LOCA2 Projections of Hourly Surface Temperatures for Stations in California

Peter Yao, David W. Pierce, Daniel R. Cayan
Division of Climate, Atmospheric Sciences, and Physical Oceanography
Scripps Institution of Oceanography, UC San Diego
February 23, 2024

1) Introduction

Hourly temperature changes are crucial for understanding energy demand and other areas such as agriculture and human health. To address the need for this information, a set of future projections of hourly temperatures from 1950-2100 was developed for 32 key meteorological stations throughout California (plus one station in Nevada) that are used by the energy sector for understanding and forecasting the state's energy demand.

The Localized Constructed Analogs (LOCA) statistical downscaling method (Pierce et al. 2014; Pierce et al. 2015a; Pierce et al. 2015b) forms the basis for these projections. An analog matching method is applied to the daily LOCA maximum and minimum temperatures (Tmax and Tmin) to construct the hourly values, using hourly observations from each station as the training data. This process produces values that better represent hourly variability compared to a traditional approach that uses an assumed climatological diurnal cycle. Data from 15 CMIP6 GCMs are included in this dataset, with between one to ten ensemble members per GCM (129 total simulations for each station). The data includes the CMIP6 historical period from 1950-2014 and up to three different Shared Socioeconomic Pathway (SSP) scenarios for the period 2015-2100, depending on the GCM and ensemble member: SSP245, SSP370, and SSP585, which are roughly medium-low, medium, and high emissions scenarios.

The projections developed for California's Fourth Climate Assessment (Pierce and Cayan 2019) first introduced the analog matching method for generating the hourly values based on data from the CMIP5 GCM archive. In this updated version based on CMIP6 GCMs, the analog matching process remains fundamentally the same; however, this new version includes several key improvements. First, the observed station temperature data, or training data used to bias correct the LOCA data and provide analog days for the matching method, has been updated to a new dataset which includes four additional stations (all in Northern California) and a longer period of record for most stations. Second, the climate model data that forms the basis of the future projections is now the Localized Constructed Analogs version 2 (LOCA2 hereafter) hybrid downscaled product (Pierce et al. 2023a), which features several notable changes as well as an expanded set of available ensemble members and future scenarios. Third, the station adjustment process has been improved to account for the difference in typical hourly progression of temperatures between wet vs. dry days. And finally, the analog matching process has been updated to ensure that the daily maxima and minima of the final hourly output exactly matches the input daily Tmax and Tmin.

This report will primarily focus on what is new in this version, including a detailed description of the updated methodology and verification of the output. Since the analog matching process is largely unchanged from the previous version, please refer to the previous report (Pierce and Cayan 2019) for the complete details of how it operates.

2) Observed station temperature data and quality control process

2.1) Station overview

Projections of hourly near-surface air temperature were developed for 33 stations selected by the California Energy Commission (CEC), including 32 stations in California and one in Nevada. Table 1 below summarizes the station metadata, and Figure 1 shows their locations. 29 stations match those from the previous version of temperature projections (highlighted in blue in Figure 1), while four stations are new additions in this version (highlighted in red). The four new stations (KACV, KRDD, KSCK, and KSMF) are all located in northern California, which was noted as an area with a lack of coverage in the previous report. Observed station data was required for two separate steps of the process (station adjustment and analog matching), which are detailed below.

Station ID	Lat	Lon	State	Name
KACV	40.978	-124.105	CA	ARCATA AIRPORT
KBFL	35.434	-119.055	CA	MEADOWS FIELD AIRPORT
KBLH	33.619	-114.715	CA	BLYTHE AIRPORT
KBUR	34.2	-118.365	CA	BURBANK-GLENDALE-PASA ARPT
KCQT	34.024	-118.291	CA	DOWNTOWN L.A./USC CAMPUS
KEED	34.768	-114.618	CA	NEEDLES AIRPORT
KFAT	36.78	-119.72	CA	FRESNO YOSEMITE INTERNATIONAL AIRPORT
KIPL	32.834	-115.579	CA	IMPERIAL CO
KLAS	36.072	-115.163	NV	MCCARRAN INTERNATIONAL AIRPORT
KLAX	33.938	-118.387	CA	LOS ANGELES INTERNATIONAL AIRPORT
KLGB	33.812	-118.147	CA	LONG BEACH / DAUGHERTY FIELD / AIRPORT
KMCE	37.285	-120.514	CA	MERCED MUNI MACREADY
KMOD	37.625	-120.955	CA	MDSTO CTY-CO H SHAM FD APT
KNKX	32.867	-117.133	CA	SAN DIEGO MIRAMAR NAS
KOAK	37.718	-122.233	CA	METRO OAKLAND INTL AIRPORT
KOXR	34.2	-119.204	CA	OXNARD AIRPORT
KPSP	33.833	-116.5	CA	PALM SPRINGS INTL
KRAL	33.95	-117.433	CA	RIVERSIDE MUNI
KRBL	40.152	-122.255	CA	RED BLUFF MUNICIPAL ARPT
KRDD	40.518	-122.299	CA	REDDING MUNICIPAL ARPT
KSAC	38.507	-121.496	CA	SACRAMENTO EXECUTIVE AIRPORT
KSAN	32.734	-117.183	CA	SAN DIEGO INTERNATIONAL AIRPORT
KSBA	34.424	-119.842	CA	SANTA BARBARA MUNICIPAL AIRPORT
KSBP	35.233	-120.633	CA	SAN LUIS CO RGNL
KSCK	37.89	-121.226	CA	STOCKTON METROPOLITAN AIRPORT
KSEE	32.833	-116.967	CA	GILLESPIE FLD
KSFO	37.62	-122.366	CA	SAN FRANCISCO INTERNATIONAL AIRPORT
KSJC	37.359	-121.924	CA	N Y. MINETA SN JO INTL APT
KSMF	38.701	-121.595	CA	SACRAMENTO INTL AIRPORT
KSNA	33.68	-117.867	CA	J. WAYNE APT-ORANGE CO APT

KTRM	33.632	-116.164	CA	DESERT RESORTS RGNL ARPT
KUKI	39.128	-123.2	CA	UKIAH MUNICIPAL AIRPORT
KWJF	34.741	-118.213	CA	GENERAL WILLIAM J. FOX AIRFIELD AIRPORT

Table 1. Table of the 33 stations included in the dataset.

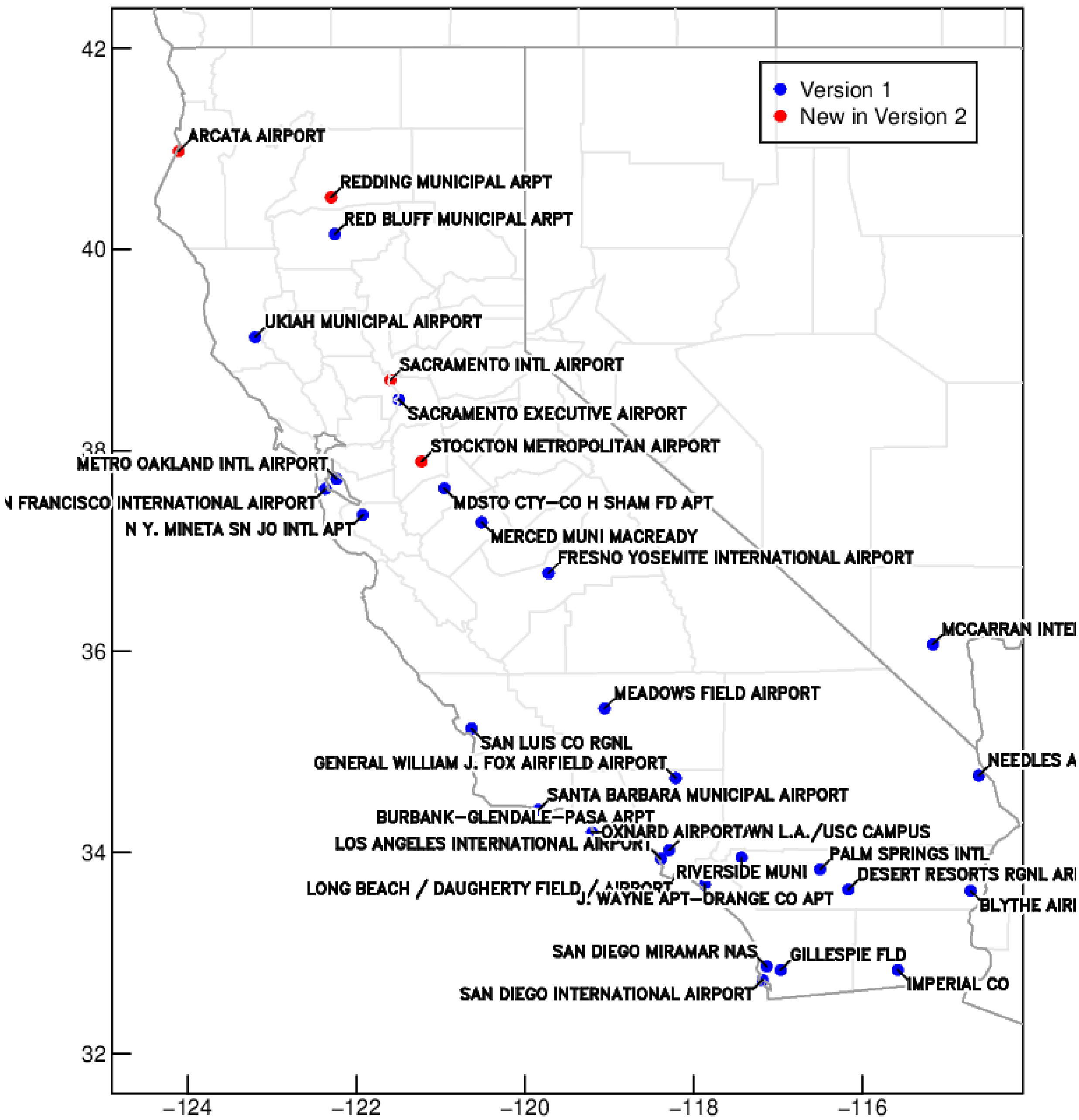

Figure 2. Map of the 33 stations included in the dataset. Stations in blue were included in the original temperature projections dataset (Pierce and Cayan, 2019). Stations in red are new additions to this version.

2.2) Overview of selected training datasets

Two different observational datasets were used for separate stages of the process: the Global Historical Climatology Network daily (GHCNd) dataset (Menne et al. 2012) was used for the station adjustment step, while HadISD, which is a global sub-daily dataset based on the ISD dataset from NOAA's NCEI (Dunn et al. 2012), was used for the analog matching process.

Two different training datasets were used because of the different requirements between the station adjustment and analog matching steps. The station adjustment process requires daily Tmax, Tmin, and precipitation values during as much of the LOCA2 historical period (1950-2014) as possible, so GHCNd was selected because it provides a more complete and reliable record of daily values than HadISD, and because HadISD did not provide a reliable precipitation record. HadISD was identified for the analog matching step because it requires hourly temperature observations.

2.3) Observed daily data used for the station adjustment process

Since LOCA2 is a gridded 3-km resolution product, the LOCA2 data from the nearest grid cell to each station was then extracted. However, there are typically systematic differences between the LOCA2 climatology at the grid cell center and the observed station climatology: for example, due to differences in elevation between the station and the LOCA2 grid cell, or in regions with a strong temperature gradient between adjacent grid cells, such as in coastal areas. The LOCA2 and station climatologies may also differ due to systematic differences between the station data and the original gridded temperature data used to train LOCA2.

To reduce these differences, a station adjustment step was performed on the LOCA2 data: for each month-of-year, this process additively adjusted the daily LOCA2 temperature climatology to match that month's observed station climatology during the LOCA2 historical period (1950-2014), and multiplicatively adjusted the daily LOCA2 anomalies to match the observed station standard deviation of anomalies. The adjustment was also performed separately on wet (precip ≥ 0.5 mm) vs. dry (precip < 0.5 mm) days to account for systematic differences in typical hourly progression of temperatures on days when there is precipitation compared to days when there is not. The station adjustment method is described in greater detail in section 3 below.

This process requires a quality-controlled daily dataset of observed Tmax, Tmin, and precipitation at each station. Here we chose to use the Global Historical Climatology Network daily (GHCNd) dataset as the training data. GHCNd was chosen primarily because of the introduction of the distinction between wet and dry days during the station adjustment step, which is new to this version. We originally tried to use the hourly resolution HadISD dataset for both the station adjustment and analog matching steps. However, this dataset was found to have unreliable precipitation data, which is required for the station adjustment but not to perform the analog matching.

The GHCNd data is already quality controlled, although several additional QA/QC steps were performed: any days with a quality control flag were excluded, as well as any days when Tmax did not exceed Tmin. Then, for each station, only observed days with valid Tmax, Tmin, and precipitation values during the LOCA2 historical period (1950-2014) were included.

Table 2 shows the GHCNd station IDs as well as the period of record for each station, represented as the starting and ending years during the LOCA2 historical period (1950-2014) when data was available and the number of valid days, rounded into years. A valid day was defined as a day within the LOCA2 historical period when the observed data meets all quality control checks and has valid values of Tmax, Tmin, and precipitation. The total number of valid days was summed for each station and then divided by 365 to approximate the number of valid years of training data used for the station adjustment.

Station ID	Name	GHCNd ID	Start Year	End Year	Valid Data (Years)
KLAX	LOS ANGELES INTERNATIONAL AIRPORT	USW00023174	1950	2014	65
KBFL	MEADOWS FIELD AIRPORT	USW00023155	1950	2014	65
KCQT	DOWNTOWN L.A./USC CAMPUS	USW00093134	1950	2014	65
KFAT	FRESNO YOSEMITE INTERNATIONAL AIRPORT	USW00093193	1950	2014	65
KSFO	SAN FRANCISCO INTERNATIONAL AIRPORT	USW00023234	1950	2014	65
KLAS	MCCARRAN INTERNATIONAL AIRPORT	USW00023169	1950	2014	65
KSAC	SACRAMENTO EXECUTIVE AIRPORT	USW00023232	1950	2014	65
KSCK	STOCKTON METROPOLITAN AIRPORT	USW00023237	1950	2014	65
KSAN	SAN DIEGO INTERNATIONAL AIRPORT	USW00023188	1950	2014	65
KBLH	BLYTHE AIRPORT	USW00023158	1950	2014	64.5
KLGB	LONG BEACH / DAUGHERTY FIELD / AIRPORT	USW00023129	1950	2014	64.5
KTRM	DESERT RESORTS RGNL ARPT	USW00003104	1950	2014	64.3
KMOD	MDSTO CTY-CO H SHAM FD APT	USW00023258	1950	2014	64.3
KEED	NEEDLES AIRPORT	USW00023179	1950	2014	64.2
KRBL	RED BLUFF MUNICIPAL ARPT	USW00024216	1950	2014	62.9
KSBA	SANTA BARBARA MUNICIPAL AIRPORT	USW00023190	1950	2014	62.3
KNKX	SAN DIEGO MIRAMAR NAS	USW00093107	1950	2014	60.7
KPSP	PALM SPRINGS INTL	USC00046635	1950	2014	59.8
KOAK	METRO OAKLAND INTL AIRPORT	USW00023230	1950	2014	48.3
KWJF	GENERAL WILLIAM J. FOX AIRFIELD AIRPORT	USW00003159	1974	2014	40.3
KIPL	IMPERIAL CO	USW00003144	1962	2014	34
KRDD	REDDING MUNICIPAL ARPT	USW00024257	1986	2014	28.4
KSEE	GILLESPIE FLD	USC00043410	1959	1979	19.4
KSBP	SAN LUIS CO RGNL	USW00093206	1998	2014	16.7
KACV	ARCATA AIRPORT	USW00024283	1992	2014	16.7
KOXR	OXNARD AIRPORT	USW00093110	1998	2014	16.6
KSMF	SACRAMENTO INTL AIRPORT	USW00093225	1998	2014	16.6
KBUR	BURBANK-GLENDALE-PASA ARPT	USW00023152	1998	2014	16.5
KSJC	N Y. MINETA SN JO INTL APT	USW00023293	1998	2014	16.4

KMCE	MERCED MUNI MACREADY	USW00023257	1998	2014	16.3
KSNA	J. WAYNE APT-ORANGE CO APT	USW00093184	1999	2014	15.8
KRAL	RIVERSIDE MUNI	USW00003171	1998	2014	15.4
KUKI	UKIAH MUNICIPAL AIRPORT	USW00023275	1954	2014	14

Table 2. Table of the GHCNd stations sorted by the length of valid data during the LOCA2 historical period (1950-2014).

2.4) Observed hourly data used for the analog matching process

Next, the analog matching process transforms the station adjusted daily LOCA2 Tmax and Tmin data into hourly temperatures. Briefly, this process (referred to as the hourly disaggregation) works by stepping through each three-day period in the LOCA2 data and finding the observed three-day sequence of Tmin and Tmax that best matches the model's sequence. The hourly values from the middle day of the observed sequence (the analog day) are then used to form the hourly output. A melding step is performed to smooth the hourly transitions between days occurring at local midnight, and the final output is scaled to match the input daily LOCA2 Tmax and Tmin.

HadISD was selected as the training data for this step because it requires hourly temperature observations. Because the HadISD data forms the "pattern library" for the analog matching process, it is important that the data is quality controlled and includes as long a period of record as possible. A long period of record is desirable here to have as many possible analog days to choose from, and the period of record can extend beyond the LOCA2 historical period because the final hourly output is scaled to match the input LOCA2 daily Tmax and Tmin. The HadISD data was quality controlled to exclude erroneous or unrealistic values, and days with missing values are excluded from the hourly disaggregation process. The quality controlled HadISD dataset was provided to us by Eagle Rock Analytics.

In addition to the four new stations in northern California, one notable improvement in this version is the longer period of record of possible analog days for most stations. In the previous version, the observed data covered the period from Jan 1, 2000 through Dec 31, 2018, equaling 19 years of data. As shown in Table 3, there are more than 19 years of valid analog days for most stations in this version (all but 6 stations). A valid analog day is defined as a day in the observed dataset when all hourly values are valid for that day as well as during the preceding and following days (because the analog matching is conducted in 3-day sequences).

Station ID	Name	Start Year	End Year	Valid Data (Years)
KLAS	MCCARRAN INTERNATIONAL AIRPORT	1948	2022	71.1
KBFL	MEADOWS FIELD AIRPORT	1941	2022	66.1
KSAN	SAN DIEGO INTERNATIONAL AIRPORT	1942	2022	64.5
KSFO	SAN FRANCISCO INTERNATIONAL AIRPORT	1948	2022	59.9
KFAT	FRESNO YOSEMITE INTERNATIONAL AIRPORT	1941	2022	59.4
KRBL	RED BLUFF MUNICIPAL ARPT	1948	2022	58.3
KBUR	BURBANK-GLENDALE-PASA ARPT	1943	2022	54.9
KSAC	SACRAMENTO EXECUTIVE AIRPORT	1947	2022	52.5
KNKX	SAN DIEGO MIRAMAR NAS	1948	2022	52.3

KOAK	METRO OAKLAND INTL AIRPORT	1943	2022	52.1
KLGB	LONG BEACH / DAUGHERTY FIELD / AIRPORT	1943	2022	50.5
KBLH	BLYTHE AIRPORT	1942	2022	50.3
KSCK	STOCKTON METROPOLITAN AIRPORT	1941	2022	49
KTRM	DESERT RESORTS RGNL ARPT	1943	2022	43.4
KWJF	GENERAL WILLIAM J. FOX AIRFIELD AIRPORT	1974	2022	42.4
KSBA	SANTA BARBARA MUNICIPAL AIRPORT	1945	2022	40.8
KLAX	LOS ANGELES INTERNATIONAL AIRPORT	1944	2022	38.5
KIPL	IMPERIAL CO	1973	2022	37.9
KEED	NEEDLES AIRPORT	1948	2022	36.1
KUKI	UKIAH MUNICIPAL AIRPORT	1949	2022	35
KSMF	SACRAMENTO INTL AIRPORT	1973	2022	31.4
KSJC	N Y. MINETA SN JO INTL APT	1968	2022	31.3
KRDD	REDDING MUNICIPAL ARPT	1944	2022	28.6
KOXR	OXNARD AIRPORT	1944	2022	27.4
KPSP	PALM SPRINGS INTL	1943	2022	24.7
KACV	ARCATA AIRPORT	1950	2022	22.3
KCQT	DOWNTOWN L.A./USC CAMPUS	1999	2022	19.6
KSNA	J. WAYNE APT-ORANGE CO APT	1942	2022	18.5
KMOD	MDSTO CTY-CO H SHAM FD APT	1998	2022	18.4
KRAL	RIVERSIDE MUNI	1998	2022	17.9
KMCE	MERCED MUNI MACREADY	1998	2022	15.6
KSBP	SAN LUIS CO RGNL	1998	2022	13.4
KSEE	GILLESPIE FLD	1973	2022	5.9

Table 3. Table of the HadISD stations sorted by the number of valid analog days.

3) Station adjustment

3.1) Station adjustment method

As described earlier, a station adjustment step is performed on the daily LOCA2 Tmax and Tmin data to match the LOCA2 climatology and standard deviation of anomalies during the LOCA2 historical period to those of the observed station data. Because the timeseries from the nearest grid cell to each station is extracted from the gridded LOCA2 projections, this step helps to account for any systematic differences between the LOCA2 climatology at the grid cell center and the observed station climatology. The climatology can be thought of as the mean historical temperature for each month-of-year, while the standard deviation of temperature anomalies, or departures from the mean, represents the historical monthly spread. Both quantities are adjusted in the LOCA2 data to match the observations more closely.

The first step in the process is to calculate the observed climatology from the training data (GHCNd) for each station. Means of Tmax and Tmin are calculated for each month-of-year, only including valid days during the LOCA2 historical period only (1950-2014). The monthly values are then converted to a day-of-year climatology via a simple linear interpolation: the monthly means are assumed to be the daily value for the 15th day of each month, and then the remaining daily values are linearly interpolated. The climatology values on February 28 and March 1 are averaged whenever the correction is applied to a leap day (February 29).

The climatology is calculated separately for wet (precip ≥ 0.5 mm) vs. dry (precip < 0.5 mm) days. The top row in Figure 2 shows example GHCNd climatologies at KLAX (Los Angeles, CA): from left to right, the climatology for all days, dry days only, and wet days only.

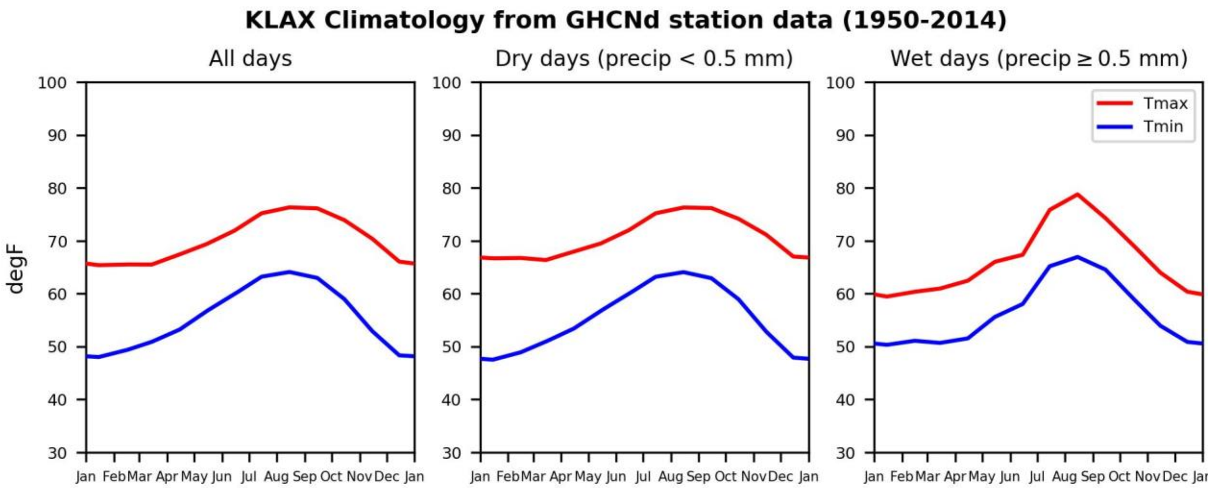

Figure 2. Left to right: climatology for all days, dry days, and wet days at KLAX (Los Angeles, CA). Tmax in red, Tmin in blue.

The same method is then repeated for the standard deviation of Tmax and Tmin anomalies. The anomalies are first calculated, which are the departures from the day-of-year climatology for each daily value. The standard deviation is then calculated for each month-of-year, and the day-of-year values are constructed in the same manner as the climatology. An example of the GHCNd standard deviation of anomalies at KLAX is shown in Figure 3.

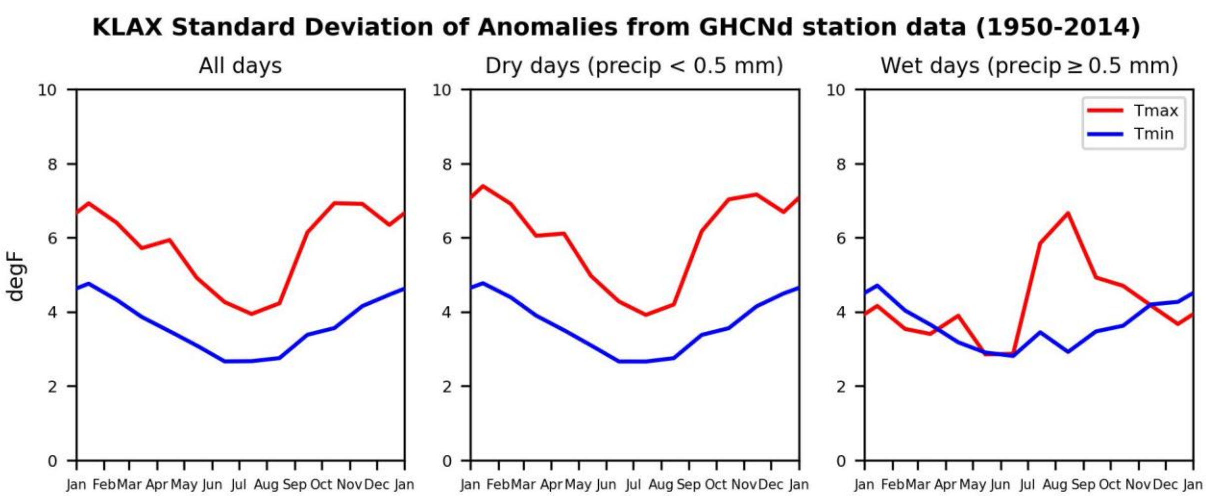

Figure 3. Left to right: standard deviation of anomalies for all days, dry days, and wet days at KLAX (Los Angeles, CA). Tmax in red, Tmin in blue.

3.2) Summer wet day correction

There are cases when there are very few wet days during a given month-of-year in an observed station record, typically during the summer months (JJA) at drier station locations. For example, there is only one wet day during the month of July at KSJC (San Jose, CA) from the entire GHCNd dataset spanning 1998-2014. Constructing a climatology based on a small sample size such as in this example can lead to large fluctuations between the monthly values (see Figure 4 for an example).

Thus, a summer wet day correction was applied: in cases when there were very few (<10) wet days for a given month-of-year across the historical record, a wet day climatology was constructed by starting with the climatology for all days (wet and dry) and subtracting the average difference between the all day and wet day climatologies during non-winter months (Mar-Nov). This process reasonably recreates what the climatology might look like during months when there is a small sample size based on the average wet day climatology across all non-winter months. An example of the correction is illustrated in Figure 4, which shows the wet day climatology before (dotted line) and after the correction (solid line) at KSJC (San Jose, CA).

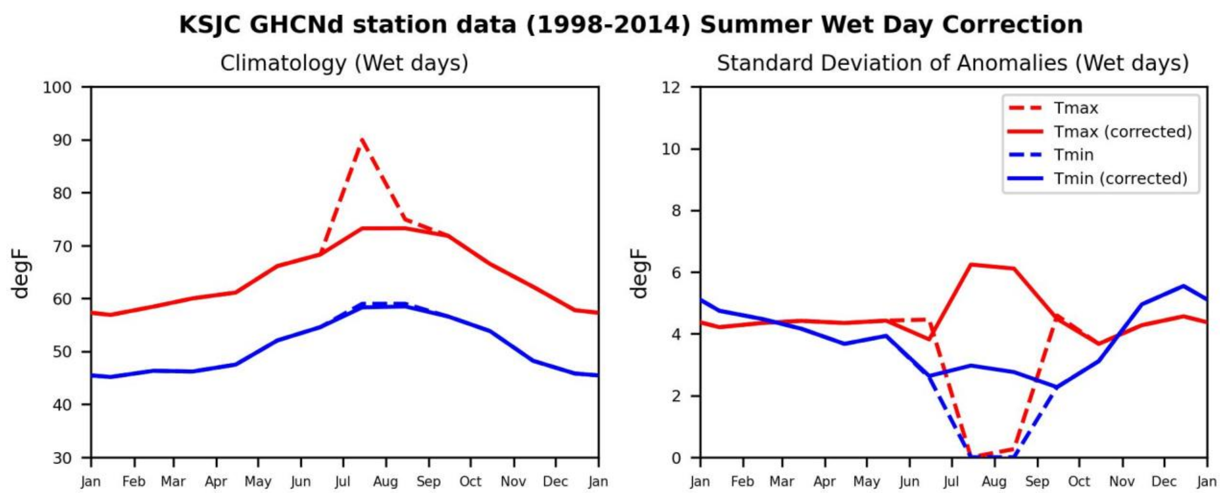

Figure 4. Left: climatology on wet days before (dotted line) and after (solid line) the summer wet day correction at KSJC (San Jose, CA). Right: standard deviation of anomalies on wet days before/after correction. Tmax in red, Tmin in blue.

A similar correction was applied to the standard deviation of anomalies: in cases when there were fewer than 10 wet days in a given month-of-year, the value for all days (wet and dry) was used instead. In the same example at KSJC, the standard deviation of anomalies for July is zero before the correction since it is based on only one value (see Figure 4).

Table 4 shows a list of stations for which the correction was applied, including the number of observed wet days during each month from May-September. No stations required a correction outside of these months. Note that just because there is a small sample size of summer wet days for a given station in the observations, that this does not necessarily reflect what the frequency of summer wet days might look like in the future as projected by LOCA2.

Station ID	May	Jun	Jul	Aug	Sep
KBFL	94	27	7	12	52
KBLH	16	5	85	118	81
KBUR	16	6	0	2	4
KCQT	74	21	9	21	52
KFAT	130	36	7	11	48
KIPL	4	1	18	31	22
KMCE	35	9	1	0	9
KMOD	131	46	9	17	54
KOXR	24	5	2	2	12
KPSP	19	7	35	52	47
KRAL	12	3	6	6	4
KSAC	166	64	8	22	71
KSBP	33	11	1	4	7
KSEE	39	19	4	10	18
KSFO	166	55	7	21	61
KSJC	33	10	1	2	10
KSMF	54	16	2	3	13
KSNA	17	5	15	2	6
KTRM	18	4	44	55	61
KUKI	48	23	3	0	17
KWJF	33	8	21	16	32

Table 4. Table of the GHCNd stations requiring the summer wet day correction. Values are the number of observed wet days (precip ≥ 0.5 mm) in a given month-of-year from May to September. Months with fewer than 10 observed days are highlighted in red.

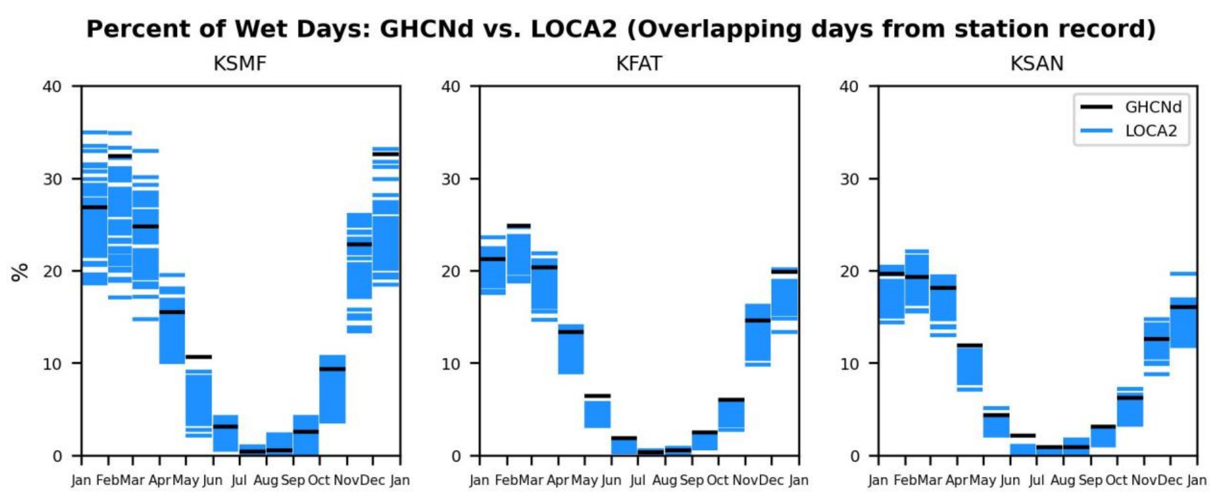

Figure 5. Percent of wet days by month of the year at 3 selected stations. From left to right: KSMF (Sacramento, CA), KFAT (Fresno, CA), and KSAN (San Diego, CA). GHCNd data in black, each LOCA2 historical simulation (only overlapping days from the station record during 1950-2014) represented as one blue line.

The wet day frequency is overall similar between the observations (GHCNd) and the LOCA2 historical data. Figure 5 shows a comparison of the percent of wet days for each month-of-year between GHCNd and the overlapping station record from each LOCA2 historical simulation at 3 example stations, one each in Northern, Central, and Southern California: KSMF (Sacramento, CA), KFAT (Fresno, CA), and KSAN (San Diego, CA). The model error varies by station, with errors generally being higher in the winter months.

3.3) Applying the station adjustment to the LOCA2 data

Once the training data is prepared, the station adjustment can be applied to each LOCA2 simulation (129 total per station). For each GCM, this step adjusts the ensemble mean climatology and standard deviation of anomalies (additively for the climatology and multiplicatively for the standard deviation of anomalies) based on the difference between the observed values and the LOCA2 ensemble mean. All days in the LOCA2 data (1950-2100) are adjusted based on the difference between the GHCNd data and the LOCA2 ensemble mean values during overlapping days from the station record. The result is that in the station adjusted LOCA2 data, the historical ensemble mean climatology and spread during days overlapping with the station record will match the training data. Then, days outside of the station record (e.g. any remaining historical days and all days in the future period from 2015-2100) are adjusted using the same factor. For example, if the observed Tmax climatology at a given station is 70 degF during June dry days and the ensemble mean June dry day climatology for a given GCM is 68 degF, all dry days in June for that GCM are adjusted by +2 degF. Similarly, if the observed Tmax standard deviation of anomalies is 3 degF over that same period while the ensemble mean value is 6 degF, the anomalies from all days during that matching period (e.g. June dry days) are divided by 2.

The first step is to calculate the climatology and standard deviation of anomalies for each LOCA2 simulation (each GCM, ensemble member, and SSP). The method is the same as for the GHCNd training data: the mean or standard deviation of anomalies is calculated for each month-of-year (separately for Tmax and Tmin and for wet vs. dry days) and then linearly interpolated to produce the day-of-year values. Only overlapping days between the GHCNd station record and the LOCA2 historical period (1950-2014) are included in the LOCA2 climatology and standard deviation of anomalies calculation. This ensures that the observed data availability does not affect the correction. For example, the GHCNd station record at KOXR (Oxnard, CA) only extends from 1998-2014 during the LOCA2 historical period. Only including the days with observations at KOXR when calculating the LOCA2 climatology ensures that the historical data from 1950-1997 is not adjusted based on non-existent observations during those years.

Although the climatology and standard deviation of anomalies is calculated for each LOCA2 simulation, only the ensemble mean climatology and standard deviation of anomalies for each GCM is used for the adjustment to retain the variability between ensemble members. Also note that any future temperature trend projected by LOCA2 is preserved during the station adjustment step.

Figure 6 illustrates an example of the climatology station adjustment at KSAC (Sacramento, CA). The top row shows the climatology of each LOCA2 simulation as a separate line (Tmax in red, Tmin in blue) before the adjustment, while the bottom row shows the adjusted climatologies. The GHCNd climatology is shown in black (unchanged between rows). Following the correction, the climatology of the full LOCA2 ensemble (including all GCMs) should be roughly centered on the GHCNd climatology while retaining the variability between individual simulations.

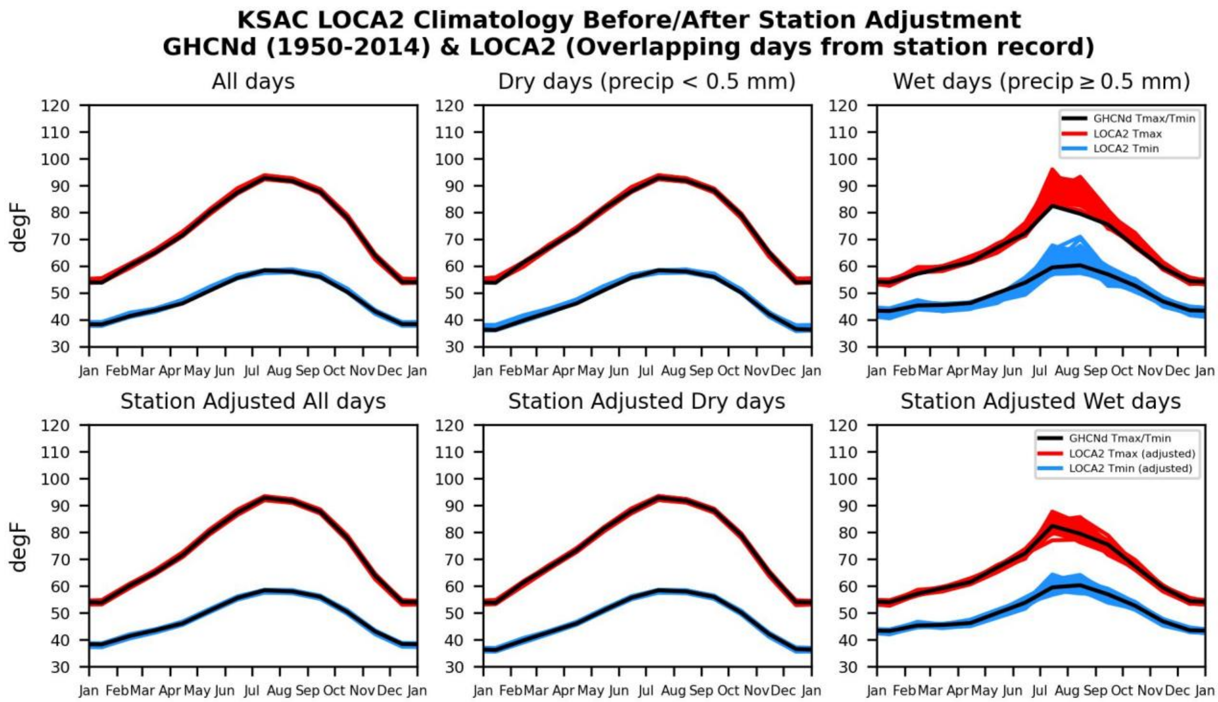

Figure 6. LOCA2 historical climatology before (top) vs. after (bottom) the station adjustment is applied at KSAC (Sacramento, CA). Left to right: climatology on all days, dry days, and wet days. Each LOCA2 line represents one simulation (e.g. one specific GCM ensemble member). GHCNd climatology in black for both Tmax and Tmin. LOCA2 Tmax in red, LOCA2 Tmin in blue.

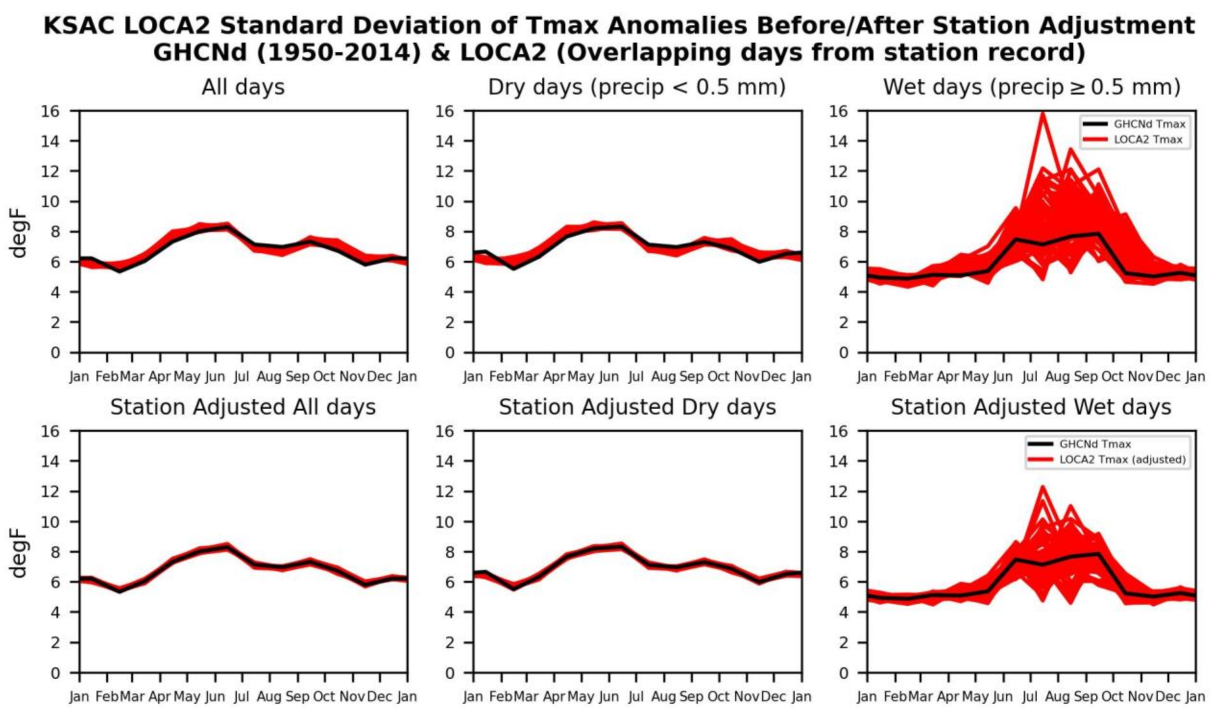

Figure 7. LOCA2 historical standard deviation of Tmax anomalies before (top) vs. after (bottom) the station adjustment is applied at KSAC (Sacramento, CA). Left to right: standard deviation of Tmax anomalies on all days, dry days, and wet days. Each LOCA2 line represents one simulation (e.g. one specific GCM ensemble member).

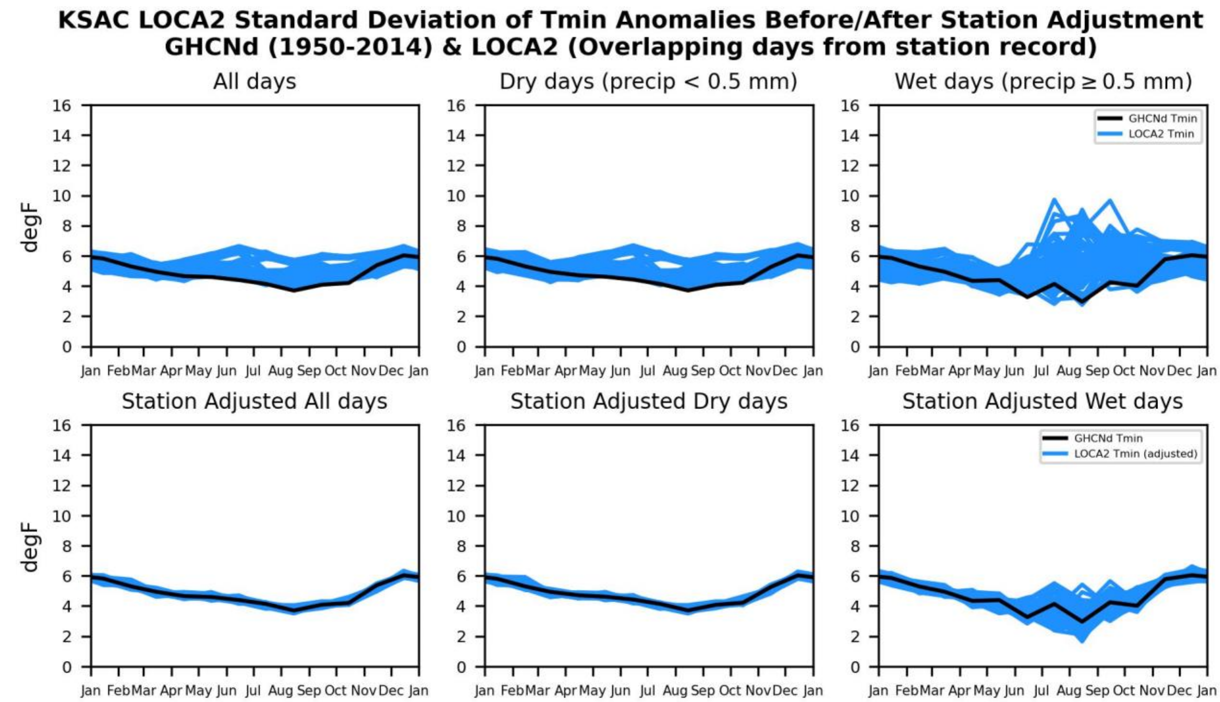

Figure 8. Same as Fig 7 but for Tmin anomalies.

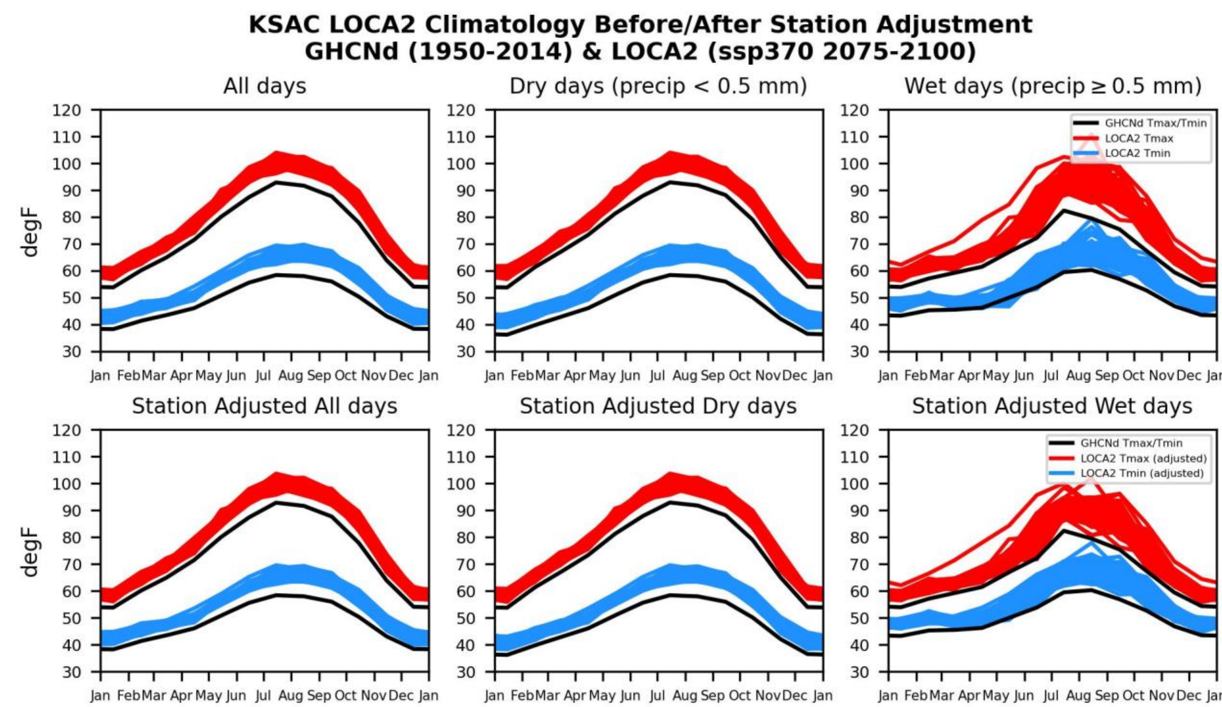

Figure 9. Same as Fig 6 but for LOCA2 ssp3370 end-of century (2075-2100) instead of the historical period.

An example of the station adjusted standard deviation of anomalies is shown in Figure 7 (Tmax anomalies) and Figure 8 (Tmin anomalies), also at KSAC. Finally, Figure 9 shows the station adjusted climatology during end-of-century (2075-2100) for the SSP370 scenario to demonstrate how the adjustment works for days outside of the historical period. Note that any trend in the climatology from the LOCA2 historical period to end-of-century is preserved following the adjustment.

3.4) DTR correction

Since Tmax and Tmin are adjusted independently, there are cases when the adjusted Tmax ≤ adjusted Tmin. In these instances, a final correction must be applied to ensure that the diurnal temperature range, or DTR, is > 0 for each day (in other words, Tmax > Tmin). A minimum DTR value is calculated for each station from the GHCNd training data: we used the 0.1 percentile DTR value for each day-of-year during the LOCA2 historical period. Then, in instances when the DTR is smaller than the minimum value, the Tmax value is retained and the final Tmin value is adjusted by subtracting the minimum DTR value from Tmax. Tmin is treated as the residual here because Tmax is of greater importance for most energy demand applications.

The distribution of adjusted LOCA2 DTR values was compared to the DTR distribution from the training data to ensure that this correction does not significantly alter the distribution. Figure 10 shows an example of the DTR distribution for the training data compared to the LOCA2 historical data before/after the station adjustment at an example station, KOAK (Oakland, CA). As shown in the example, the DTR distribution was not significantly altered by the station adjustment.

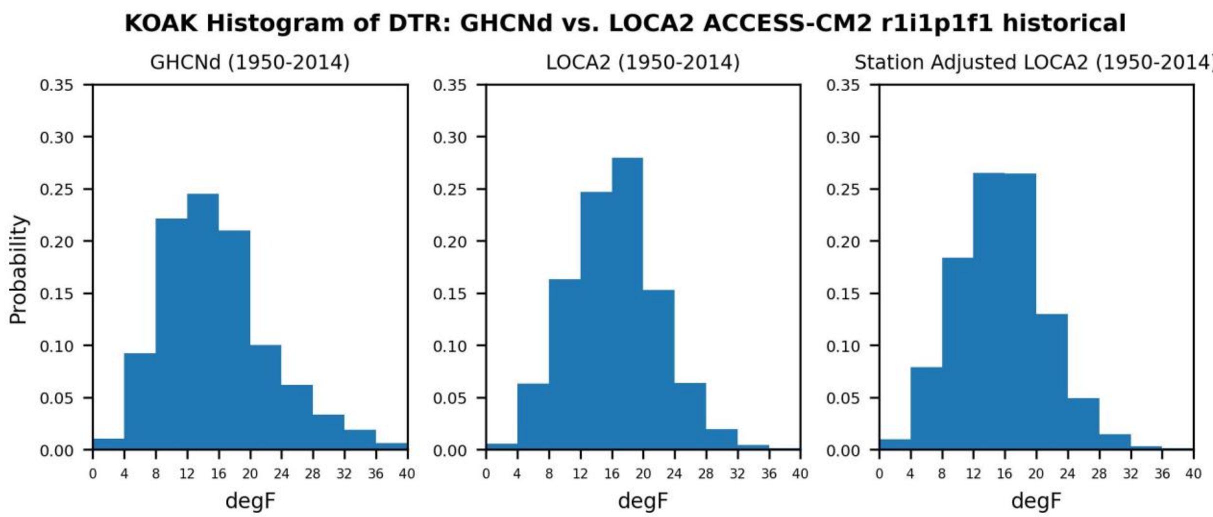

Figure 10. Histograms of DTR at KOAK (Oakland, CA) for one simulation (ACCESS-CM2 r1i1p1f1). Left to right: GHCNd data, LOCA2 daily historical data before station adjustment, LOCA2 data after the station adjustment was applied.

4) Hourly disaggregation

4.1) Hourly disaggregation method

The final step is to transform the station adjusted daily LOCA2 Tmax and Tmin into hourly temperature projections. This process, or the hourly disaggregation method, remains largely the same as in the previous version. Please refer to the previous report (Pierce and Cayan 2019) for the full details of how the method works.

Briefly, the method works by stepping through each 3-day period in each LOCA2 simulation and finding the best analog period from the training data, HadISD, based on the 3-day sequence in the station record with the lowest RMSE between the 6 Tmax/Tmin values during those days. The middle day of the observed sequence with the lowest RMSE is selected as the analog day, and those hourly values are used to construct the values for that day in the LOCA2 output. Once all the LOCA2 days are constructed by this method, a final melding step adjusts the transitions between days (at local midnight) to remove any sharp discontinuities. An example of the first week of the hourly result for one LOCA2 simulation at KSAC is shown in Figure 11. The LOCA2 daily Tmax and Tmin are shown in red and blue, respectively, while the constructed hourly output is shown in black.

There are several small differences that are new in this version. First, the way in which missing hourly values in the training data are handled is slightly different. In the previous version, any missing values were filled in with values from the ASOS dataset. In this version, when Tmax and Tmin are calculated from the HadISD data, any missing values are set to a very large number (e.g. 999) to ensure that day cannot be selected as an analog day. This step also ensures that only 3-day sequences with all 72 hourly values valid are considered when finding the best analog day.

Second, the process has been updated to ensure that the daily Tmax and Tmin in the final hourly projections exactly match those of the station adjusted LOCA2 data. In the previous version, the melding step adjusted the values such that in some cases the daily Tmax or Tmin were then altered. In this version, the hourly values for each day undergo a final scaling step to match the daily maxima and minima with the LOCA2 Tmax and Tmin data.

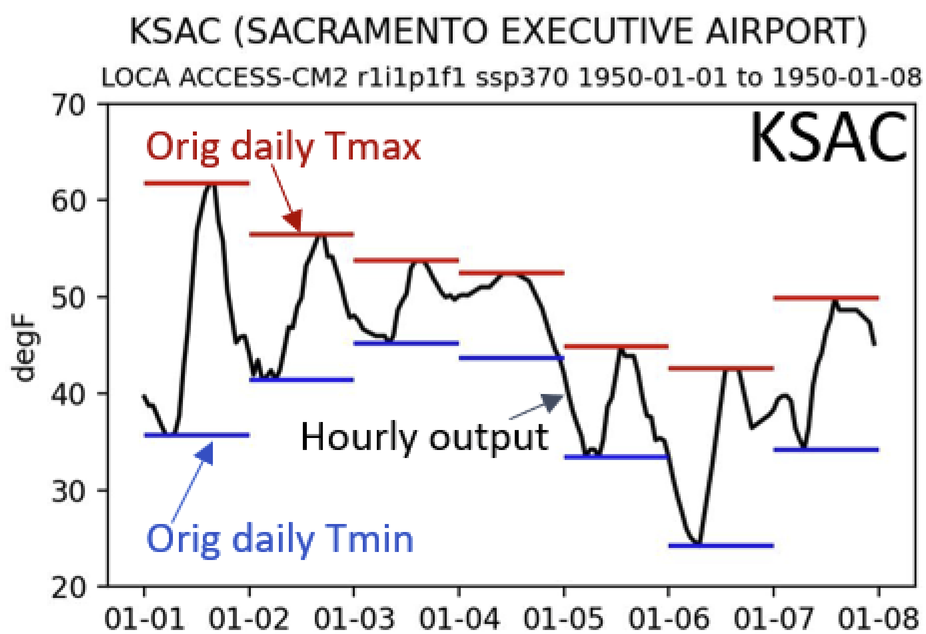

Figure 11. Example time series plot of the generated hourly LOCA2 output at KSAC (Sacramento, CA). Red/blue lines are the input daily LOCA2 Tmax and Tmin values, respectively. The black line is the hourly LOCA2 temperature projection.

4.2) Verification of hourly output

Once the hourly projections were completed, several verification plots were produced to check the quality of the output. The first step was to plot a random week of hourly values in the training data, HadISD, compared to the hourly LOCA2 projections for each station. The objective here is to visually inspect the output and make sure that the constructed values look "real" – in other words, if the plots were not labeled, one should not be able to identify which plots are observed and which ones are modeled.

An example for 3 selected stations, one each in Northern, Central, and Southern California, is shown in Figure 12. A random week starting during the month of January (to account for seasonal differences) was selected from each of the training data, LOCA2 historical period, and LOCA2 ssp370 future period.

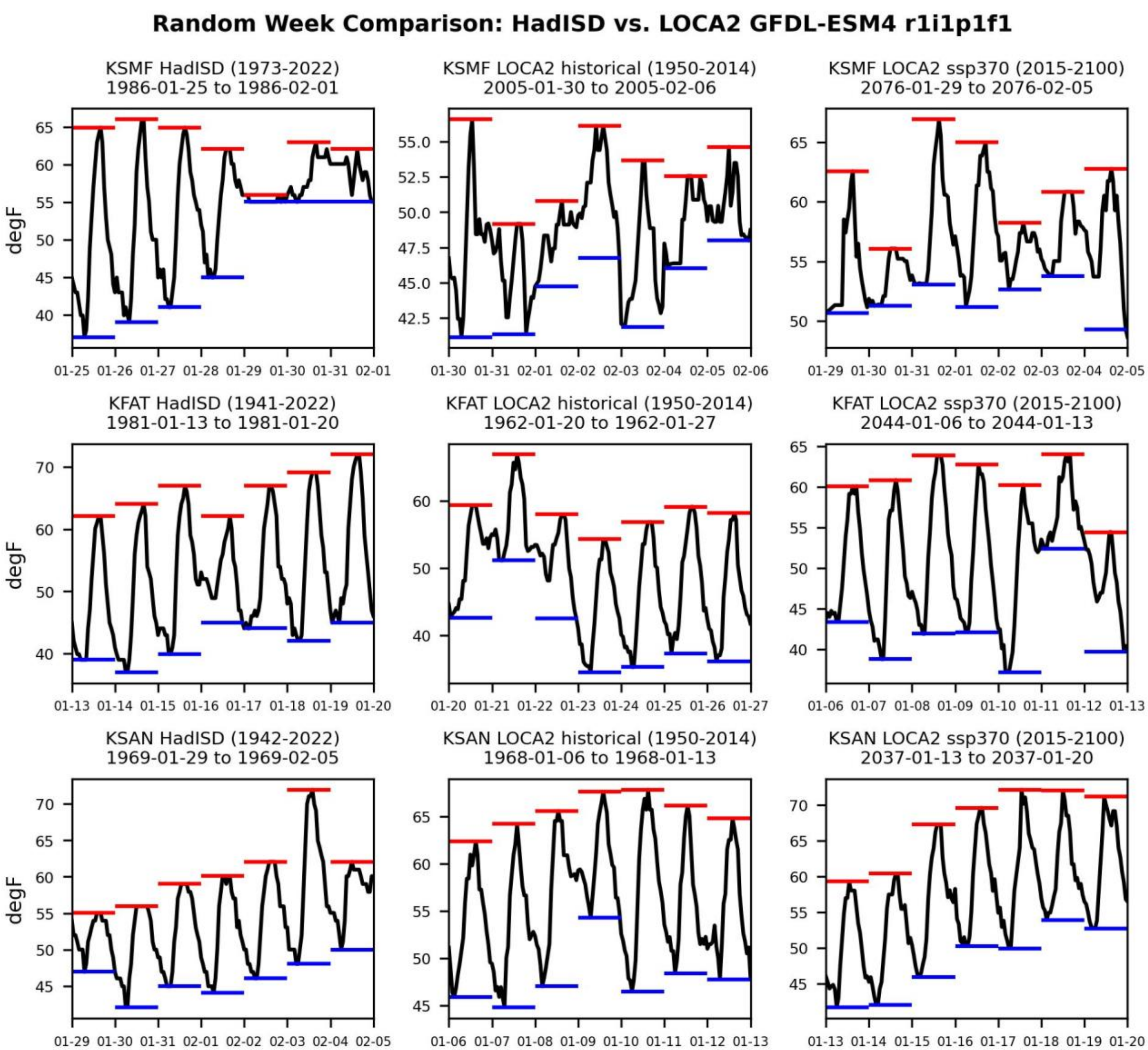

Figure 12. Example of randomly selected weeks starting in the month of January at 3 selected stations: KSMF (Sacramento, CA), KFAT (Fresno, CA), and KSAN (San Diego, CA). Left to right: HadISD data, LOCA2 historical period, and LOCA2 ssp370 future period. Hourly values in black, daily Tmax in red, daily Tmin in blue.

Next, several outlier values showing the largest changes in Tmax were compared between HadISD and the LOCA2 historical and future periods. Like the previous check, the objective here is to verify that the constructed hourly projections are comparable to the observations during periods when the temperature changes significantly. Again, Tmax was considered here because of its importance to energy demand applications.

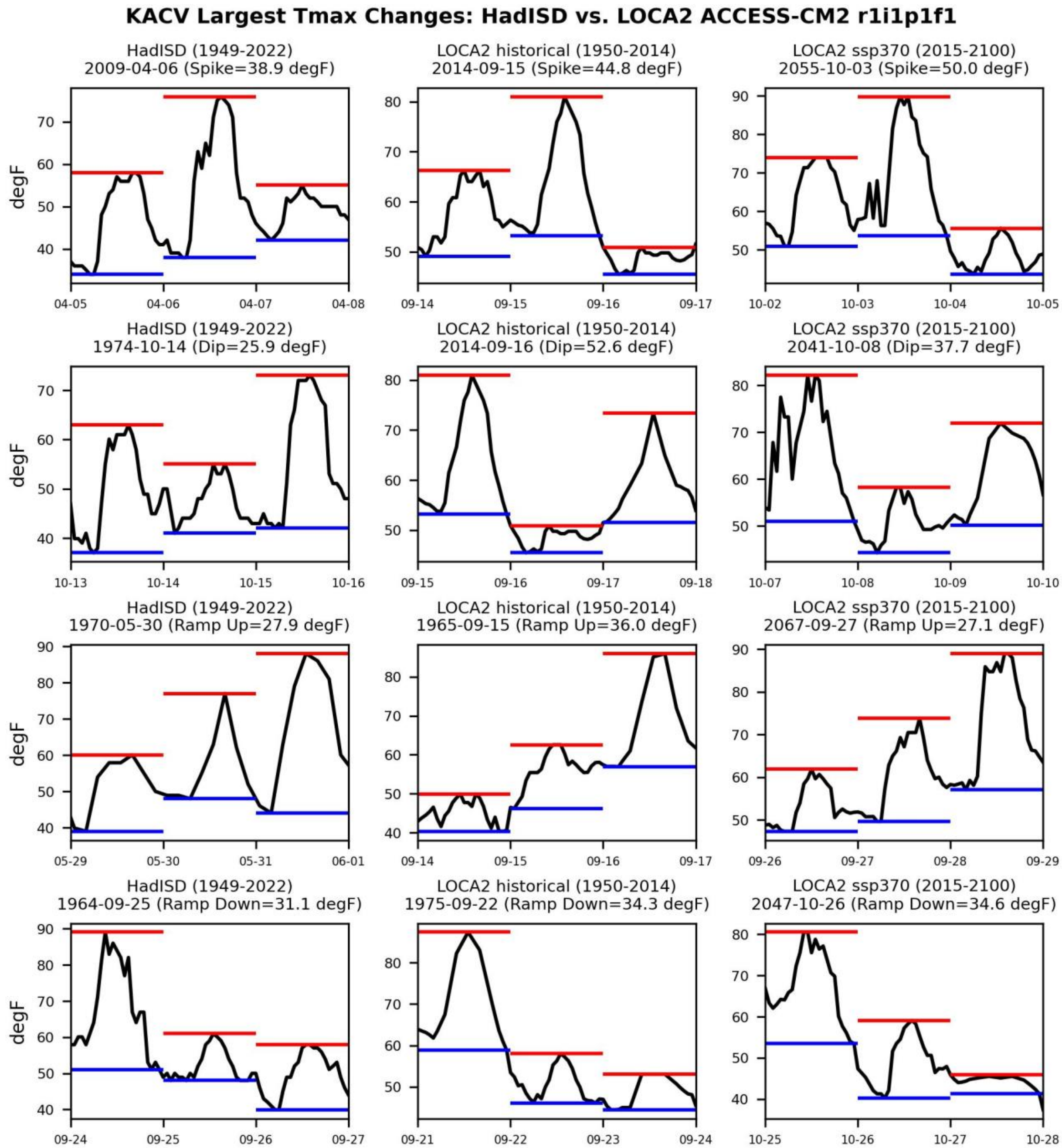

Figure 13. Largest changes in Tmax at KACV (Arcata, CA). Top to bottom: largest single day spikes in Tmax, largest single day dips, largest ramp ups, and largest ramp downs. Left to right: HadISD, LOCA2 historical period, and LOCA2 ssp370 future period.

For a given 3-day sequence, let the Tmax values on those days be represented as Tmax1, Tmax2, and Tmax3, respectively. Four quantities are then defined according to the equations below: single day spikes in Tmax, single day dips, ramp ups, and ramp downs.

$$Tmax\ single\ day\ spike = (Tmax2 - Tmax1) + (Tmax2 - Tmax3)$$

$$Tmax\ single\ day\ dip = (Tmax1 - Tmax2) + (Tmax3 - Tmax2)$$

$$Tmax\ ramp\ up = (Tmax2 - Tmax1) + (Tmax3 - Tmax2)$$

$$Tmax\ ramp\ down = (Tmax1 - Tmax2) + (Tmax2 - Tmax3)$$

The largest single day spike in Tmax is thus the largest increase in Tmax relative to the preceding and following days, while the largest single day dip is the largest decrease. The largest Tmax ramp up is the 3-day period with the greatest Tmax increase, while the largest Tmin ramp down is the period with the greatest decrease. Figure 13 shows an example of these four quantities for one LOCA2 simulation at KACV (Arcata, CA). Once again, the results look comparable such that it would be difficult to distinguish between the observations and modeled results simply by looking.

Next, days with the largest and smallest DTR were compared, similarly, to ensure that the observations and modeled results are comparable. An example at KSAN (San Diego, CA) for one LOCA2 simulation is shown in Figure 14. Note that the minimum DTR threshold applied to the projections is based on GHCNd, not HadISD. Thus, it is possible for the smallest DTR observed in the projections to be less than the smallest DTR observed in HadISD.

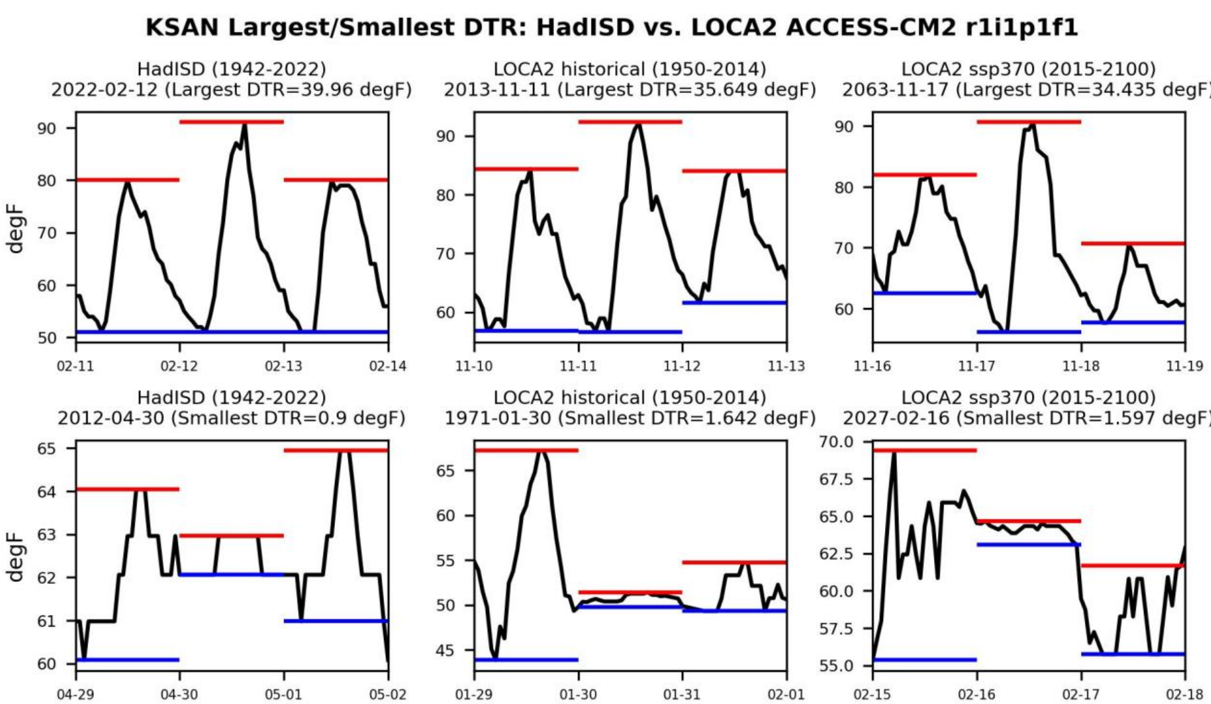

Figure 14. Largest/smallest DTR days at KSAN (San Diego, CA). Top: days with largest DTR, bottom: days with smallest DTR. Left to right: HadISD, LOCA2 historical period, and LOCA2 ssp370 future period.

After examining the hourly timeseries of the final projections, the overall hourly temperature progression was considered by plotting a few distributions: the distribution of 1-hour temperature changes and the distribution of the warmest hour of the day. These distributions for the LOCA2 projections are expected to be similar to those from HadISD, since the hourly values are constructed using the training data. Comparing the distributions simply ensures that the hourly temperature progressions in the projections were not significantly altered somehow during the process.

The distributions are plotted for the same 3 stations as in earlier examples (KSMF, KFAT, and KSAN). The distribution of hourly changes is shown in Figure 15 while the distribution of the warmest hour of the day is shown in Figure 16: in both cases the LOCA2 historical results are very comparable to the training data.

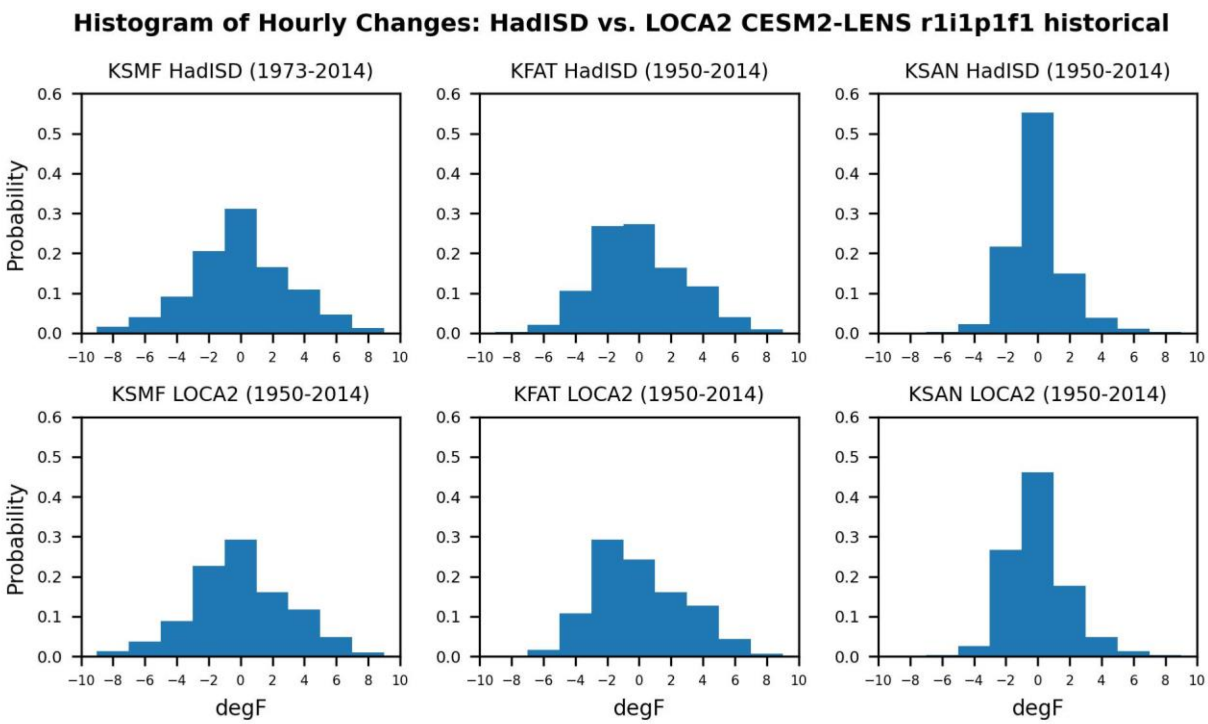

Figure 15. Histograms of hourly temperature changes for one LOCA2 historical simulation at 3 selected stations: KSMF (Sacramento, CA), KFAT (Fresno, CA), and KSAN (San Diego, CA). Top: HadISD, bottom: LOCA2 historical period.

One final consideration is the number of times each analog day from the training data is used in the hourly output. For example, it is desirable for variability of the output to ensure that the hourly projections are not being constructed from the same handful of repeated analog days.

A few example histograms are shown in Figure 17, for 3 stations with varying station records (number of valid analog days): KLAS (Las Vegas, NV) has 71.1 years of valid analog days, KSBA (Santa Barbara, CA) has 40.8 years, and KMCE (Merced, CA) has 15.6 years. As expected, the number of times a given analog day is selected tends to increase when there are fewer possible analog days to choose from. While this did not measurably impact the variability of the output here, a potential consideration for a future version of these projections is to limit the number of times that each day in the training data can be chosen as an analog day.

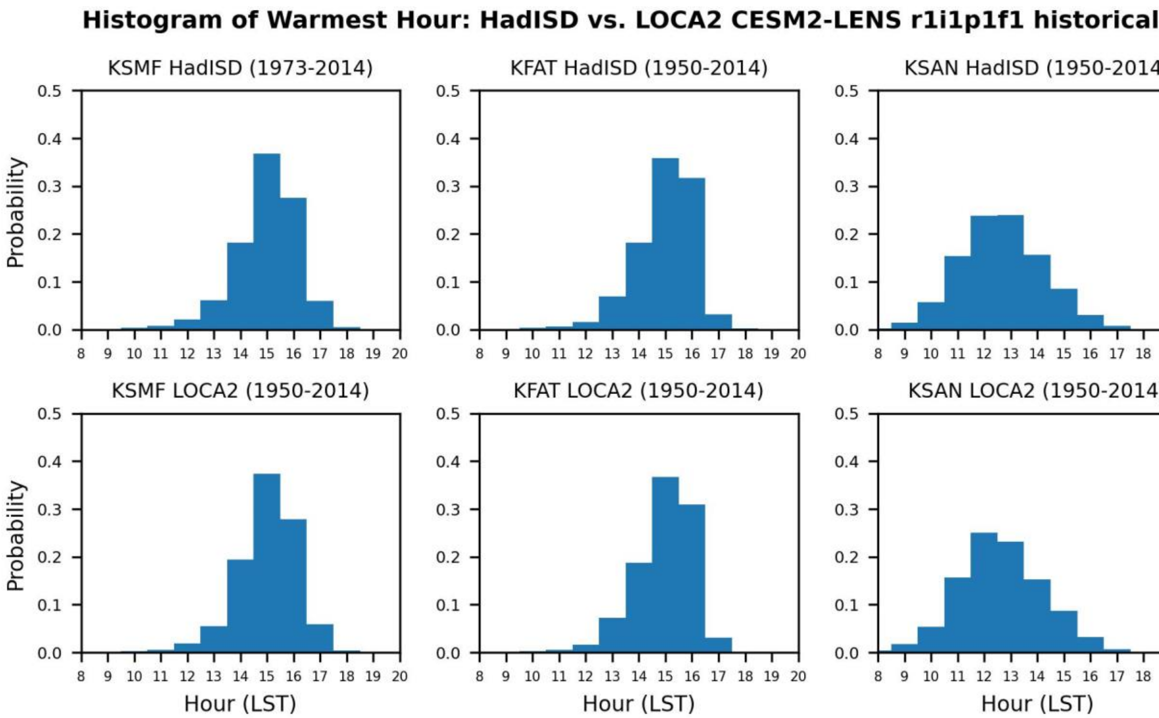

Figure 16. Histograms of the warmest hour of the day in Local Standard Time (LST) for one LOCA2 historical simulation at 3 selected stations: KSMF (Sacramento, CA), KFAT (Fresno, CA), and KSAN (San Diego, CA). Top: HadISD, bottom: LOCA2 historical period.

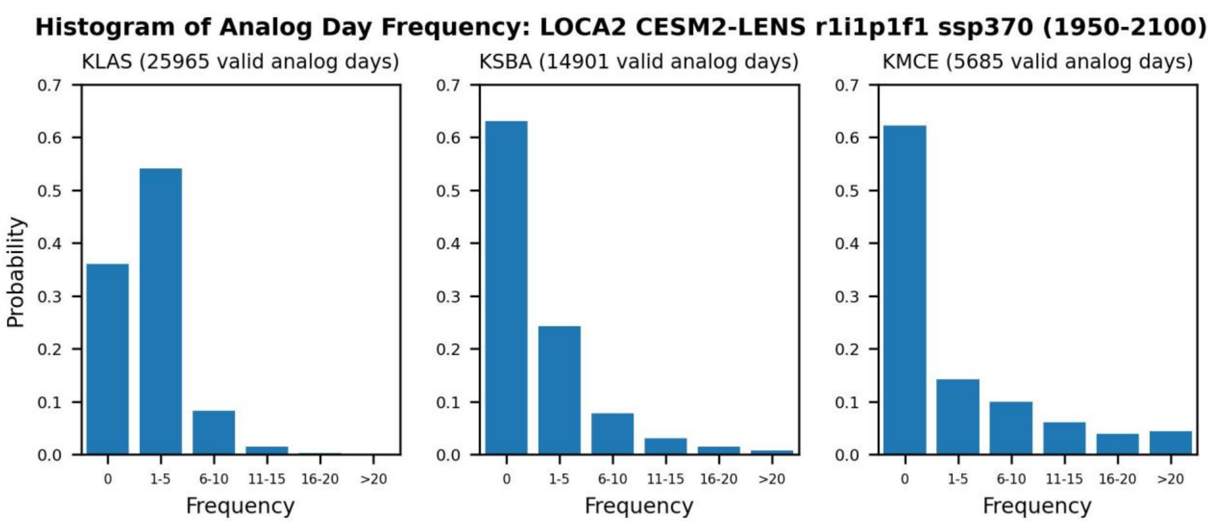

Figure 17. Histograms of the number of times each analog day was selected for one LOCA2 simulation (1950-2100) at 3 selected stations with varying numbers of valid analog days: KLAS (Las Vegas, NV), KSBA (Santa Barbara, CA), and KMCE (Merced, CA).

Acknowledgments

This project would not have been possible without the support of my mentors, David Pierce and Dan Cayan, as well as the California Energy Commission.

References

Dunn, R. J. H., et al. (2012), HadISD: A Quality Controlled global synoptic report database for selected variables at long-term stations from 1973-2011, Climate of the Past, 8, 1649-1679.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily Database. Journal of Atmospheric and Oceanic Technology, 29, 897-910, doi.10.1175/JTECH-D-11-00103.1.

Pierce and Cayan 2019: Future projections of hourly surface temperatures in California. Available online at https://cirrus.ucsd.edu/~pierce/tmp/Hourly_data_interpolation_2019-05-23.pdf

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., D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, 2015a: Improved bias correction techniques for hydrological simulations of climate change. J. Hydrometeorology, v. 16, p. 2421-2442. DOI: http://dx.doi.org/10.1175/JHM-D-14-0236.1

Pierce, D. W. and D. R. Cayan, 2015b: Downscaling humidity with Localized Constructed Analogs (LOCA) over the conterminous United States. Climate Dynamics, DOI 10.1007/s00382-015-2845-1.

Pierce, D. W., D. R. Cayan, D. R. Feldman, and M. D. Risser, 2023a: Future Increases in North American Extreme Precipitation in CMIP6 downscaled with LOCA. J. Hydrometeor., https://doi.org/10.1175/JHM-D-22-0194.1.