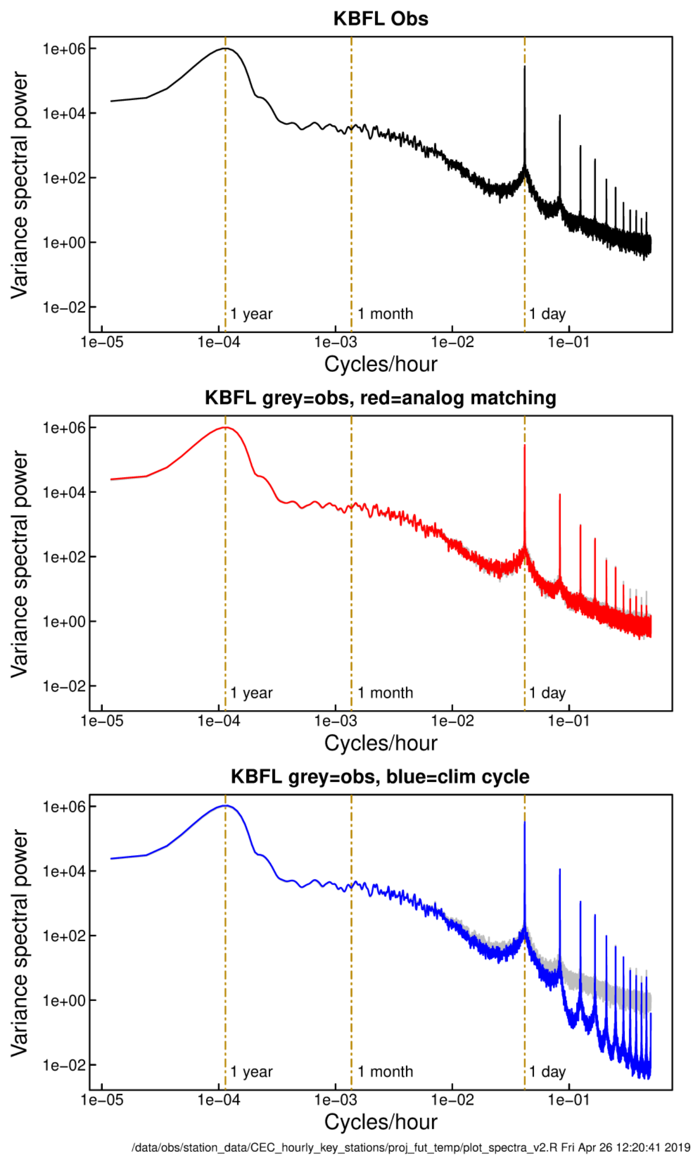

Figure 13. Spectra of the hourly temperature time series at station KBFL, from the observations (top panel, black), analog matching approach (middle panel, red), and the climatological cycle approach (bottom panel, blue). In the middle and bottom panels, the grey line repeats the observed spectrum, for comparison.

Another way of evaluating the variability of the disaggregated hourly values is to examine the distribution of hottest and coldest hour of the day. This is illustrated in Figure 14 for station KBFL.

Considering the warmest hour of the day first (red lines and dots), the observations (top row) show that 4 PM LST is most often the warmest hour of the day, with an occurrence of about 40%. However, in all seasons any hour between 2 PM and 5 PM LST has a chance of being the warmest hour of the day. The analog matching approach preserves this distribution extremely well (middle row), while the climatological curve approach, by its nature, has much less variability in which hour of the day is the warmest hour (bottom row). Results from the coldest hour of the day (blue) are similar, but the match between the observations and analog day disaggregation is degraded due to the splicing together of adjacent days, which influences the hours before and after midnight. Nonetheless, the analog day solution preserves variability in the coldest hour of the day much better than the climatological curve approach, which has little variability in this measure.

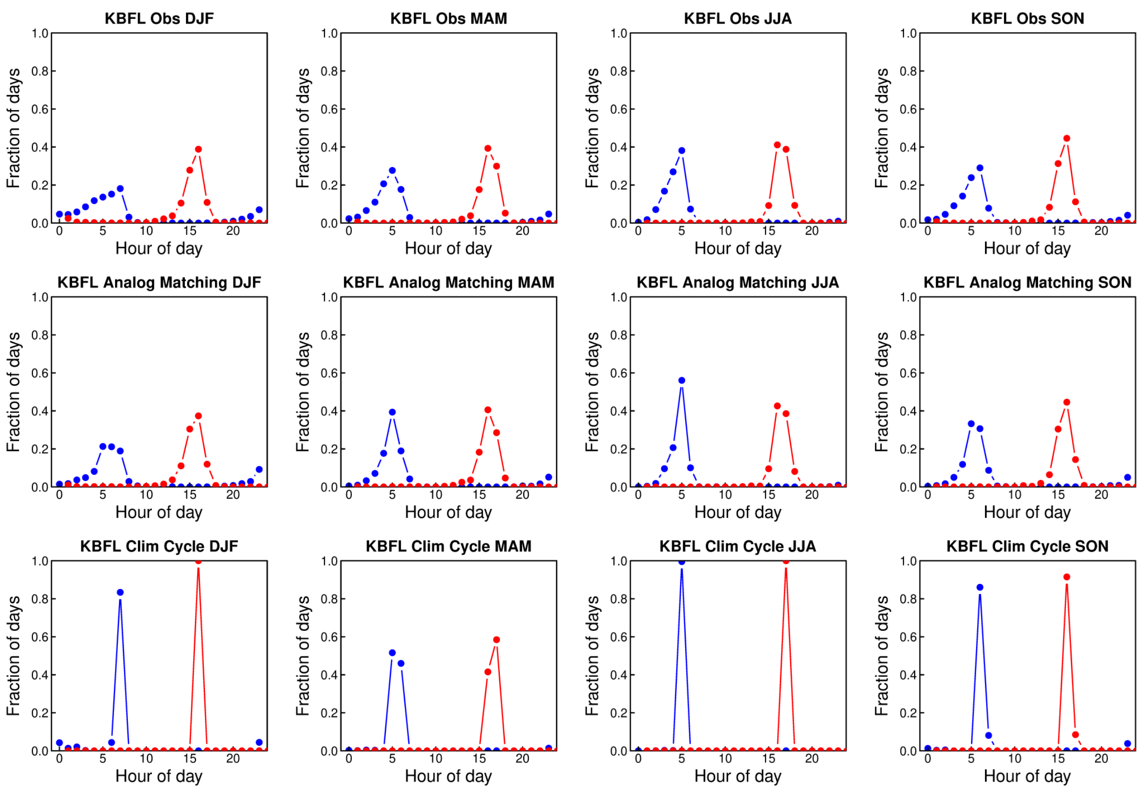

Figure 14. Histograms of the fraction of time that each hour is the warmest (red) or coldest (blue) hour of the day. Values are for Bakersfield (KBFL), and stratified by season (from left to right, the columns are Dec-Jan-Feb, Mar-Apr-May, Jun-Jul-Aug, Sep-Oct-Nov). The top row is from the observations, the middle row is from the analog matching approach, and the bottom row is from the climatological curves approach.

A more complete examination of how the temperature at different hours compare to each other is shown in Figure 15, which illustrates the percent of days where the temperature at hour *A* (vertical axis) is greater than the temperature at hour *B* (horizontal axis). So, for instance, in the observations (top left panel) hour 1500 is warmer than hour 0700 more than 99% of the time, while hours 1000 and 2000 are close to being evenly matched, with approximately equal chances of being the warmest hour of the two. The errors in replicating this result in the analog day disaggregation (top right panel) are less than 1 percentage point over most hour pairs, with the largest errors appearing in the early morning hours, likely due to the splicing required. The climatological curve solution, by contrast, has a much larger range of hour pairs over which errors greater than 1 percentage point occur. Errors are particularly noticeable for pairs of hours that are nearly matched, with a clear tendency for the climatological curve solution to amplify (have the same sign as) the observed pattern. In other words, if hour *A* has a mild tendency to be warmer than hour *B* in the observations, this tendency is strongly exaggerated in the climatological curve approach. Such behavior is understandable as a consequence of the reduced hourly variability obtained when using climatological curves.

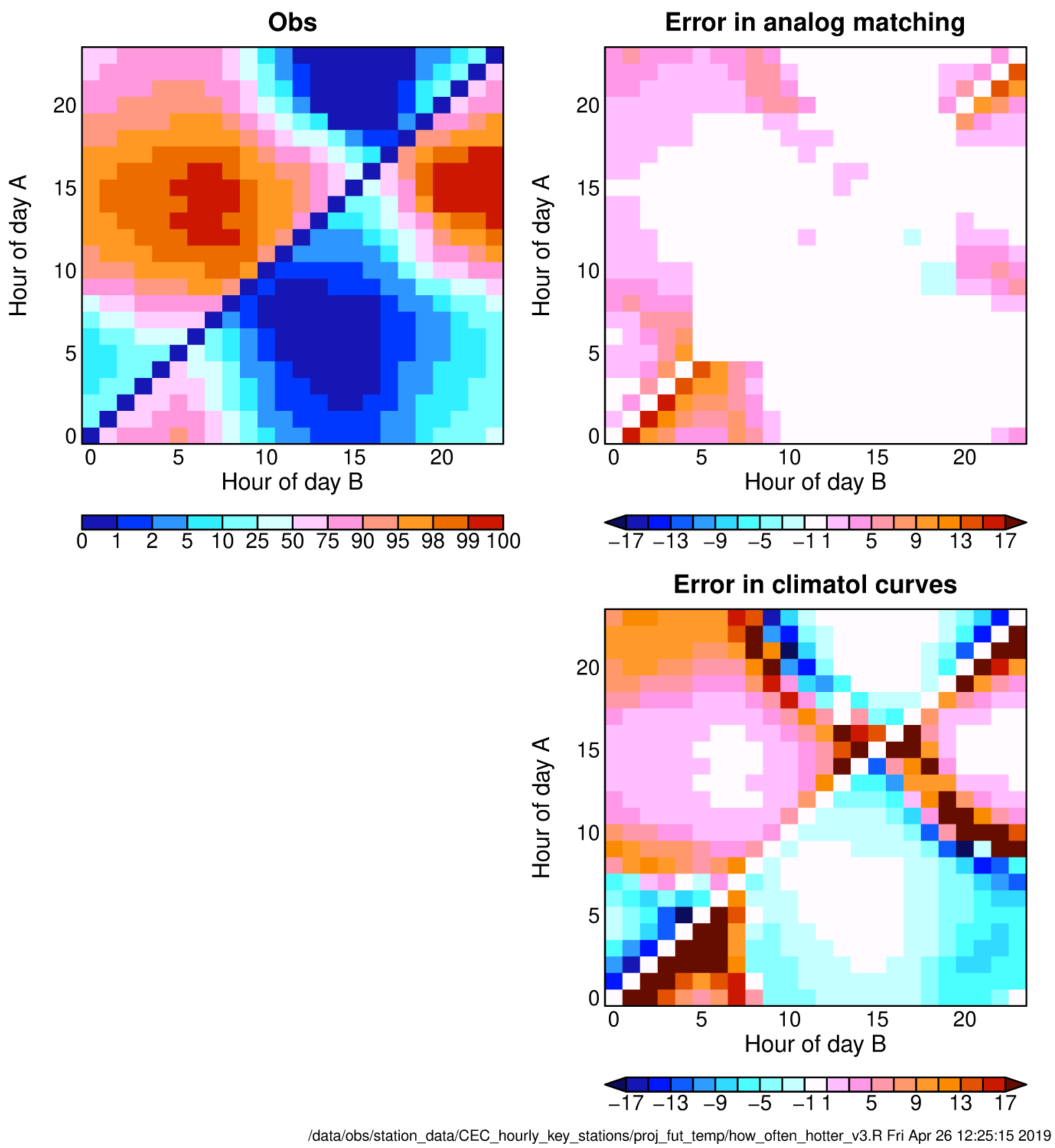

Figure 15. Top left: percent of time that hour-of-day A (vertical axis) is warmer than hour-of-day B (horizontal axis) in a day. Top right: error with respect to the observational result obtained using the analog day approach. Bottom right: error w.r.t. observational result obtained using the climatological curves approach.

4. Future projections

Future projections of hourly temperature are obtained using the analog matching method to disaggregate daily Tmin and Tmax from 10 global climate models (GCMs) downscaled using the Localized Constructed Analogs technique (LOCA; Pierce et al. 2015). The GCMs cover the period 1950 to 2100 (2099 for some models) and include two scenarios of future greenhouse gas emissions: a business as usual scenario termed RCP 8.5, and a scenario where moderate cutbacks in global greenhouse gas emissions are achieved by mid-century, termed RCP 4.5. The 10 GCMs used here are those identified by the California Department of Water Resources climate change technical advisory group as being the best set to use for future projections of California's climate (DWR-CCTAG). The analog matching method was used for the hourly disaggregation in preference to the climatological curve approach due to its superiority in capturing variability, as documented in the previous section.

The LOCA time series of Tmin and Tmax for each station were taken from the closest gridpoint to the station being processed. Since the LOCA gridcells are 6 km on a side, there can be systematic differences between LOCA gridcell-mean climatology and climatology at a specific point (station) in regions where there is a strong gradient of mean temperature with distance from the coast (i.e., coastal locations in California), or places where the mean LOCA gridcell elevation differs from the elevation of a particular station. There can also be differences between the LOCA and station climatologies due to systematic differences between the station data used here and the original temperature data used to train LOCA (which was Livneh et al., 2015). To reduce these biases, a simple monthly bias correction was first performed on the LOCA data that additively adjusted the monthly mean temperature of the LOCA data to match the mean monthly observed value over the historical period of the hourly observations (2000-2018), and multiplicatively adjusted the daily LOCA anomalies, by month, to match the observed standard deviation of daily anomalies by month.

4.1 Exceedance of thresholds

The multi-model ensemble average projected change in mean number of hours per year above and below selected temperature thresholds is shown in Figure 16 (note that the figure continues over several pages). The observed values over the historical period are shown by the purple dots and 95% confidence interval bars, and generally agree with the model values over the same time period.

The projected changes in hourly temperature are substantial at most of the stations. For example, under the RCP 8.5 emissions scenario, Sacramento is expected to increase from an average of < 50 hours/year above 100 °F to almost 250 hours/year by the end of this century, a five-fold increase. Increases in some locations are dramatic, with the number of hours exceeding 100 °F rising by an order of magnitude, particularly by the end of the century for the RCP 8.5 (high emissions) scenario. At many of the stations, greenhouse gas emissions consistent with the RCP 8.5 scenario would result in more hours above 110 °F by the end of this century than were seen above 100 °F at the end of the 20^{th} century.

Such changes would have a substantial impact on daily electricity demand for cooling, which dominates over electricity demand for heating over the majority of the state. For example, historical electrical demand information from the California Independent Systems Operator (CalISO) over the period 2001-2011 shows that electrical demand for Pacific Gas and Electric (PG&E), San Diego Gas and Electric (SDG&E), and Southern California Edison (SCE) increases by approximately 300, 60, and 325 MW per degree F, respectively, above 80 °F. These represent increases of approximately 2% per degree F. A 10 °F increase in the extreme warm hours, such as projected at many station locations by the end of this century under the RCP 8.5 emissions scenario, would result in approximately 20% increase in energy demand. This is assuming linearity, which might underestimate actual increases in demand at the extreme, when generating capacity is stressed.

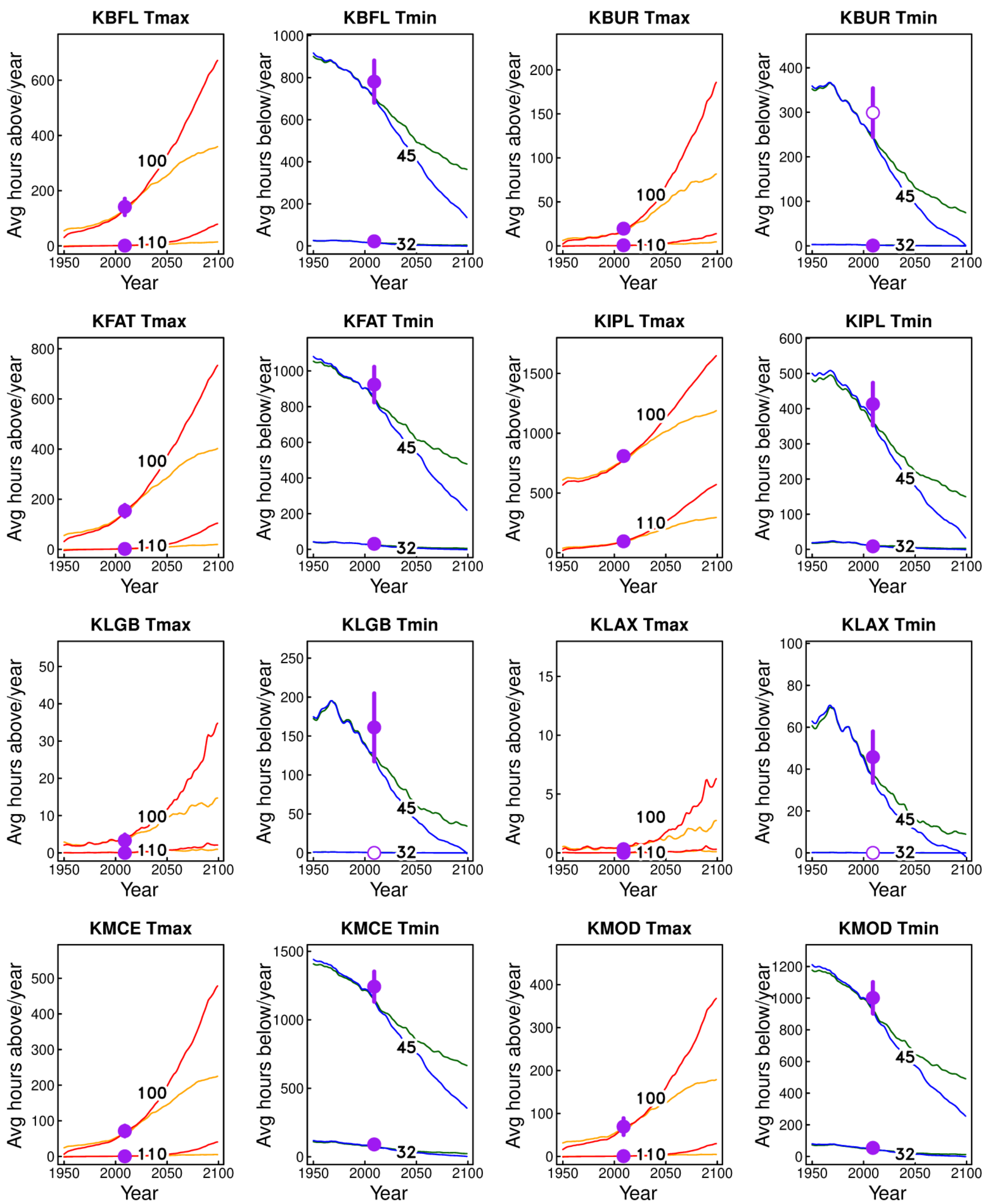

Figure 16. Multi-model ensemble averaged projected changes in the average number of hours per year either above (red and orange lines) or below (blue and green lines) the indicated temperature threshold in degrees F (100 and 110 °F for high temperature thresholds; 32 and 45 °F for low temperature thresholds). The red and blue lines show results from RCP 8.5; the orange and green lines are for RCP 4.5. The purple dot and whisker show the observed mean value and 95% confidence interval over the historical period (2000-2018). The purple dot is solid where the observed and model values agree when taking their mutual uncertainties into account, and open where they disagree. Figure continues.

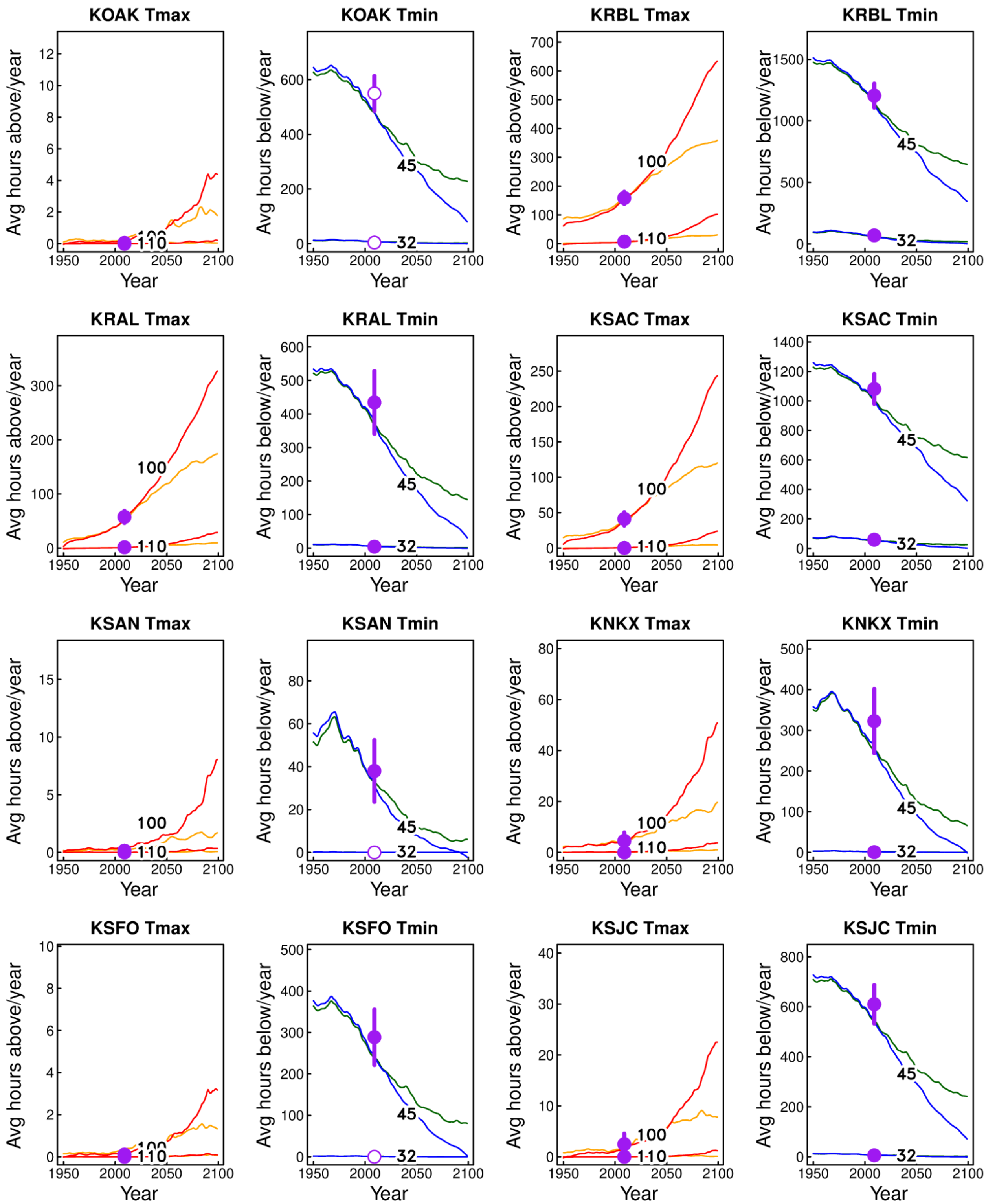

Figure 16, continued. Figure continues.

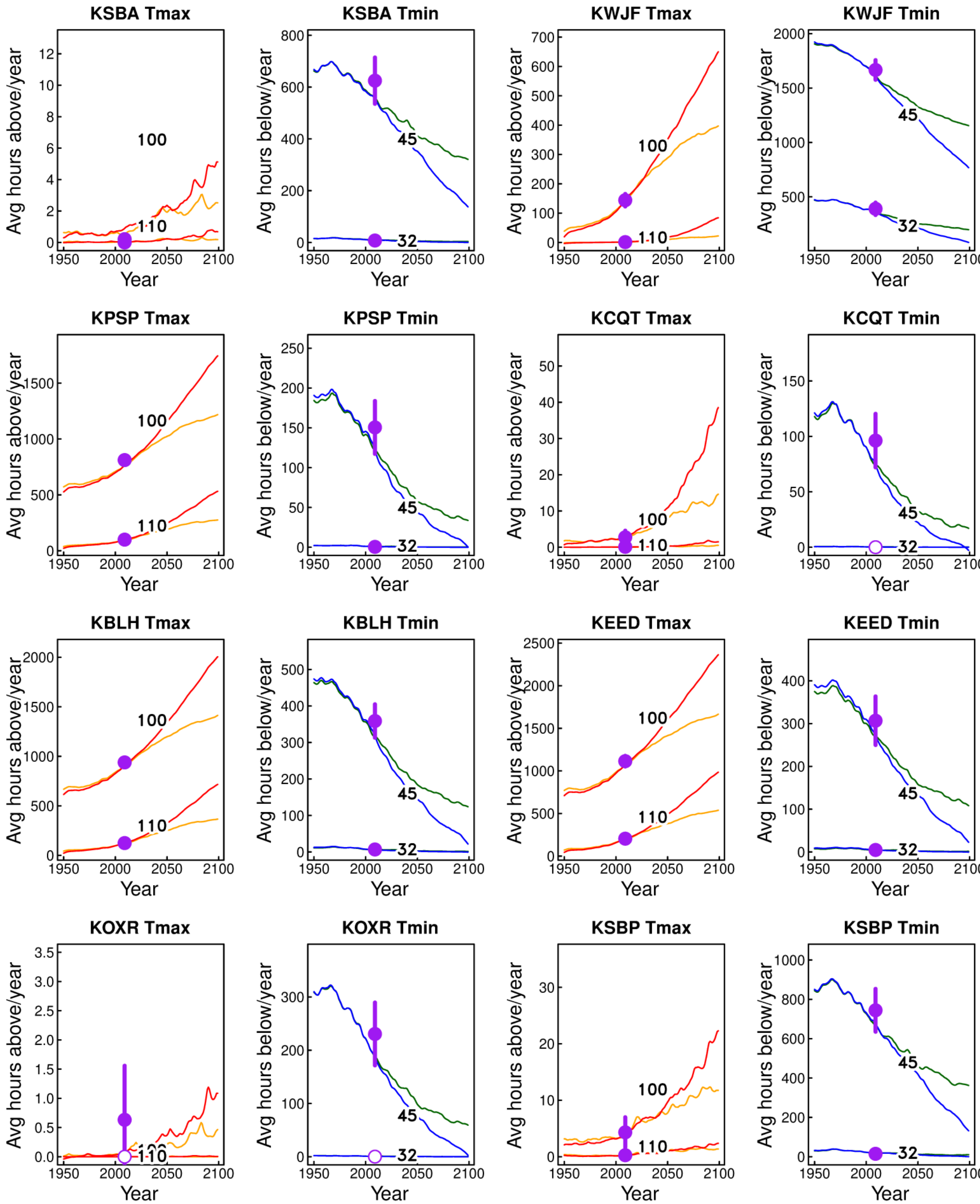

Figure 16, continued. Figure continues.

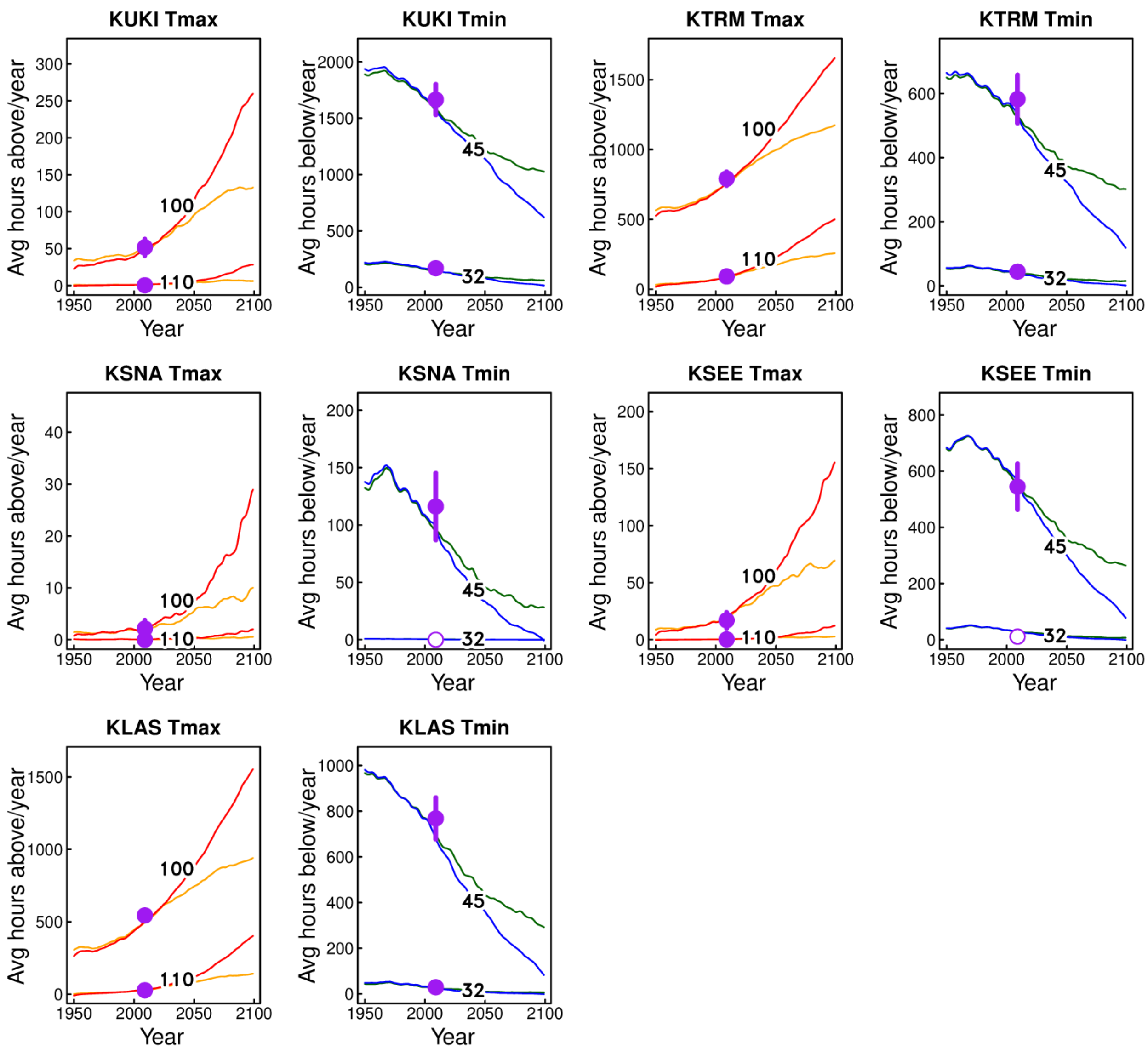

Figure 16, continued.

Strong projected declines in the number of hours below 45 °F (chill hours) are also seen. For example, at Bakersfield, chill hours decline from about 800 to about 150 by the end of the century under the RCP 8.5 emissions scenario. Chill hours are important because they are needed by various fruit and nut crops grown in California, such as almonds, apples, cherries, pistachios, and walnuts. The large and relatively rapid changes in chill hours seen here could be disruptive to farms that depend on trees that take a substantial length of time to grow to maturity.

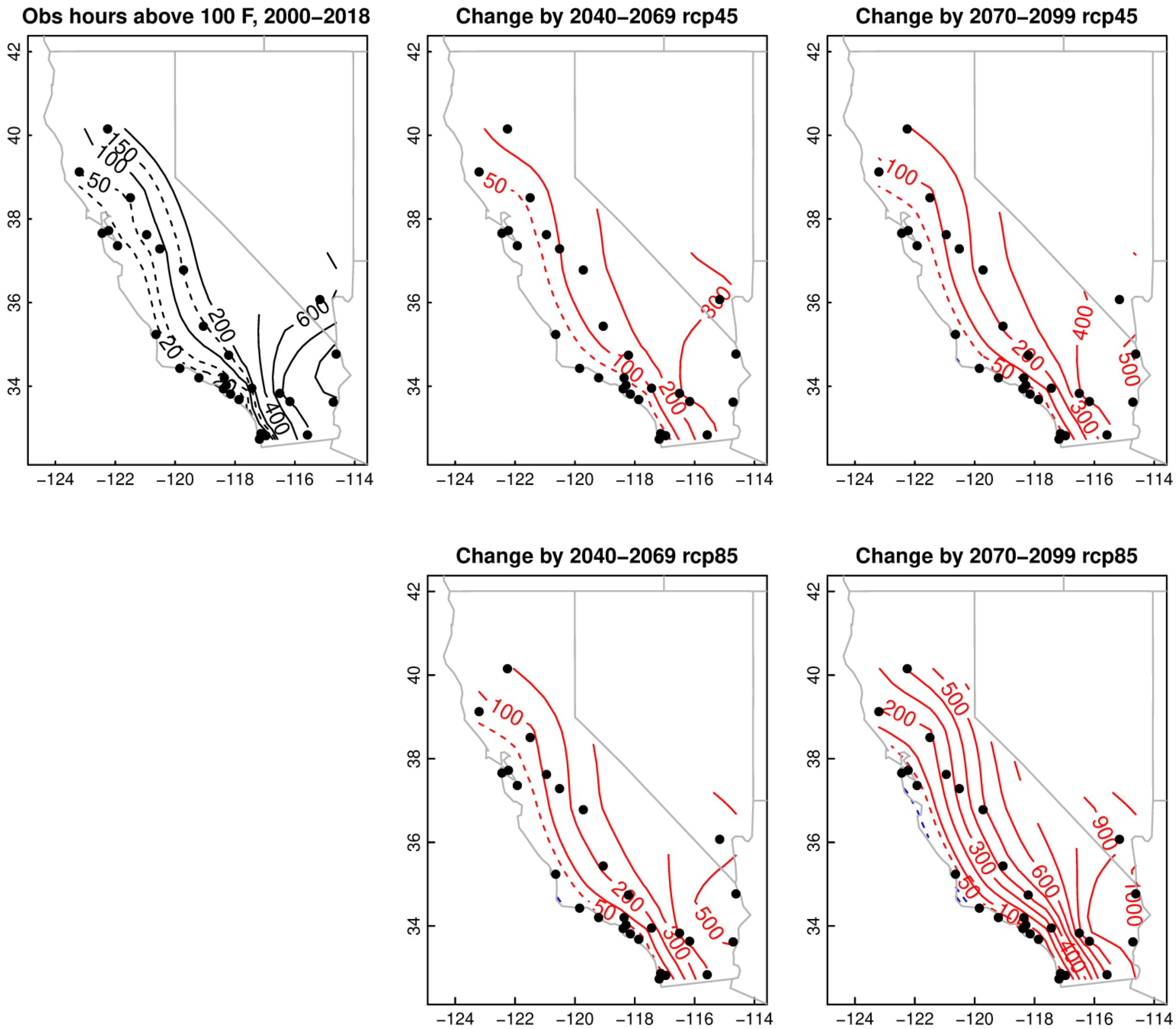

Figure 16. Top left: observed average number of hours/year with temperature >= 100 °F. Top middle and top right: change in number of hours/year by mid (middle) and end (top right) century, in the RCP 4.5 scenario (medium emissions). Bottom middle and bottom right: same, for the RCP 8.5 scenario (business as usual emissions). Contours of positive change (increases) are shown in red; negative changes are shown in blue.

The historical (2000-2018) number of hours/year with temperature >= 100 °F, and the model-projected change in hours/year, are shown in Figure 16. Changes are damped near the coast, and reach large values inland, in general agreement with the pattern of increased warming away from the coasts expected in California (Pierce et al. 2018). The differences between the RCP 4.5 and RCP 8.5 values are mild at mid-century but substantial by the end of this century, indicating that a reduction in greenhouse gas emissions would have a substantial mitigating effect on climate warming were it to happen.

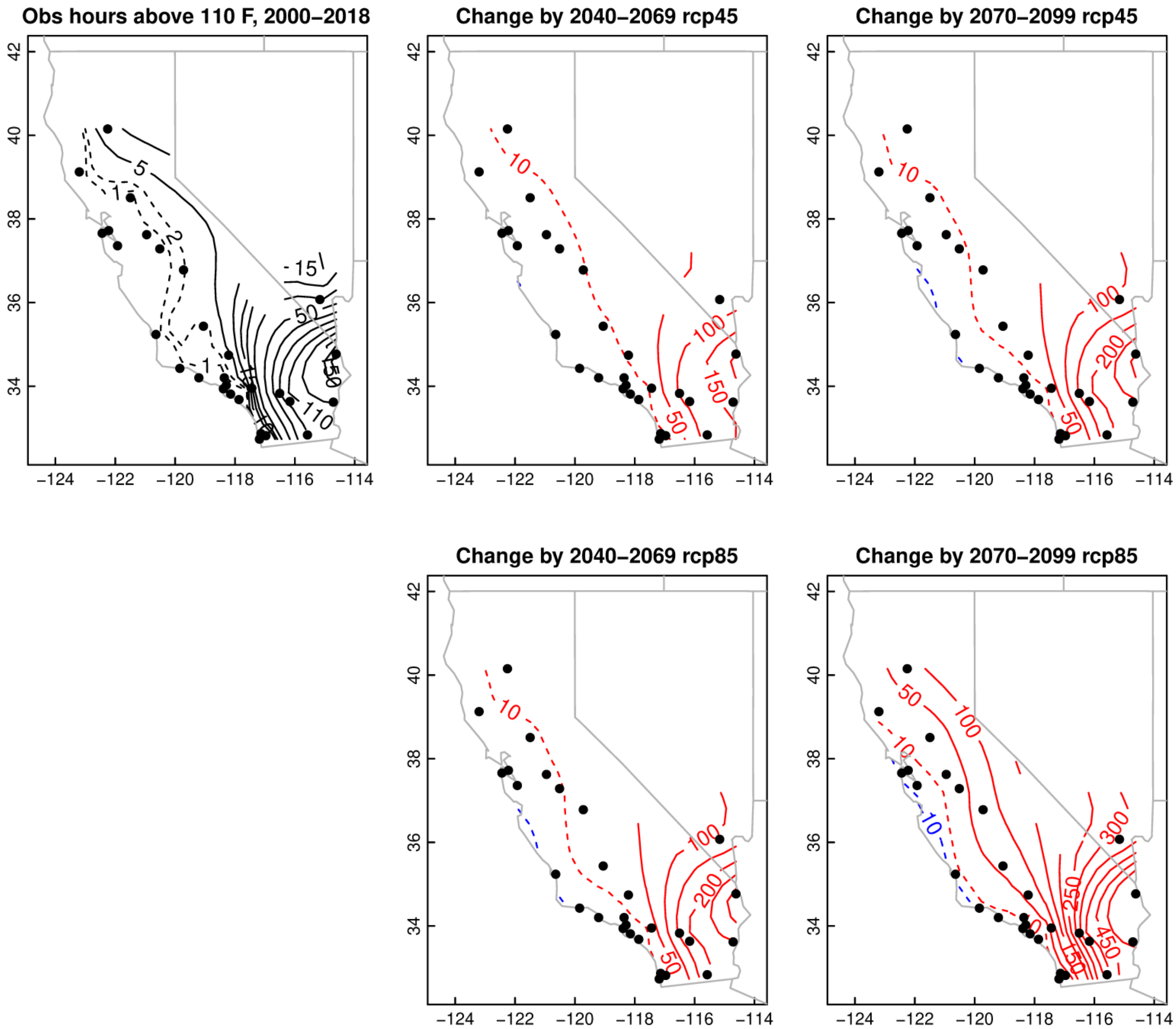

Figure 17. Similar to Figure 16, but for a threshold of 110 °F.

The climatological pattern and model-projected changes in the number of hours >= 110 °F are shown in Figure 17. It is interesting to note that fitted surface of model-projected changes shows negative values right along the coast at the end of century. None of the stations show negative values, however, so it is likely due to the attempt to fit a surface to the station values. However it should be noted that the stronger heating in the interior might contribute to an increase in sea breeze frequency or strength in the future, which could conceivably reduce the incidence of extremely high temperatures right along the coast. A more spatially detailed look at this result may be warranted.

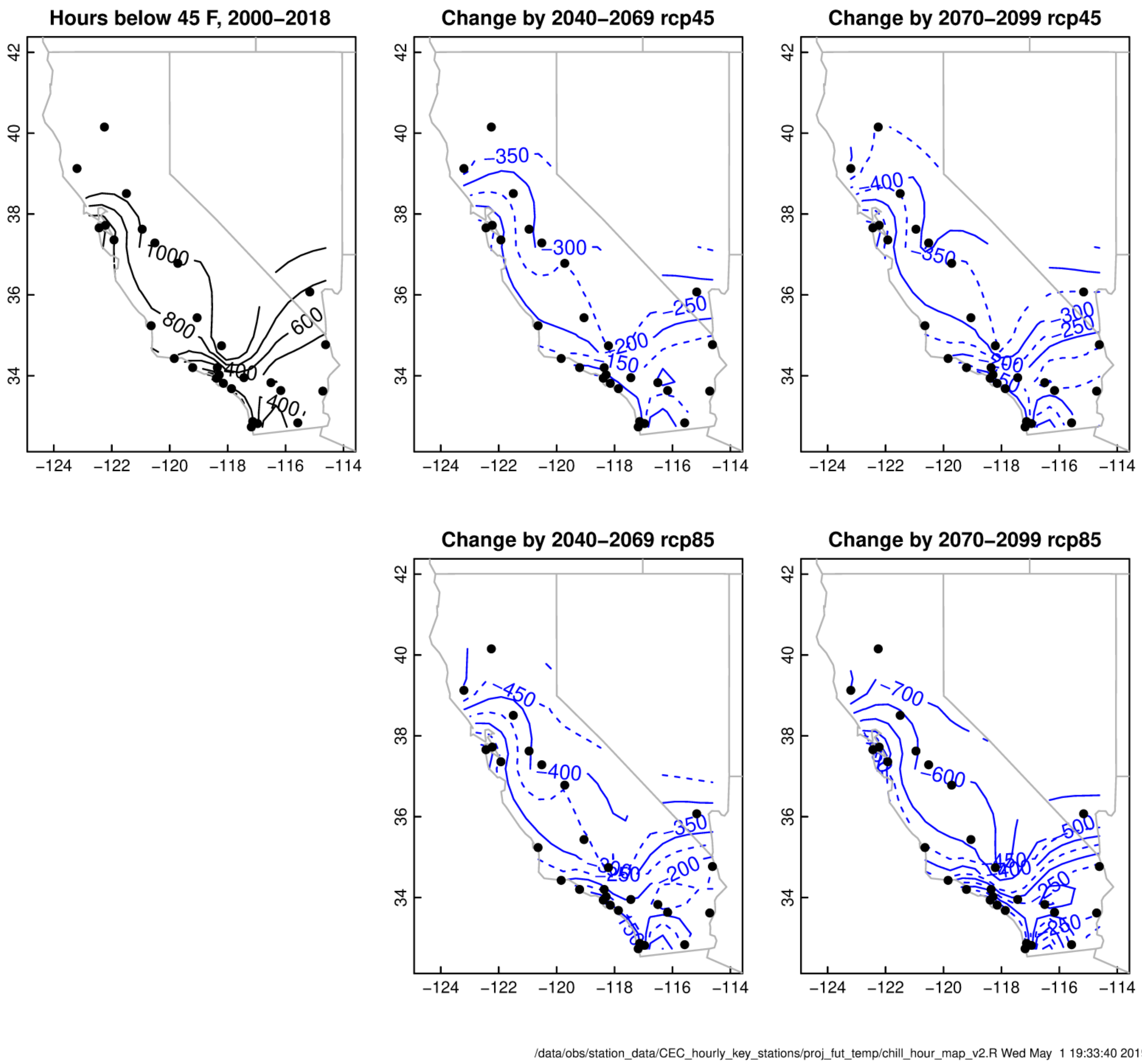

Figure 18. Similar to Figure 16, but for the average number of hours/year <= 45 °F.

The climatological pattern and model-projected change in the average number of hours per year <= 45 °F is shown in Figure 18. This metric is sometimes used as chill hours, which are required for successfully farming certain fruits and nuts. Declines in the number of chill hours are on the order of 20-

30% by mid-century, and reach 60% or more by the end of century in the RCP 8.5 scenario. Such changes would be associated with substantial disruptions to crops that depend on chill hours.

4.2 Change in the diurnal cycle

The statistical disaggregation process uses historically observed diurnal cycles, so it might be thought that future projections cannot show a change in the shape of the diurnal cycle (i.e., that it is a stationarity assumption of the method). This is incorrect because there could be a change in the frequency of Tmin/Tmax sequences that are linked to different diurnal cycles in the observational record. For instance, Figure 11 shows that days with a large diurnal range have a systematically different diurnal cycle shape than days with a small diurnal range. In that case, a change in the future distribution of diurnal ranges would trigger a change in the mean shape of the disaggregated diurnal cycle. There are still changes that cannot be captured by the stationarity assumptions of the method, however. For example, the process could not project a future diurnal cycle shape that has not been observed. More realistically, changes in the shape of the diurnal cycle that are not reflected in changes to daily Tmin/Tmax would not be captured.

To evaluate changes in the projected shape of the diurnal cycle we use a simple measure: the maximum absolute difference between the normalized historical and future seasonal mean diurnal cycle. (Referring to Figure 11 again, this is analogous to the maximum absolute difference between two curves illustrated there, but in this case the two curves are the historical and future mean diurnal cycles.) Using the RCP 8.5 scenario by end of century (2070-2100), we find that the normalized differences are small, on the order of 1-2%. However, although small, the projected seasonal changes do have a consistent geographical signature in the multi-model ensemble average (MMEA), as shown in Figure 19. The future diurnal cycles are most different from historical in autumn (September-October-November), and generally larger in the southern part of the state, particularly along the Southern California coast.

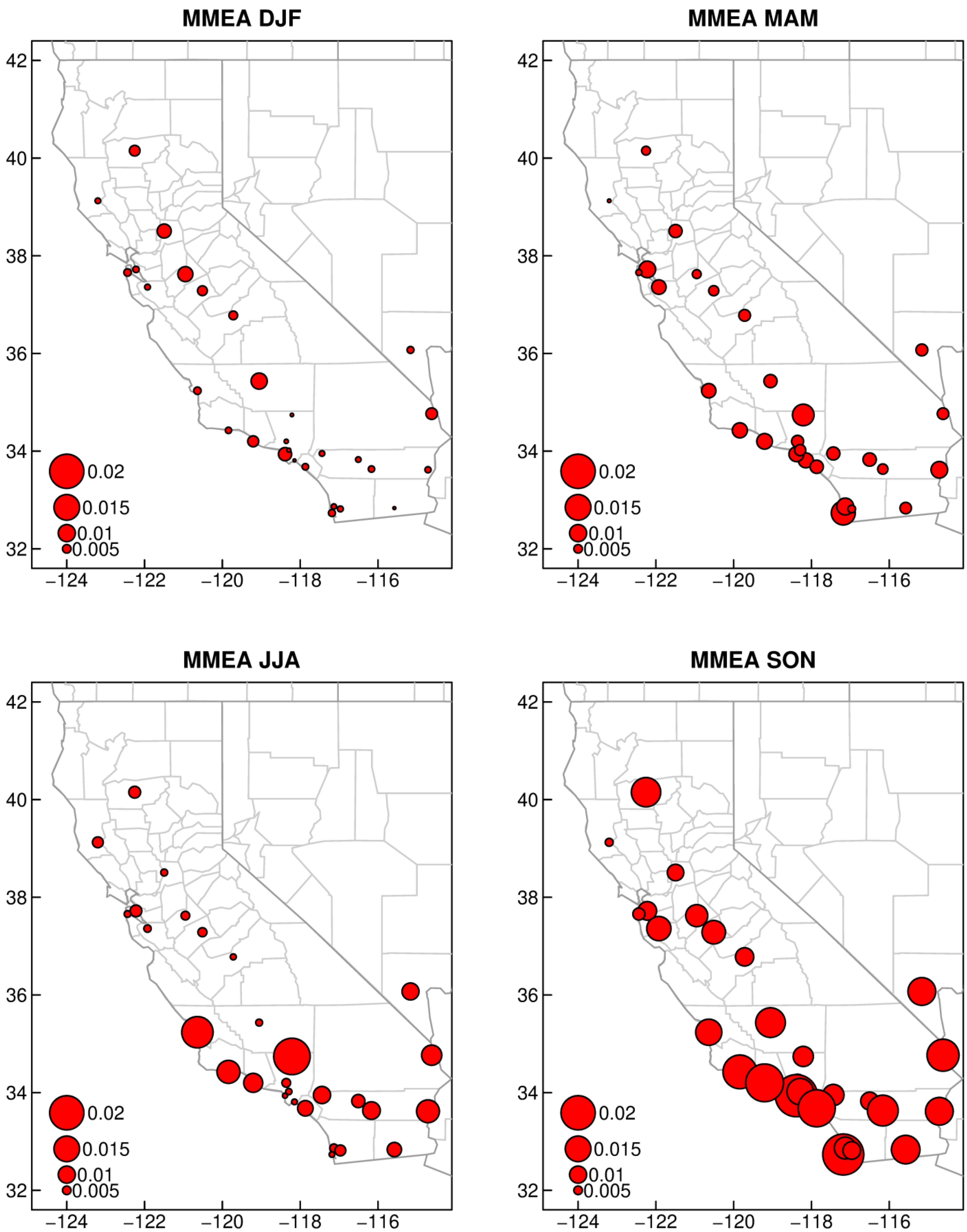

Figure 19. Maximum absolute difference between the normalized future and historical diurnal cycles, by season. Values shown are the multi-model ensemble average (MMEA) across all 10 models.

Given that the projected future differences are small, it is worth asking whether this result is consistent across models or arises from noise across the 10 models that does not quite cancel to zero in the MMEA. This can be evaluated by forming the so-called pseudo-PC, i.e., the projection (or dot product) of each individual model's anomaly pattern on the MMEA anomaly pattern. If the individual models disagree with the MMEA result, then the pseudo-PC values will be scattered around zero. Results for each model are shown in Figure 20. All but one of the models positively express the MMEA pattern, which suggests that the projected changes in diurnal cycle, although small, are generally consistent across models and not the result of noise. The two Hadley center models (HadGEM2-ES and HadGEM2-CC) show the pattern the most, with ranks 1 and 2, while the two NCAR-derived models (CCSM4 and CESM1-BGC) show it the least, with ranks 9 and 10. The fact that models with similar physics have similar results is consistent with the idea that model physics underlie the projected changes in the shape of the diurnal cycle, rather than random noise.

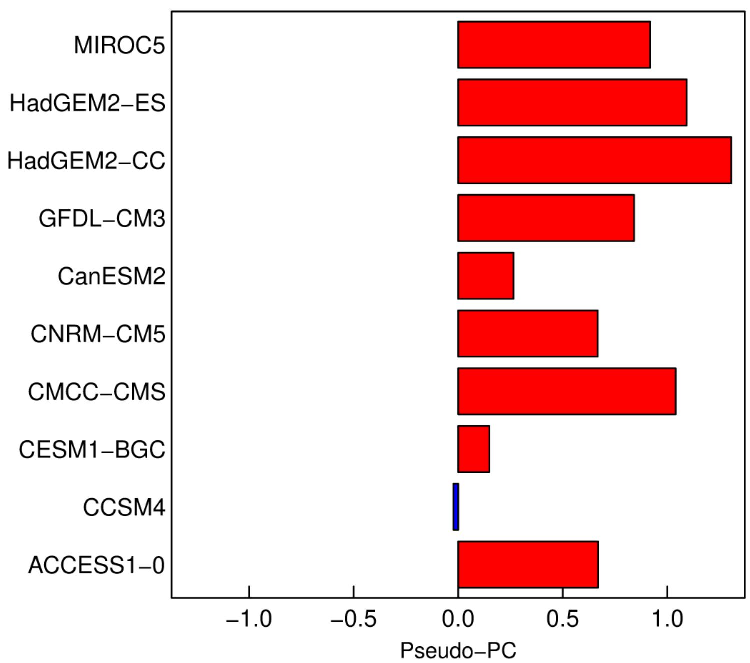

Figure 20. Pseudo-PCs calculated by projecting (forming the dot product) of each model's deseasonalized spatial-temporal pattern of maximum change in diurnal cycle with the multi-model ensemble average result. Values are normalized such that the projection of the MMEA pattern upon itself is 1.

5. Discussion

GCMs do not forecast actual temperatures on a particular day in the future – say, July 1st, 2070 – rather, they strive to simulate the statistics of temperature on future days (mean, variance, exceedance thresholds, etc.). In other words, the actual projected temperature values are not relevant to the applications of interest here, only their statistical properties. The analog day method captures the statistical properties we examined well, and noticeably better than did the climatological curve approach, which was (by construction) deficient in hourly timescale variability.

There are other applications of hourly temperature data that require the actual values. An example is forensic or accident investigation, where the temperature (say, above or below freezing) at a specific hour and day in the past may be of interest. If only Tmin and Tmax were observed on that day, could the hourly values be reliably inferred by the analog matching technique developed here?

We can evaluate this by calculating the RMSE (with respect to the hourly observations) of the disaggregated hourly temperatures computed using the cross-validation of the historical data, as described in section 2, since we have both the observed hourly temperatures and the values disaggregated from observed Tmin and Tmax. We find that the RMSE of the climatological curve and matching analog day approaches is similar. This suggests that there is insufficient information in the sequence of Tmin and Tmax values to reliably estimate hourly values on any particular day. The best that can be done is to properly simulate the statistics of the hourly values, which we have shown here the analog matching method does well.

A related question is whether the Tmin and Tmax values saved in the model projections are sufficient to capture future changes in the diurnal cycle on an hourly time scale. We have already shown (section 4.2) that the projections exhibit a consistent future change in the diurnal cycle, but it is small (~2%). Given that the RMSE results show that information on actual hourly temperature values is not completely contained in the Tmin and Tmax sequences, we can infer that the diurnal cycle changes shown in section 4.2 may be underestimated. Our results suggest that temperatures must be recorded more often than the twice a day resolution afforded by saving Tmin and Tmax to fully answer this question.

6. Conclusions

Future increases in hourly near-surface air temperature due to climate change are likely to impact areas such as agriculture, energy demand, and human health. A major impediment to understanding such future changes is that global climate models (GCMs) used to project future climate changes have not often saved hourly temperatures, likely due to the volume of data this would entail, instead generally opting to save daily minimum and maximum temperature (Tmin and Tmax).

In this work we have developed and evaluated a method using observed matching analog days to construct an hourly temperature sequence from a model's daily Tmin and Tmax. To construct hourly values for a given model day, the three-day sequence of Tmin and Tmax from the model (centered on the day being processed) is used to find the best matching observed 3-day sequence of Tmin and Tmax. The actual hourly values of the central matching observed day, scaled by the model's Tmin and Tmax on the day being processed, is the hourly time sequence used.

We compared this "analog matching day" approach to a more traditional approach consisting of constructing climatological curves of hourly diurnal temperature variability by season and quartile of the diurnal temperature range. As one would expect, the climatological curve approach is deficient in hourly variability. By contrast, the analog matching day method captures the observed hourly variability quite well, as evaluated by the spectrum, histograms of the warmest and coldest hours of the day (by season), and how often any particular hour *A* is warmer than another hour *B*.

The analog matching day method was applied to time series at 29 meteorological stations extracted from 10 GCMs well suited to future climate projections over California, with simulations starting in 1950 and extending to the end of the current century using two greenhouse gas emissions scenarios, one moderate (RCP 4.5) and one business-as-usual (RCP 8.5). The stations analyzed are used for energy demand forecasting in the state.

The projected hourly data shows strong future changes in the average number of hours per year that are likely to fall above and below defined thresholds. For example, a 45 °F threshold can be indicative of chill hours necessary for important fruit and nut crops in California. The results show strong declines in chill hours at most locations in the state, with reductions of 300 to 900 chill hours per year (out of a historical range of roughly 800-1200), depending on location, time horizon, and greenhouse gas emissions scenario. On the other hand, the average number of hours per year above thresholds of 100 and 110 °F increase substantially at most stations, with some locations showing an order of magnitude increase. This would have a significant impact on energy demand for cooling, which historically has increased about 2% per °F at the large utilities in the state. Under the RCP 8.5 (high) greenhouse gas emissions scenario, by the end of this century there are projected to be more hours above 110 °F at many locations than were above 100 °F at the end of the 20^{th} century.

The models show a consistent indication of future changes in the shape of the diurnal cycle, with changes largest in the autumn, especially along the Southern California coast. The changes are small (on the order of 2% of the amplitude of the diurnal cycle by the end of this century in the RCP 8.5 scenario), but analysis of the cross-validated results over the historical period suggest that this change may be underestimated. More frequent temperature data than the twice-daily Tmin, Tmax values is needed for further progress on this front.

Acknowledgements

This work was funded under a California Energy Commission grant for the project “Development of hybrid downscaling techniques for climate scenario modeling to improve projections for the renewable sources of energy and energy infrastructure.” Their support is gratefully acknowledged.

References

Pierce, D. W., D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using localized constructed analogs (LOCA). J. Hydrometeorology, v. 15, p. 2558, doi:10.1175/JFM-D-14-0082.1

Pierce, D. W., J. F. Kalansky, and D. R. Cayan, 2018: Climate, drought, and sea level rise scenarios for California's fourth climate change assessment. California's Fourth Climate Change Assessment, California Energy Commission and California Natural Resources Agency. 78 pp.

California Department of Water Resources Climate Change Technical Advisory Group (DWR-CCTAG), 2015: Perspectives and guidance for climate change analysis. California Department of Water Resources report, 142 pp.