# Memorandum on LOCA2 – Hourly Surface Temperature Projections for Stations in California

**Interim Deliverable for EPC-20-006, Prepared by:**
Peter Yao, David W. Pierce, Daniel R. Cayan, Scripps Institute of Oceanography, La Jolla, California
January, 2024

- This research is funded by the California Energy Commission (CEC) through its Electric Program Investment Charge (EPIC) Program, which invests in scientific and technological research to accelerate the transformation of the electricity sector to meet the state's energy and climate goals.
- The applied research grant, EPC-20-006, will integrate the latest downscaling approaches applied to the recently produced global climate models (GCMs) with an engagement process to develop a robust, usable, set of climate projections applicable for California.
- This memo is being shared to support transparent and timely consideration of interim deliverables that are relevant for energy stakeholders and all those interested in California's next generation of climate projections.

**Methods and Prior Relevant Work**

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 (Figure 1).

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.

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 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.

Constructing the projections begins with identifying the training dataset at each station. 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.

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 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.

Finally, the analog matching process transforms the station adjusted daily LOCA2 Tmax and Tmin data into hourly temperatures (Figure 2). 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.

## QA/QC and Uncertainty

Because this set of projections is based on the LOCA2 hybrid downscaled product, it inherits any uncertainties from the original product, such as uncertainties in the original GCM-WRF data or those arising from the bias correction process (Pierce et al. 2023b).

One of the main uncertainties associated with the creation of this dataset is the lack of a perfect observational station record. 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. 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 HadISD QA/QC step was performed by Eagle Rock Analytics. The GHCNd data used in the station adjustment process 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. However, any remaining uncertainties resulting from the quality of the training data will carry over to the hourly output.

It is also worth noting that during the station adjustment process, a summer wet day correction was applied to account for the lack of wet days during summer months (JJA) in the historical record at most stations. 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 correction was necessary to ensure that a climatology based on a very small sample size was not used in the adjustment process, which can be subject to large fluctuations or uncertainties.

A number of checks were performed on the hourly projections to assess their quality. For example, randomly selected weeks from the training data and generated hourly values were compared visually to check that there aren't systematic differences in the hourly time series between the two datasets. Similarly, hourly values during the top ten largest increases or decreases in Tmax between consecutive days (spikes/dips) were plotted, as well as the top ten largest Tmax increases/decreases over a 3-day period (ramp ups/ramp downs) and the largest/smallest DTR. Distributions of hourly temperature changes, DTR, and the warmest hour of the day were also compared between the training data and output. The temperature trend from the historical to future periods was also plotted to verify that any trends found in the LOCA2 projections were preserved following the station adjustment process.

## Guidance or Caveats on Best Practices for Use of Data Products

Several best practices specific to this dataset are detailed below; however, refer to the guidance outlined in the "LOCA2 Hybrid Downscaling" (Pierce et al. 2023b) memo for a fuller set of considerations on how to use the LOCA2 projections, which apply here as well.

The major advantage of this product is its hourly resolution compared to the daily gridded LOCA2 projections. However, hourly temperature projections are also available from the UCLA WRF dynamically downscaled runs (Rahimi-Esfarjani 2022a and 2022b). There are many factors to consider when deciding which product is best suited for a specific use case. For example, the WRF data, which is a gridded product, has a wider spatial coverage and includes a number of other hourly variables in addition to temperature. The LOCA2 station projections include a wider range of GCMs, ensemble members, and SSPs (129 total LOCA2 simulations for each station vs. 9 WRF runs), and all of the LOCA2 projections are bias corrected, unlike the WRF data. The LOCA2 station projections also underwent a station adjustment process to account for systematic differences between the gridded projections and the observed station record, which makes them better suited for station-based applications, compared to simply extracting the nearest grid point to a given location from a gridded dataset. Due to the complexity of these considerations, there is no one size fits all answer to which data should be used.

Another consideration to keep in mind is that temperatures can vary widely between different spatial locations even when nearby: for example, factors such as elevation, slope/aspect, and vegetation can all influence local temperatures. Therefore, there may be applications for separate projections even for nearby stations (e.g. KSAC vs. KSMF, which are both located within the Sacramento metro area).

One caveat that bears repeating is that GCMs do not forecast actual temperatures on a particular day in the future (e.g. July 1st, 2070) – rather, they strive to simulate the statistics of temperature on future days. In this case, the actual projected hourly temperature values on a given day should not be treated as such, rather their statistical properties can be evaluated.

A final consideration is that the analog matching process cannot project a future diurnal cycle shape that has not been observed before. More realistically, changes in the shape of the diurnal cycle that are not reflected in changes to daily Tmin/Tmax would not be captured.

## 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.

Pierce, D. W., D. R. Cayan, S. Rahimi, and J. Kalansky, 2023b: LOCA2 Hybrid Downscaling**.** Data memo for the California Energy Commission, 4 pp.

Rahimi-Esfarjani, S.: Memo on the Evaluation of Downscaled GCM Using WRF, Report to California Energy Commission. https://cal-adapt.org/files/01_Memo_Evaluation_of_Downscaled_GCMs_Using_WRF_CEC_final.pdf, 2022a.

Rahimi-Esfarjani, S.: Memo on the Development and Availability of Dynamically Downscaled Projections Using WRF, Report to California Energy Commission. https://www.energy.ca.gov/sites/default/files/2022-09/20220907_CDAWG_MemoDynamicalDownscaling_EPC-20-006_May2022-ADA.pdf, 2022b.

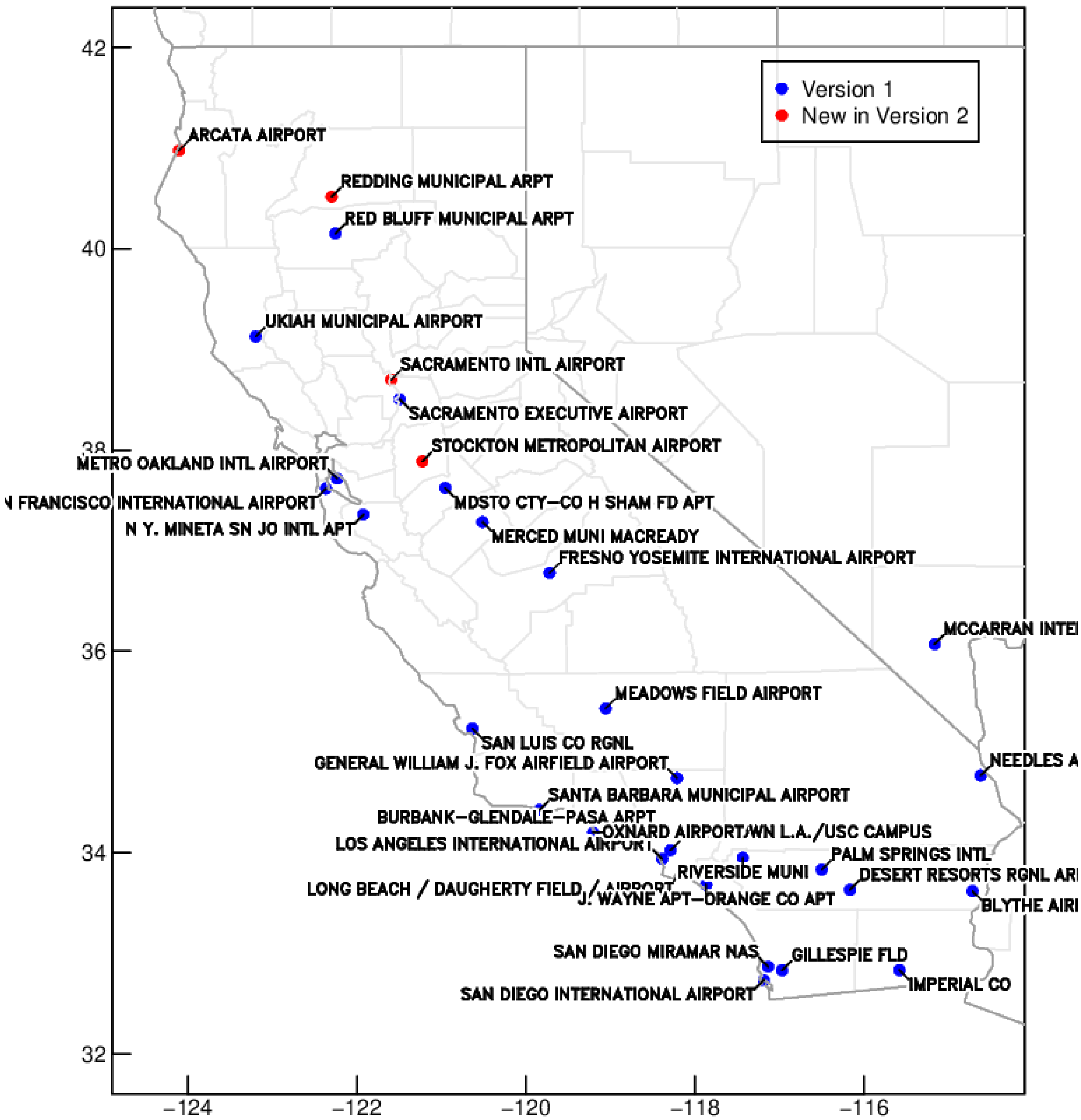


Figure 1. Map of the 33 stations included in the projections dataset. Stations in blue were included in the original temperature projections dataset (Pierce and Cayan, 2019). Stations in red are new additions to this version.

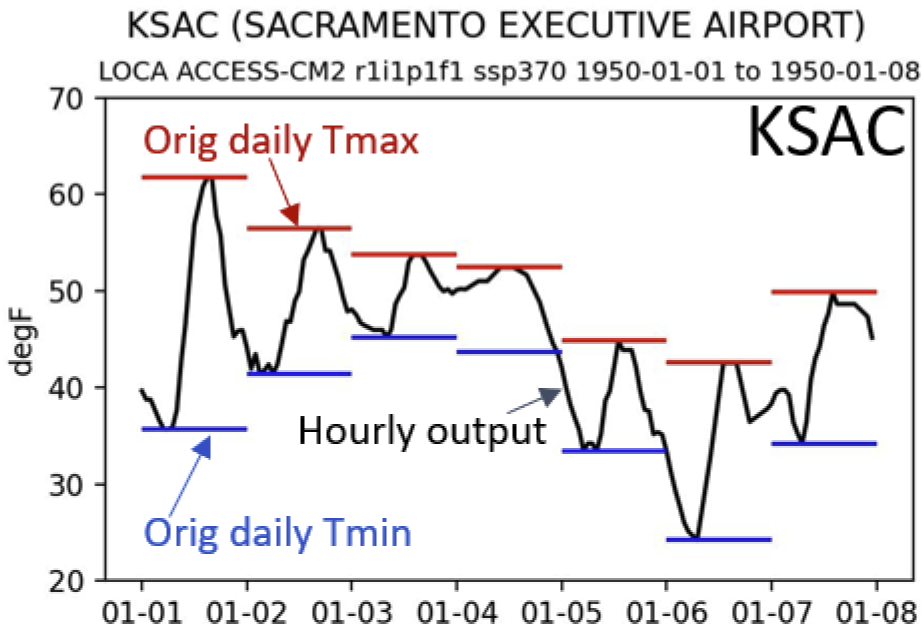


Figure 2. Example time series plot of the generated hourly LOCA2 output at KSAC. Red/blue lines are the input daily LOCA2 Tmax and Tmin values, respectively. The black line is the hourly LOCA2 temperature projection.