CALIFORNIA ENERGY COMMISSION

Energy Research and Development Division

FINAL PROJECT REPORT

Weather and Climate Informatics for the Electricity Sector

Subdaily Observations and the Predictability of Extreme Heat Events

Gavin Newsom, Governor
June 2020 | CEC-500-2020-039

PREPARED BY:

Primary Author:
Owen Doherty, Ph.D.

Eagle Rock Analytics
Sacramento, CA 95820
(916) 936-0199
Eaglerockanalytics.com

Contract Number: EPC-15-036

PREPARED FOR:
California Energy Commission

Susan Fischer Wilhelm, Ph.D.
Project Manager

Jonah Steinbuck, Ph.D.
Office Manager
ENERGY GENERATION RESEARCH OFFICEE

Laurie ten Hope
Deputy Director
ENERGY RESEARCH AND DEVELOPMENT DIVISION

Drew Bohan
Executive Director

ACKNOWLEDGEMENTS

The authors thank and acknowledge the topical, administrative, and scientific guidance provided to this work by the Commission Agreement Manager Dr. Susan Wilhelm. The authors are grateful for the early leadership of Guido Franco, initial Commission Agreement Manager on this project, and for guidance and feedback from Dr. David Stoms and Dr. Martine Schmidt-Poolman of the Energy Commission. This work benefited from direct feedback and commentary from the Sacramento Municipal Utility District, facilitated by Kathleen Ave. The authors are grateful for comments and guidance from the technical advisory committee members; in particular, comments from the Demand Forecast Office were critical to the success of this work. The work benefited from the technical contributions of Chris Torsiello, who built from scratch and at low cost a high-performance workstation to support the computational efforts undertaken in this work. The authors are grateful for the assistance of Ms. Eva Chavez in the preparation and formatting of this manuscript. The authors acknowledge the editorial assistance of Dr. Anne Cooper Doherty.

PREFACE

The California Energy Commission's (CEC) Energy Research and Development Division supports energy research and development programs to spur innovation in energy efficiency, renewable energy and advanced clean generation, energy-related environmental protection, energy transmission and distribution and transportation.

In 2012, the Electric Program Investment Charge (EPIC) was established by the California Public Utilities Commission to fund public investments in research to create and advance new energy solutions, foster regional innovation and bring ideas from the lab to the marketplace. The CEC and the state's three largest investor-owned utilities — Pacific Gas and Electric Company, San Diego Gas & Electric Company and Southern California Edison Company — were selected to administer the EPIC funds and advance novel technologies, tools, and strategies that provide benefits to their electric ratepayers.

The CEC is committed to ensuring public participation in its research and development programs that promote greater reliability, lower costs, and increase safety for the California electric ratepayer and include:

- Providing societal benefits.
- Reducing greenhouse gas emission in the electricity sector at the lowest possible cost.
- Supporting California's loading order to meet energy needs first with energy efficiency and demand response, next with renewable energy (distributed generation and utility scale), and finally with clean, conventional electricity supply.
- Supporting low-emission vehicles and transportation.
- Providing economic development.
- Using ratepayer funds efficiently.

Weather and Climate Informatics for the Electricity Sector is the final report for the Probabilistic Seasonal and Decadal Forecasts Using Linear Inverse Modeling project (Contract Number EPC-15-036) conducted by Eagle Rock Analytics. The information from this project contributes to the Energy Research and Development Division's EPIC Program.

For more information about the Energy Research and Development Division, please visit the CEC's research website (www.energy.ca.gov/research/) or contact the CEC at 916-327-1551.

ABSTRACT

Weather conditions are an important driver of demand for electricity. Very warm days during the summer season are associated with peak electricity consumption. As California has moved toward a zero-carbon, high-renewables electricity system, the need for improved weather data and information has increased. Hourly weather data are highly variable ("noisy"), with frequent observational errors and instrumentation failures, making data use difficult. In this project, the research team designed and implemented a series of tests to produce a stable record of hourly weather data for use by the energy sector. After careful data quality review, the team produced a curated repository of hourly weather observations at 39 locations across California for 1973–2019. This report discusses the utility of this product and provides recommendations for how best to use the data and supporting documentation. California's warming trend is asymmetrical — stronger in the late afternoon and weaker shortly after sunrise. The report discusses regional and seasonal differences from the statewide trend, with a focus on implications for California's electricity supply and demand. Data products used in this work form the basis for the development of a statewide, multilevel subdaily repository and represent a clear pathway forward to providing energy sector stakeholders with regular ultra-high-resolution data products that are critically needed to help California meet its renewable energy and climate goals. Last, the report quantifies the value of Pacific Ocean surface conditions in informing predictions of temperature in California and outlines an approach for making such predictions operational. The data produced and analyses performed in this project provide investor-owned utilities as well as public utilities and state agencies with insights into the effects of subdaily weather on the electrical system.

Keywords: weather, hourly, subdaily, trends, temperature, seasonal predictions

Please use the following citation for this report:

Doherty, Owen and Amato Evan. 2020. *Weather and Climate Informatics for the Electricity Sector*. California Energy Commission. Publication Number: CEC-500-2020-039.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

EXECUTIVE SUMMARY

Introduction

Weather conditions are an important driver of demand for electricity, with very warm days during the summer season associated with peak electricity consumption. As California moves toward a diverse energy future focused on renewables, the need for improved weather data and information has increased. Renewables largely rely critically on weather — solar generation varies with cloud cover, and wind power generation varies with wind speed. Additionally, hydroelectric resources — which have historically provided a rapid-response, dispatchable, low-carbon resource for meeting peak demand — vary with changes in the type, frequency, and intensity of precipitation. Thus, to make accurate predictions of energy consumption, the California Energy Commission (CEC) needs accurate weather data.

The traditional approach to electricity sector planning relies on daily weather data, often considering minimum and maximum temperatures as key indicators. Such daily snapshots convey the gist of the weather conditions that occurred but are not sufficient in scope to resolve changes in energy use as a function of the time of day.

Despite striving to position itself as a leader in climate science, California lags behind most other states in collecting subdaily weather observations in a central location. This data void hampers the ability of the energy sector to plan for and respond to ongoing climate change and limits the CEC's ability to meet the needs of investor-owned utilities and others for high temporal resolution climate information. In this work, the authors outline a plan for the CEC to lead efforts toward a statewide mesonet (a network of automated weather and environmental monitoring stations), which can enhance the stability of the electrical system and also act as a critical source of information for future climate modeling efforts.

Project Purpose

The aim of this work is to provide the CEC with a curated, stable, long-term weather record with subdaily observations appropriate for typical electricity sector analysis. This record has been carefully examined to ensure it is of sufficient length and quality for typical studies and analyses. Stakeholders recognize the need for more frequent and finer resolution weather observations. This work provides a pathway forward for the CEC to transition from daily to hourly, and from regional to local scales, in support of California's upcoming Fifth Climate Change Assessment.

The data and analyses produced in this project are designed to support the activities of the CEC's Demand Forecast Office, science and engineering teams of publicly and investor-owned utilities, and other energy sector stakeholders through dissemination via Cal-Adapt, a web-based climate adaptation planning tool. This project provides California's electricity sector with the needed data to make the step from daily to hourly analyses. Beyond curating and organizing a stable hourly weather record for the energy sector, the authors discuss how this work supports ongoing CEC efforts and energy sector needs.

- This project seeks to improve understanding of how temperature is changing in California by analyzing the change on hourly rather than daily timescales. This information is critically important as utilities begin to shift demand toward off-peak hours and would not be possible without a stable long-term record of hourly weather observations across California.

- Temperature data are needed in many forms by the energy sector. In this work, the authors seek to provide a reliable standardized record of temperature, including typical daily formulations for modeling work that require long-term average conditions by the hour, and raw and transformed records that highlight various aspects of temperature, including how temperature deviates from daily or hourly long-term climatological average.
- As renewables become more prominent in California's electricity market, highly accurate, high temporal resolution weather data is required to ensure a stable power supply during extreme weather conditions. This work addresses seasonal predictability by assessing the capacity of the Pacific Ocean to explain and predict temperatures in California and provides guidance on a pathway forward to making accurate predictions on seasonal (and shorter) timescales.

Weather data are notoriously difficult to process for quality control. Weather stations can move, instrumentation used to make observations can change and bias the record, stations can abruptly shut down, and many stations operate only during daytime hours. Historical weather datasets are full of errors — from incorrect recording of observations into logs, to errors in the transmission of observations to central repositories, to errors in digitization of weather archives. A major effort of this work is to identify through systematic testing weather observations that are of sufficient quality to serve the sensitive needs of the energy system and, moreover, to ensure that this record is representative of the complex meteorology in California.

Stakeholders, including investor-owned utilities, have requested that the CEC produce records of temperature at a resolution of approximately two kilometers, on hourly timescales, extending back at least 30 years. Further, researchers responsible for downscaling global climate model predictions to smaller, localized scales require a longer-term, high-resolution temperature product to train their models. No such record exists presently to meet these needs. Support for this effort by the CEC was necessary because no federal or private funding source is available to provide the resources needed to produce these data.

Project Approach

This project was managed by Dr. Owen Doherty of Eagle Rock Analytics, with Dr. Amato Evan of the University of California, San Diego providing scientific and technical leadership. Work tasks associated with this project were primarily computer-driven data and statistical analysis of atmospheric data; this work was performed on high-performance computing clusters and workstations in San Diego and Sacramento, respectively. Approaches used varied from statistical modeling via linear algebra to calculations of ratios of events, albeit repeated many billion times. The specific approaches and the barriers they overcame are briefly described as follows:

- The authors developed and applied a rigorous quality control protocol to ensure a stable and reliable product for every weather station within California.

- To ensure that the hourly temperature database is applicable to a wide range of purposes and accessible in formats that are commonly used, raw and transformed data are stored as NetCDFcdfs and R database files.
- The temperature record was smoothed through a low-pass filter to produce a clean "typical day" profile. Each daily profile contains the long-term mean value for each hour of the day, as a raw value and as an anomaly relative to the daily mean temperature.
- The authors applied a linear trend analysis for each hour in the day, using monthly mean temperature for each station and year in the record. To account for the potential influence of instrumentation changes, the authors repeated the analysis for the manual and automated weather observer periods.
- The predictability of air temperature in California by the Pacific Ocean was assessed through linear inverse modeling of ocean temperature and singular-value decomposition analysis of air temperature/ocean temperature–coupled data. This is a computationally light statistical approach that negates the need for expensive supercomputer modeling.

This approach was guided by, and greatly improved through feedback from, the project's technical advisory committee, comprised of representatives from investor-owned and publicly owned utilities, the CEC Demand Forecast Office, and the California Independent System Operator to ensure broad stakeholder input and support. Two meetings of the technical advisory committee were held to refine and improve upon methods and products. The technical advisory committee explicitly advised that subseasonal (including subdaily) timescales were important, necessitating a deep and comprehensive examination of the sources and quality of subdaily weather information. Further, input from the technical advisory committee was critical in determining the need to quantify the rate of change in air temperature in California on an hourly basis due to anthropogenic climate change.

Project Results

Stakeholders, including investor-owned utilities, public utilities, and CEC staff, clearly and directly indicated that subseasonal, reliable, and stable weather information is critically needed to achieve climate-related goals. Some highlights of this work related to the four main thrusts of this project include:

1. The generation of a stable, long-term hourly weather database.
2. The creation of typical day formulations for temperature as a resource for energy-sector modeling studies that use climatological mean conditions.
3. The careful assessment of the rate of change of temperature in response to anthropogenic climate change on an hourly basis.
4. The assessment of the potential for generating seasonal predictions of temperature across California.

While all these results are of interest to at least some stakeholders, it is the opinion of the authors that the hourly trend analysis — specifically the results suggesting late afternoon as the period of the day warming most rapidly — has the greatest effect on the energy sector overall.

Knowledge Transfer

This project has yielded two types of products: data and actionable, utility-relevant knowledge that will be disseminated to key stakeholders and the public.

Data products will be made publicly available on Cal-Adapt. Data have been output for transfer in stakeholder-requested or endorsed formats (that is, R databases and NetCDF). Data are structured and annotated (via metadata) to meet the best practices of Cal-Adapt and more broadly, industry standards. Included in this report are recommendations and suggestions of how best to automate additional observations to keep the records current, as well as guidance on how best to host, visualize, and transmit the data.

Knowledge will be transferred to stakeholders and investor-owned utilities at a forthcoming Cal-Adapt User Needs Assessment, tentatively focused on sub-seasonal data. The results of this work will be presented, along with next steps and recommended courses of action. Additional outreach has occurred within the CEC, ensuring that the knowledge generated will be shared with multiple divisions.

Recommendations and Future Directions

This report discusses the next steps to progress to a more granular, data-intensive approach to weather data that meets energy-related needs (for example, demand forecasting), as articulated by CEC Commissioner Andrew McAllister at the March 4, 2019, Integrated Energy Policy Report Workshop on Data Inputs and Assumptions for Modeling and Forecasting Activities. This report lists concrete steps, along with background information and context to allow the CEC to improve the quality of and access to hourly weather data, leverage hourly data to create additional products that utilities need, and make seasonal predictions operational. The future directions include:

1. Expanding the current hourly weather database (greater than 46 years of hourly observations at 39 locations in California) to cover regions of California presently underrepresented.
2. Improving the scope (higher spatial resolution on subdaily timescales) and the quality of the temperature record, as requested by utilities and energy sector stakeholders to meet climate-related goals.
3. Leveraging California’s great opportunity to improve the understanding of temperature variability across the state by merging observations that are presently disparate (across various state and local agencies, as well as utilities) into a functional, uniform mesonet. The CEC can continue its historical role of climate leadership at the state level through:
    - Centrally collecting, quality controlling, and assimilating weather observations across the state into a uniform data product, and disseminating this information in near-real time on Cal-Adapt, providing hyperlocal weather information to utilities to improve reliability in the face of extreme weather events.
    - Using information from this collection to build a high-resolution gridded product that would meet the spatial and temporal needs articulated by energy sector stakeholders.

    - Merging these disparate networks into a single network, which would be of great value to the electricity sector, via the protection of life and property from extreme weather events and improved system reliability through better weather forecasting.
4. Developing a seasonal prediction scheme for air temperature in California, using a hybrid approach, including information from linear inverse models and seasonal climate modeling efforts.

Benefits to California

California's electricity system already relies on a substantial portion of renewable generation and will continue to incorporate more electricity from renewable sources of energy, such as solar and wind. Concurrently, California's peak demand for electricity during extreme heat events will increase as the state continues to warm. Historically, investor-owned utilities and CEC Demand Forecast Office personnel have used daily climate information to assess and plan for energy use in response to weather. With the growth in renewable energy sources, and the rising availability of per-user subdaily consumption data, subdaily weather information is increasingly necessary. This work provides a stable, quality-controlled record curated for California's energy sector. Providing this information to energy system stakeholders through Cal-Adapt will enable a more stable energy system for California ratepayers, by allowing utilities to better understand how diurnal weather variability and electricity use are related. Improved weather data has the potential to allow for reduced electricity costs if utilities can leverage the information provided to improve supply and electricity acquisition decisions. Further, benefits to ratepayers are likely to emerge as subdaily data can significantly better inform infrastructure developments compared to daily data. This work will provide a centralized repository of subdaily weather data, which will reduce utilities' expenses in acquiring expert guidance to produce their own records, and ultimately lower ratepayers' electricity costs.

California is warming asymmetrically, with different warming rates across the day as a function of location and month in year. When during the day and how temperature is changing at a location are becoming more important as utilities seek to implement peak demand shifting. This information will help utilities anticipate and plan for near-term future changes in demand, leading to a more stable electrical system at a lower cost, and is likely to be of greater benefit to publicly-owned and smaller utilities, whose ability to add staff with expertise in climate and weather is limited.

Improvements to energy use projections, through improved weather and climate forecasts, will increase electricity stability and reduce rates by allowing utility companies to make informed decisions about infrastructure and market purchases. Accurate predictions of temperatures will ensure sufficient electricity is available to meet peak demands. As California moves toward more diverse sources of energy in the future, including higher shares of renewables, accurate weather forecasts are going to become increasingly necessary for the CEC and energy suppliers to accurately predict energy generation and consumption. Through more accurate seasonal and decadal predictions, utility companies can ensure adequate zero-carbon electricity supplies are available to ratepayers.

CHAPTER 1:
Introduction

Weather and Climate Information for California's Energy Sector

Weather conditions are an important driver of demand for electricity, with very warm days during the summer season associated with peak electricity consumption. As California has moved toward a diverse energy future, with a focus on renewables, the need for improved weather data and information has increased. Renewables largely rely critically on weather — solar generation varies with cloud cover and wind power generation varies with wind speed. Additionally, hydroelectric resources — which have historically provided a rapid-response, dispatchable, low-carbon resource for meeting peak demand — vary with changes in the type, frequency, and intensity of precipitation. Thus, to make accurate predictions of energy consumption, the CEC needs accurate weather data.

This report discusses potential sources of weather data, as well as methods for and results from a careful assessment of suitability of the data and changes in the daily structure of temperature at the monitoring sites, and provides robust recommendations for how these data can be disseminated via Cal-Adapt and used as the basis for products that address the needs of California's electricity sector. Beyond this, the report presents an approach for producing seasonal predictions of temperature for California, the limitations of this approach, how such projections are likely to be used, and recommendations for making seasonal predictions available to the energy sector and the public.

Emerging Needs for Subdaily Weather Data

Historic Uses of Weather Data by the Energy Sector

The traditional approach in electricity sector planning relies on daily weather data, often considering daily maximum and minimum temperatures as key indicators. Such daily snapshots convey the gist of the weather conditions that occurred but are not sufficient in scope to resolve changes in energy use as a function of the time of day. As renewables become more prominent in California's electricity market, highly accurate, high temporal resolution is required to ensure a stable power supply during extreme weather conditions. Such high-resolution information is needed to support the CEC efforts to progress to a more granular, more temporal, more data-intensive approach and meet the need for higher-resolution weather data in support of SB 100 and SB 350 goals.[1]

Stated Need for Subdaily Weather Data

The project's initial focus was on assessing the ability of statistical models to predict mean monthly temperatures on seasonal and decadal timescales. Through meetings facilitated by the CEC, the TAC explicitly advised that sub-seasonal (including subdaily) timescales were

---

[1] Commissioner J. Andrew McAllister. March 4, 2019. IEPR Commissioner Workshop on Data Inputs and Assumptions for 2019 IEPR Modeling and Forecasting Activities. Link to transcript: https://ww2.energy.ca.gov/2019_energypolicy/documents/#03042019

important, necessitating a deep and comprehensive examination of the sources and quality of subdaily weather information. This required refining the approach to data acquisition and quality control because neither the state nor any stakeholders had a reliable repository of subdaily weather data available. This stated need for higher-resolution weather and climate information included a desire for sub-seasonal predictions beyond the monthly mean temperature, which is often the predicted value produced by the weather and climate community but not actionable for utilities.

Meeting the Need for Subdaily Weather Data

Weather data have been systematically recorded on a subdaily (hourly) basis in California since 1933. For nearly 60 years, skilled weather observers took observations each hour, encoded the observations, and either archived them manually or transmitted them to a central repository for safekeeping. In the late 1990s, weather observations began to be taken by automated machinery, and human weather observers were stationed only at the largest airports in the state. While this record is long, there are many challenges to overcome in using it, including accounting for the movement or closure of stations and changes in instrumentation and observational practices. In this report, the methods for sourcing observations are described, which yield a curated dataset of stable long-term observations suitable for the energy sector.

Generation of a collection of weather stations that are representative of the diverse climatic zones of California will enable new and transformative analyses to be performed in the electricity sector. One such analysis requested by the TAC is an hour-by-hour, day-by-day analysis of changes in temperature across the state. Utilities advised us that subdaily trends are poorly understood and of great interest in the energy sector. Globally, it is well established that nighttime temperatures are warming faster than daytime temperatures (IPCC 2014); however, regionally some differences may occur, particularly in marine-influenced regions. Asymmetric warming across the day has implications for electricity production and demand, and must be quantified.

On the Predictability of Temperature at Seasonal Timescales in California

The grant-funding opportunity that was the funding mechanism for this work laid out the case and need for seasonal predictions of temperature and how such work would benefit ratepayers and the energy sector.[2] Funding was predicated on a unique requirement — that the approach must be computationally light — that is, capable of running on a standard desktop computer. Thus, statistical means are sought to make predictions, rather than relying on computationally heavy weather or climate models. The chosen statistical approach relies upon three key factors: (1) a source of predictability of air temperatures in California, (2) a statistical transform of the source of predictability of air temperature, and (3) a method of predicting future conditions of the source of predictability. In this work, ocean conditions are relied upon as the source of predictability; singular value decomposition (SVD) as the statistical transform;

---

[2] California Energy Commission, Reduce the Environmental and Public Health Impacts of Electrical Generation and Make the Electricity System Less Vulnerable to Climate Impacts: Phase II, October 2015. Publication Number: GFO-15-309.

and linear inverse modeling as the source of predictability. Briefly, these concepts are introduced below, and expanded upon in later sections.

Predictability and Sea Surface Temperature

The atmosphere has low or no memory — or statistically speaking, the autocorrelation of atmospheric variables rapidly declines on daily to weekly scales. What this means for forecasting is that the current state of the atmosphere tells us very little about its future state. In contrast, the high heat capacity of the ocean results in significant oceanic memory and potential predictability. The surface of the ocean drives atmospheric circulation through heat fluxes. This makes ocean surface temperature (or SST), which affects the atmosphere and is predictable, a strong choice for a predictor variable.

Generating Predictions via Singular Value Decomposition

In this work, SVD is used as the multivariate statistical analysis of covariance between SST and the temperature. SVD analysis identifies the time series (principle components or PCs) and corresponding spatial patterns (loading patterns) that describe the maximum covariance between the datasets. In other words, SVD analysis can reduce noise in the data and find the most energetically important coincident patterns of SST and predictor variable anomalies. SVD analysis is analogous to what is sometimes referred to as maximum covariance analysis and is similar to the commonly used empirical orthogonal function (EOF) analysis, in that it produces time series (principle components) and spatial patterns (loading patterns). However, SVD works through decomposition of a cross-covariance matrix, whereas EOF analysis decomposes a covariance matrix comprised of one spatiotemporal field. Therefore, SVD is the advantageous approach for predictive work because users can take advantage of its multiple covariance matrices to sum the predictions across all spatial-temporal modes identified as significant.

Linear Inverse Modeling

Linear inverse modeling (LIM) is the method of choice because it is considered the state-of-the-science approach to time series modeling and allows us to efficiently reduce an otherwise unwieldy number of degrees of freedom (time, space, and atmosphere-ocean physics) to a quantity manageable on a desktop computer. LIM is currently used by the National Oceanic and Atmospheric Administration’s (NOAA's) Climate Prediction Center in their seasonal and annual predictions of the El Niño Southern Oscillation. Further, LIM represents an optimal way to forecast a time series. Consider, as an example, the task of forecasting temperature everywhere in California, on a regular 0.5 x 0.5–degree resolution grid, requiring approximately 100 grid points. To model this directly it is necessary to train and run a forecast model on each pixel independently. However, SVD analysis identifies the top 5 to 10 principle component time series (sufficient to explain 85 to 90 percent of the variance of the predictor variable). By focusing on forecasting these 5 to 10 time series, this approach can effectively “solve” the forecast (through the LIM model) for all 5 to 10 time series simultaneously, because each time series is associated with a known spatial pattern. In doing so, the dimensionality of the problem can be reduced from 100 to 5. What is removed by this reduction in dimensionality is atmospheric and oceanic data noise, which is by definition unforecastable. As is evident by this example, LIM allows for computational time and efforts to focus on predicting the key 5 to 10 modes by which the atmosphere and ocean are actually covarying.

It is noted that over the course of proposing and performing this work and ultimately writing this final report, LIM has come to be used in many more applications and its utility in predicting ocean state and responsive air temperature has become better known. Dias et al. (2019) provide an excellent overview of the utility of this approach, and an open-source programming package for Python, which is capable of generating predictions of Pacific Ocean SSTs, has been made freely available (https://github.com/frodre/pyLIM). These and other external efforts allowed us to redirect time and effort to providing subdaily weather information, in support of needs stated by the project TAC and CAM.

CHAPTER 2:
Project Approach

Subdaily Weather Data for the Energy Sector

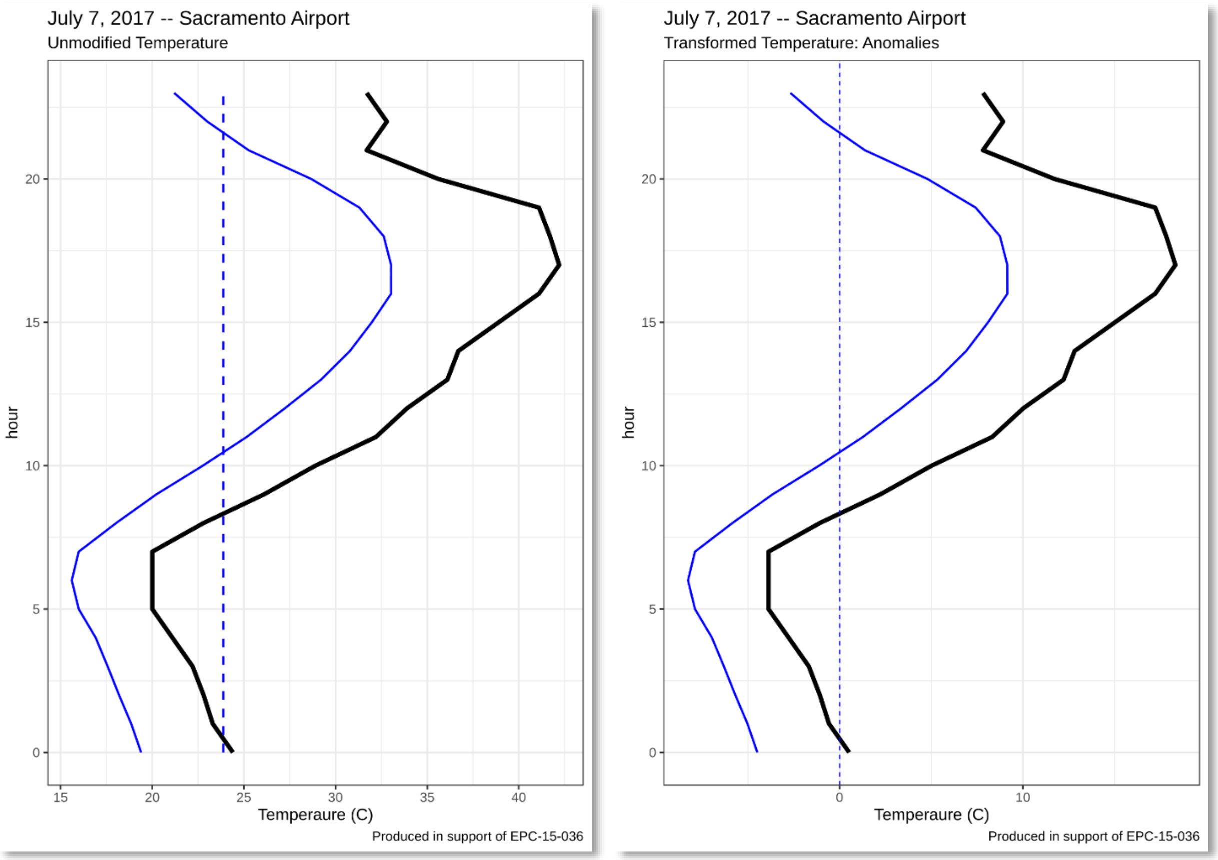
On the Presentation of Temperature by Hour

Temperature can be represented in multiple formats through normalization or statistical transformations, and such formats should be matched to the data's intended end use purpose. In this study, three formulations of temperature at each hour in the record for which data exists are provided: raw temperature data, daily anomalies, and hourly anomalies. The concept of raw data is straightforward — the temperature (in degrees Celsius) as recorded by instrumentation. Daily anomalies consist of the temperature minus the long-term daily mean value. Hourly anomalies are the temperature minus the long-term hourly mean value for that day of the year. Hourly and daily anomalies are calculated relative to the 30-year period 1981 to 2010. All hours reported are converted to local Pacific time, with adjustments made for daylight savings.

Energy sector uses for hourly weather data vary widely, and users should spend some time considering the type of temperature information that most directly matches their end use. Some engineering applications require raw temperature to understand climate impacts on buildings and infrastructure. Demand forecasts might require information about the diurnal cycle — that is, the daily shape of hourly weather data — to, for example, understand impacts of compound heat events on load. More granular analyses might require hourly anomalies to assess short-term impacts of extreme conditions that deviate from typical values found at a given hour and day of the year. Figure 1 shows two methods of visualizing different formulations of the same observations.

A large NetCDF database of hourly temperature data has been provided along with this final report to the CEC. The data are provided for the period from 1973 to the present (July 2019). In addition, R databases have been provided in recognition of the importance of this programming language to the Demand Forecast Office (DFO). The NetCDF database was created to meet Cal-Adapt standards and expectations for data formatting and metadata. Chapter 4 outlines these data products and provides a plan for their immediate transfer to Cal-Adapt. Chapter 5 provides recommendations for how this information can be distributed via Cal-Adapt's API and data-download tool, incorporated into existing products, and ideas for tool support for this product. Further, this database can easily be amended when new data become available. Chapter 5 discusses how this can be achieved.

Figure 1: Temperature Profiles



Raw (left) and transformed (right) temperature data for July 7, 2017 in Sacramento.

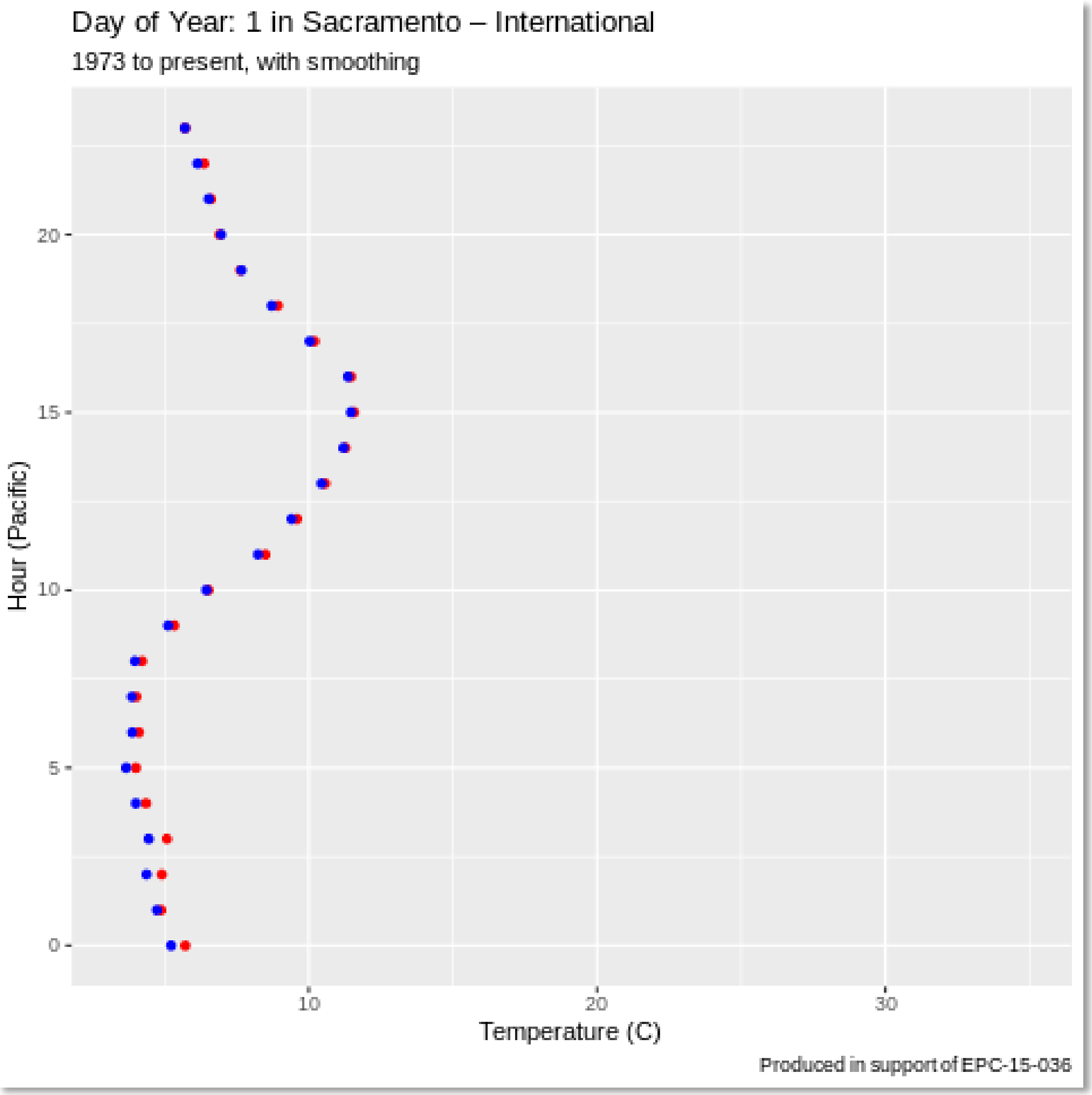
Source: Eagle Rock Analytics

Formulation of Typical Day Profiles

At any given location in California, the rate of warming and cooling that a location typically experiences changes day by day and hour by hour in response to solar forcing, regional weather, and microclimatological factors. Characterizing the typical thermal profiles, and how temperature changes hour by hour, is helpful in several planning and design applications and is often a necessary input to many physical and mechanical models. For a subset of locations (39) in California that possess a long-term, high-quality hourly temperature record, "typical day" profiles are produced for each day in the year.

Weather data, even over long periods of time, often contains significant noise from natural (for example, day-to-day weather variability) and manmade (for example, shifts in instrumentation or changes to surrounding environment) factors. To ensure the profiles are representative of the local climatology with low impacts from noise, a low-pass filter is applied to the hourly data. This filter (locally estimated scatterplot smoothing, or LOESS) considers the surrounding 18 days (~5 percent) and gently smooths the diurnal curves to be representative of long-term forcing. An example of such a profile is presented in Figure 2. Such profiles are appropriate for studies seeking the most common conditions for a given day, but the raw data should be used for studies needing variability associated with weather events.

Figure 2: Typical Day Formulations



The long-term mean temperature by the hour in Sacramento (blue) and a smoothed long-term mean temperature after LOESS filtering for Sacramento (red).

Source: Eagle Rock Analytics

A database containing typical day diurnal curves of temperature at locations across California has been produced and will be submitted to CEC along with this final report. This record contains the raw and daily temperature anomalies for 39 stations, 366 days and 24 hours.

Assessing Changes in the Diurnal Structure of Temperature in Response to Climate Change

To determine how a change in climate is affecting the diurnal characteristics of temperature, the linear trend in each hour of every day of the year is estimated. Trends are calculated on a monthly basis, by averaging across all days in a month at a given hour and location. Regional and statewide assessments were performed by averaging all sites within the state/region, and then performing a linear trend analysis. Linear analyses are calculated over the entire record (1973–present) at each station. To understand if changes in instrumentation (that is, installation of automated systems in the late 1990s) were important, the record is divided into the "manual observation period" of 1973 to 1996 and the "automated observation period" of 1997 to 2019. Many, but not all, stations in the record have manual observations augmenting automated data in the later period.

Linear models were constructed for each month-hour combination at each location. Summaries of key results are presented in Chapter 3, with complete monthly analyses provided as supplementary figures. Complete model repositories are saved as an R database for later use at the DFO, as described in Chapter 4.

Seasonal Predictions of Climate for Energy Sector

Developing a Linear Inverse Model

A linear inverse model was developed in MATLAB and trained on monthly ERSST v3 sea surface temperature data. These data were detrended in preprocessing before model training occurred. The domain selected for predictions spanned the region from -10°S to 65°N, and from the international dateline (at 180°E) to 150°W. A number of different forecast length periods were tested, and ultimately the LIM was developed for zero to nine months out. Propagation matrices were generated for each month. Subsequent to this work, previously referenced open-source software packages performing this analysis were developed and potential users are referred to this package (see Chapter 1).

Singular Value Decomposition

Code to perform the SVD analysis was developed in Python, in a domain centered over California, with ocean points excluded from the analysis. Maximum daily temperature values in a number of reanalysis products were assessed for covariation with SST over the LIM domain. The top 10 modes explaining the most variance were determined for each month in the year. Predictions of air temperature for future months were calculated for each month by applying output from the LIM to each of the sets of monthly modes.

CHAPTER 3:
Project Results

Stakeholders, including investor-owned utilities (IOUs), publicly owned utilities (POUs) and CEC staff, clearly and directly indicated that sub-seasonal, reliable, and stable weather information is critically needed to achieve climate-related goals. Some highlights of this work related to the four main thrusts of this project are briefly outlined below: (1) producing a stable, curated, long-term subdaily weather product, (2) formulation of typical day and hour profiles, (3) an hourly trend analysis, and (4) assessment of the capacity for seasonal temperature predictions.

1. A stable, long-term hourly weather database was generated:
    - Extensive quality assurance and quality control was performed on the record.
    - An hourly, 47-year record (1973 to present) was produced at 39 locations across California.
    - Representative hourly stations were identified to replace the 19 daily stations used by the DFO.
    - Across California, the quality of weather observations has declined since Reagan-era federal cost-savings programs replaced trained weather observers with automated equipment.
    - The authors identified which quality assurance approaches that are applied to global-observation datasets are inappropriate for use in California.
2. Typical day formulations for temperature were created as a resource for energy-sector modeling studies that use climatological mean conditions:
    - A long-term mean is generated at each of the 39 stations for each hour-of-day, day-in-year combination.
    - Smoothing across the year was applied at each hour to reduce high-frequency noise and more faithfully reproduce long-term conditions.
    - Temperature information is presented in multiple formats, including raw values and anomalies relative to daily means, with guidance as to best usage practices.
3. The rate of change of temperature in response to anthropogenic climate change is carefully assessed on an hourly basis:
    - Statewide, the rate of warming is maximized in the hours just before sunset; this increase in temperature will frustrate utilities’ efforts to shift peak demand away from this part of the day.
    - Across California, the minimum increase in temperature is occurring just after sunrise; this is in contrast to many other locations in North America where nighttime lows (which typically occur at or just before sunrise) are warming more rapidly than daytime temperatures.
    - Important differences in warming rates are noticed across regions of California, but most stations fall into one of two categories: (1) stations where warming is

occurring fastest in the late afternoon and (2) coastal stations where warming is occurring rapidly in the morning and late afternoon.
    - Utilities should consider multiple stations in a given region to capture the climate signal and rate of warming and reduce the impacts of local non-climate signals.
4. The potential for generating seasonal predictions of temperature across California is assessed:
    - The Pacific Ocean has sufficient predictive capacity, so that if the future state of the ocean can be accurately predicted, so too can temperature in California.
    - LIM is able to make skillful predictions of the future state of the ocean under certain conditions, but not others.
    - A hybrid approach in which LIM is used when predictive skill is expected, and seasonal climate models when it is not, could be used to improve overall skill.
    - Utilities are interested in sub-monthly predictions because monthly mean temperatures (currently predicted by atmospheric scientists) are not necessarily the most useful variable for making energy sector predictions.

A Stable, Curated, Long-Term Subdaily Weather Product for California's Energy Sector

Subdaily Weather Observations

The Integrated Surface Database (ISD), maintained by NOAA, is the preeminent hourly weather repository. The most commonly collected weather parameters are temperature, dewpoint, wind (direction, speed, and gustiness), cloud cover, present weather, sea-level pressure, precipitation, and snow depth. Daily summary variables (for example, maximum and minimum temperatures) are also included in the record. The ISD has 54 QA/QC checks applied to it (Smith et al., 2011); however, it is designed to be conservative and significant amounts of spurious temperature results pass through unflagged. In this study, temperature data are from HadISD, a quality-controlled version of the ISD record (Dunn et al., 2012; Dunn et al., 2016; Dunn 2019). HadISD is maintained by the Hadley Centre in the United Kingdom. The ISD and HadISD are peer reviewed, with significant technical documentation as well as in some cases open-source code available.

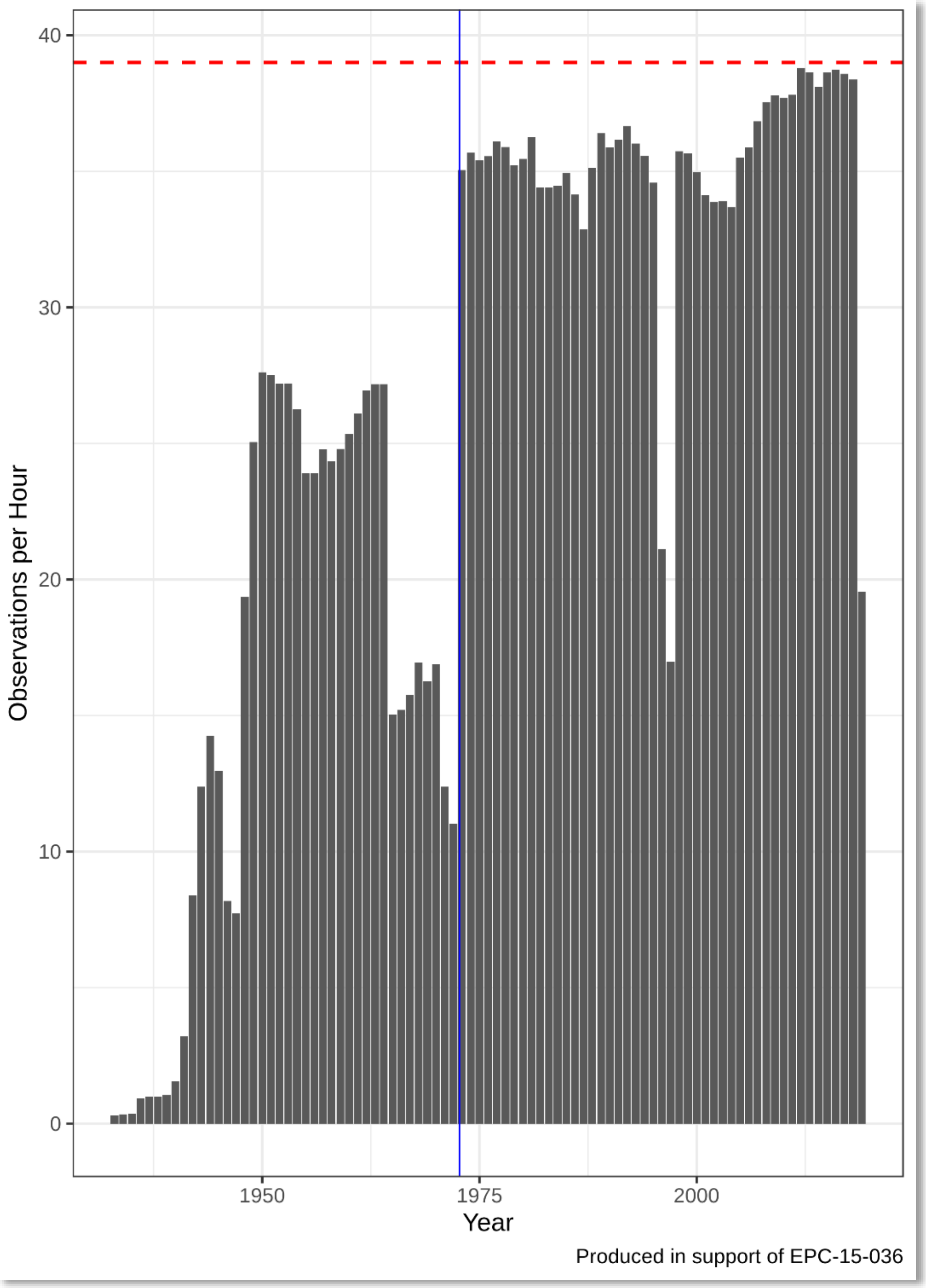
Ensuring a Stable, Quality Record

Sources and History of Hourly Weather Data

Primarily, hourly weather observations are taken at airports due to a need for high-quality, frequently updated conditions for aviation safety. Standardization of weather reporting dates back to the late 19th century, when the National Weather Service was organized to centralize and store weather observations disseminated by telegraph. In California, the hourly weather observation record begins in 1933 (Figure 3) at a handful of airfields, primarily those operated by the military. With the rise in civil aviation after World War II, the number of hourly records increased in California, with data stored locally on paper and archived. In the 1960s, the number of stations for which there is access to records declined as the transmission system transitioned from analog to digital data, with some stations' records being lost in the process.

In 1973, the Federal government began directly receiving and archiving hourly weather reports via the Global Telecommunications System, which resulted in a sharp rise in the number of stations with records. Engaged stakeholders noted that a 30-year record is the typical length considered for electrical sector studies. The dataset used here begins coincident with the 1973 onset of hourly data collection to ensure a record that is stable and of appropriate length.

Figure 3: Historical Observations

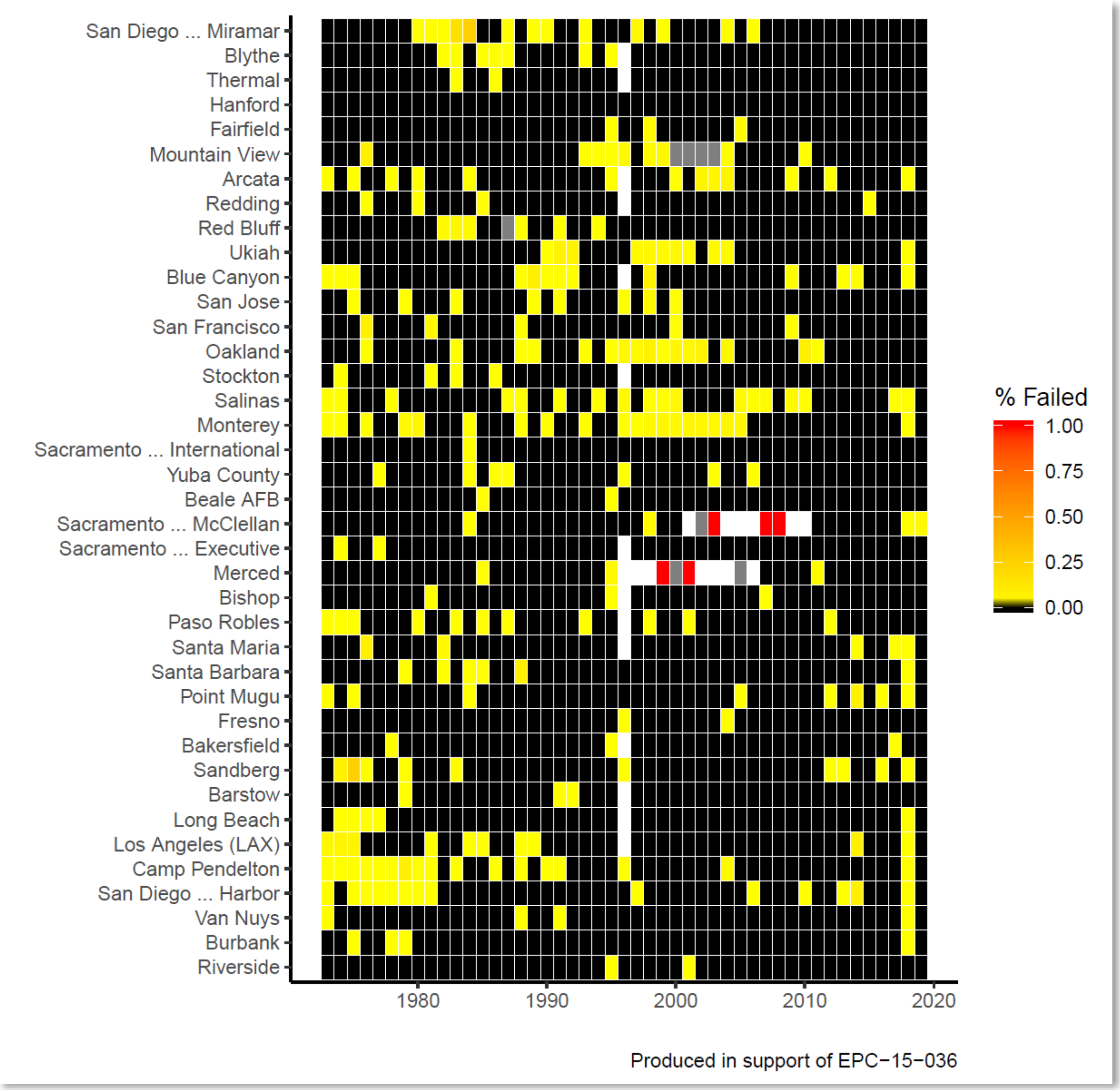


Number of observations in database at select stations across California. Vertical blue line denotes 1973, after which records are more consistently archived. Horizontal dashed redline represents the maximum number of observations.

Source: Eagle Rock Analytics

From the onset of recording until the 1990s, weather observations were taken manually, usually by trained and certified weather observers. Beginning in the late 1980s as part of a Reagan-era cost-cutting program, efforts began to automate weather observations, augmenting with manual weather observers only at the most critical airports. As a result, the quality of weather observations began to decline. Air traffic controllers are nominally required to perform basic augmentation and quality control on weather reports, but in practice are very poorly trained and too busy with more critical tasks to perform anything but basic checks on the weather reports. Moreover, a time-stressed air traffic controller may correct for a missing visibility or ceiling reading (weather variables critical for aviation) but are almost assuredly not going outside to manually record a temperature reading if the automated sensor is not functioning. As a result, the frequency of missing and poor-quality temperature observations increased beginning in the late 1990s (Figure 4).

Figure 4: Failed Temperature Observations

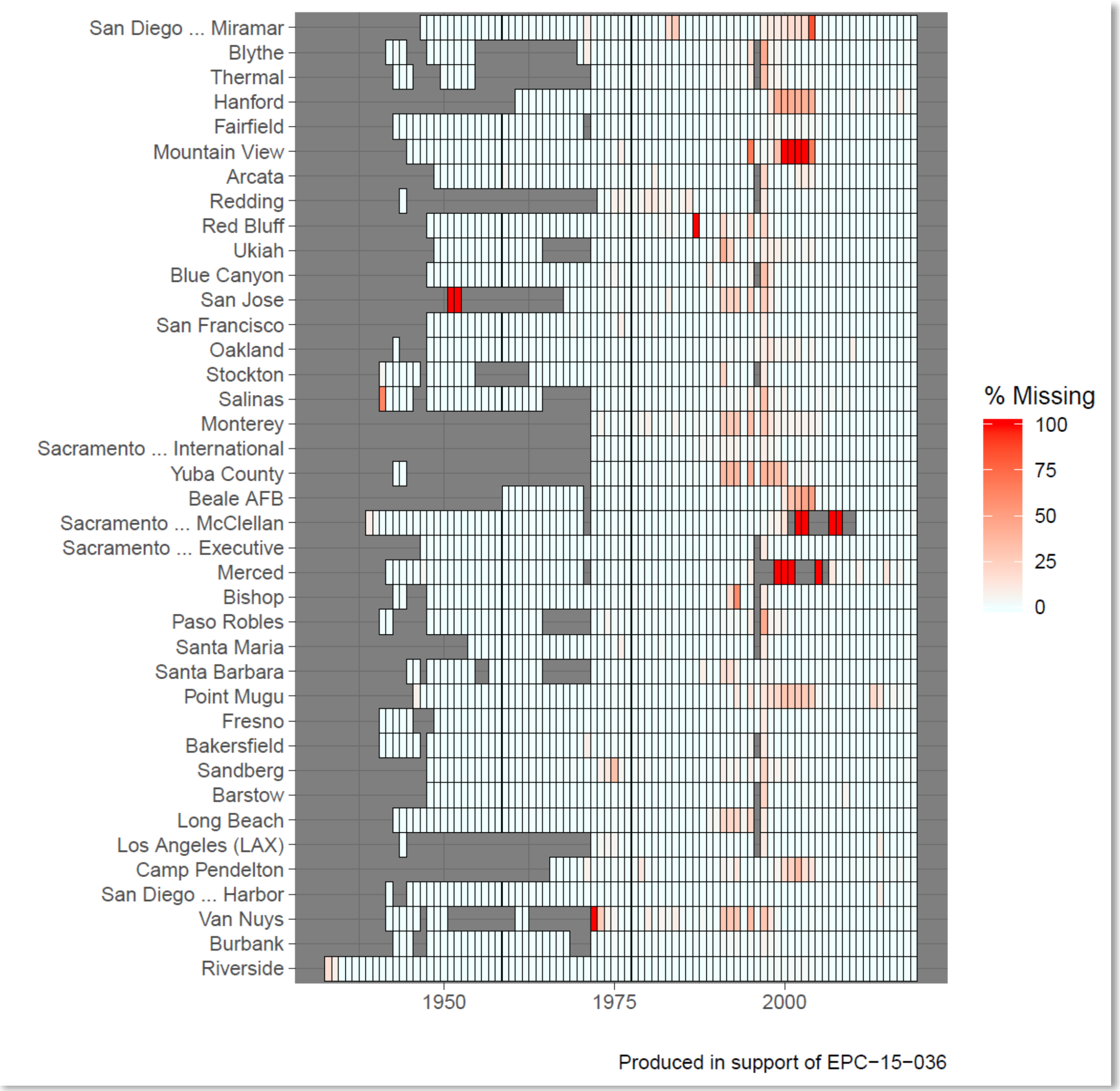


Fraction of observations failing QA/QC at each station over the period 1973 to 2019.

Source: Eagle Rock Analytics

Large chunks of the record are missing during the years 1998 to 2002, when instrumentation was installed and systems were temporarily taken offline (Figure 5). Changes in instrumentation can lead to jumps or nonphysical trends in the record. HadISD applies a pairwise homogeneity test, however does not bias-correct for trends introduced by station location changes (Dunn et al., 2014).

Figure 5: Missing Observations



Fraction missing at each station over the period 1973 to 2019. The frequency of missing observations increases in the late 1990s and early 2000s during the transition from manual observations to automated observations.

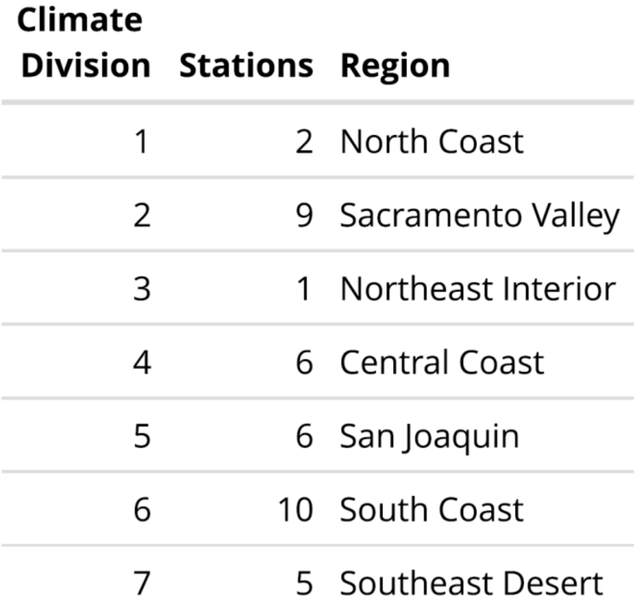
Source: Eagle Rock Analytics

Weather Stations in California

Weather stations come on and offline as population distributions and infrastructure change. Gathering long-term stable records can be a challenge, particularly away from population centers. Weather stations in underrepresented locations that are poorly resolved by climate and weather models (such as Mount Wilson and Mount Shasta) have come offline in recent years, critically reducing the understanding of diurnal temperature characteristics in rural regions. While demand forecasting focuses on urban, highly populated regions, much of California's vulnerable transmission infrastructure is sited in locations for which insufficient weather data exist.

The database produced in this study is large, and at 39 stations with a 47-year record contains significant actionable information; however, not all climatic regions, demand forecast zones; or utility districts are well represented by the database. Table 1 shows the number of long-term stations divided into California's climate zones as defined by NOAA. Densely populated regions such as the Sacramento Valley and the Central and Southern coasts are well represented; however, the North Coast and Northeastern Interior are poorly represented.

Table 1: Stations by Climate Division



Climate Division	Stations	Region
1	2	North Coast
2	9	Sacramento Valley
3	1	Northeast Interior
4	6	Central Coast
5	6	San Joaquin
6	10	South Coast
7	5	Southeast Desert

Number of stations in each climate division, as defined by NOAA.

Source: Eagle Rock Analytics

Matching Sites to the Daily Sites Used by the Demand Forecast Office

Historically, the CEC's Demand Forecast Office used 19 stations that reported daily maximum and minimum temperatures across the state in various analyses and forecast applications. Presented here is a list of curated hourly weather stations that represent the best climatological fit to replace daily stations that lack concurrent hourly observations. Twelve of 18 stations shared siting with daily observations. Appropriate stations that experience roughly similar climate conditions for six of the remaining stations (El Centro, Merced, Modesto, Riverside, Paso Robles, and Santa Barbara) are identified. Note that there is a large geographical distance between El Centro and San Luis Obispo replacement stations, but these hourly stations experience similar diurnal variability and sensible weather features. El Centro and Merced had similarly sited hourly records, but these stations are deemed to be of low quality and using them is not advised. One daily station, El Cajon, did not have an analog hourly station that would be an appropriate replacement (however, for studies requiring a shorter record, Gillespie Field could be used as a source of hourly records, but is inadequate for long-term analysis). A summary of recommended sites is presented in Table 2.

Table 2: Recommendations for Hourly Stations

Daily Station	Hourly Station	Distance (km)	Nearest City	Airport ID
BAKERSFIELD AP	Meadows Field Airport	0.00	Bakersfield	KBFL
BURBANK VLY PUMP PLT	BURBANK-GLENDALE-PASA ARPT	1.83	Burbank	KBUR
EL CAJON	NA	NA	NA	NA
El CENTRO NAF	DESERT RESORTS RGNL ARPT	116.46	Thermal	KTRM
FRESNO YOSEMITE INTL AP	FRESNO YOSEMITE INTERNATIONAL AIRPORT	0.04	Fresno	KFAT
LONG BEACH DAUGHERTY FLD	LONG BEACH/ DAUGHERTY FIELD/ AIRPORT	0.05	Long Beach	KLGB
MERCED	CASTLE AFB	11.88	Merced	KMER
MODESTO CITY CO AP	Stockton	38.12	Stockton	KSCK
OAKLAND METRO INTL AP	METRO OAKLAND INTL AIRPORT	0.05	Oakland	KOAK
RED BLUFF MUNI AP	RED BLUFF MUNICIPAL ARPT	0.04	Red Bluff	KRBL
RIVERSIDE FIRE STN 3	MARCH AIR RESERVE BASE	13.97	Riverside	KRIV
SACRAMENTO EXECUTIVE AP	SACRAMENTO EXECUTIVE AIRPORT	0.01	Sacramento	KSAC
SAN DIEGO LINDBERGH FLD	SAN DIEGO INTERNATIONAL AIRPORT	0.05	San Diego-Harbor	KSAN
SAN DIEGO MIRAMAR NAS	MARINE CORPS AIR STATION	0.04	San Diego-Mesa	KNKX
SAN LUIS OBISPO POLY U	PASO ROBLES MUNICIPAL ARPT	40.68	Paso Robles	KPRB
SANTA BARBARA	SANTA BARBARA MUNICIPAL AIRPORT	14.60	Santa Barbara	KSBA
UKIAH MUNI AP	UKIAH MUNICIPAL AIRPORT	0.03	Ukiah	KUKI
SAN JOSE INTERNATIONAL AIRPORT CA US	N.Y. MINETA SN JO INTL APT	0.02	San Jose	SKJC

Recommended hourly weather stations to use in place of daily reporting stations.

Source: Eagle Rock Analytics

Quality Assurance Tests for California

In this study, 39 stations suitable for analyses that require long-term, stable hourly temperature records are identified (Table 3, Figure 6).

Table 3: Table of Stations

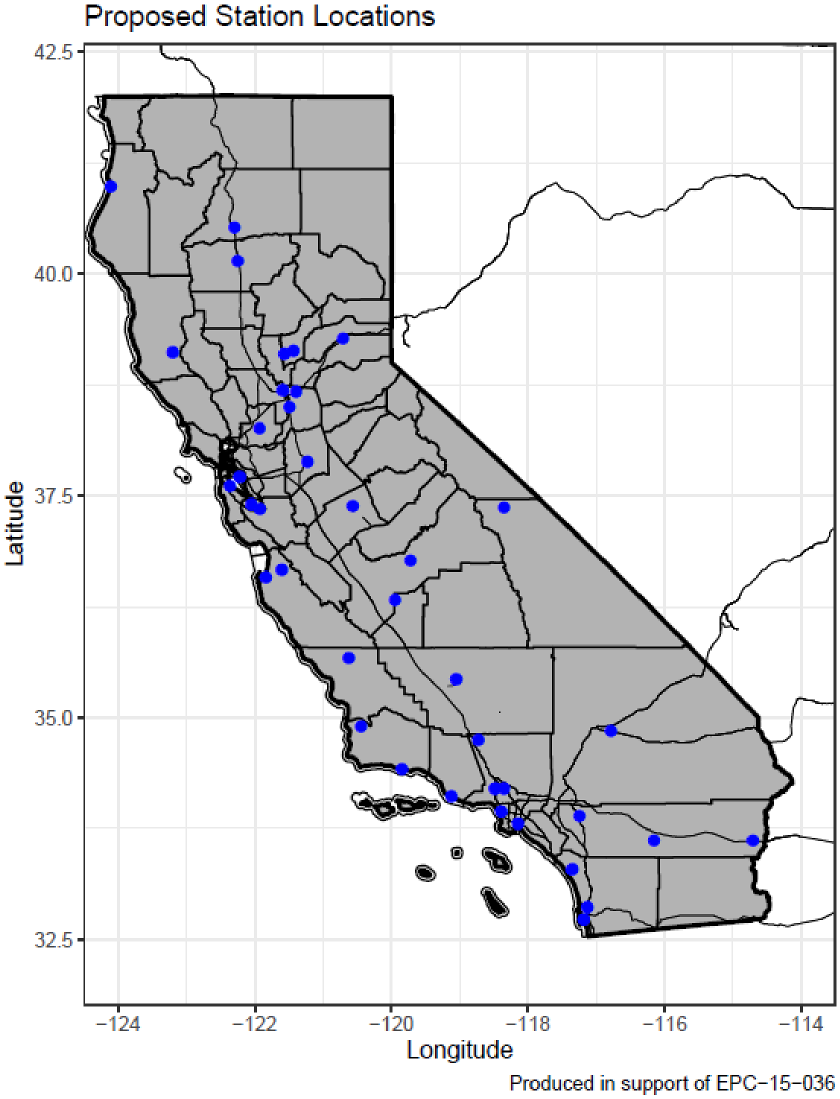
Station	Latitude	Longitude
Riverside	33.900	-117.250
Burbank	34.201	-118.358
Van Nuys	34.210	-118.489
San Diego- Harbor	32.734	-117.183
Camp Pendleton	33.300	-117.350
Los Angeles (LAX)	33.938	-118.389
Long Beach	33.812	-118.146
Barstow	34.854	-116.786
Sandberg	34.744	-118.724
Bakersfield	35.434	-119.054
Fresno	36.780	-119.719
Point Mugu	34.117	-119.117
Santa Barbara	34.426	-119.843
Santa Maria	34.899	-120.449
Paso Robles	35.670	-120.628
Bishop	37.371	-118.358
Merced	37.383	-120.567
Sacramento- Executive	38.507	-121.495
Sacramento- McClellan	38.667	-121.400
Beale AFB	39.133	-121.433
Yuba County	39.100	-121.567
Sacramento- International	38.696	-121.590
Monterey	36.583	-121.833
Salinas	36.664	-121.608
Stockton	37.889	-121.226
Oakland	37.721	-122.221
San Francisco	37.620	-122.365
San Jose	37.359	-121.924
Blue Canyon	39.277	-120.710
Ukiah	39.126	-123.201
Red Bluff	40.152	-122.254
Redding	40.518	-122.299
Arcata	40.978	-124.109
Mountain View	37.406	-122.048
Fairfield	38.267	-121.933
Hanford	36.333	-119.950
Thermal	33.627	-116.159
Blythe	33.619	-114.714
San Diego-Miramar	32.867	-117.133

Location of the 39 stations used in this study.

Source: Eagle Rock Analytics

These stations meet the following qualifications: (1) possess a long-term stable record, (2) observations taken over full 24-hour periods, (3) observations consistently taken throughout the year, (4) observations are rarely missing, and (5) observations are rarely erroneous. How these aforementioned conditions are systematically tested for is discussed below.

Figure 6: Map of Station Locations



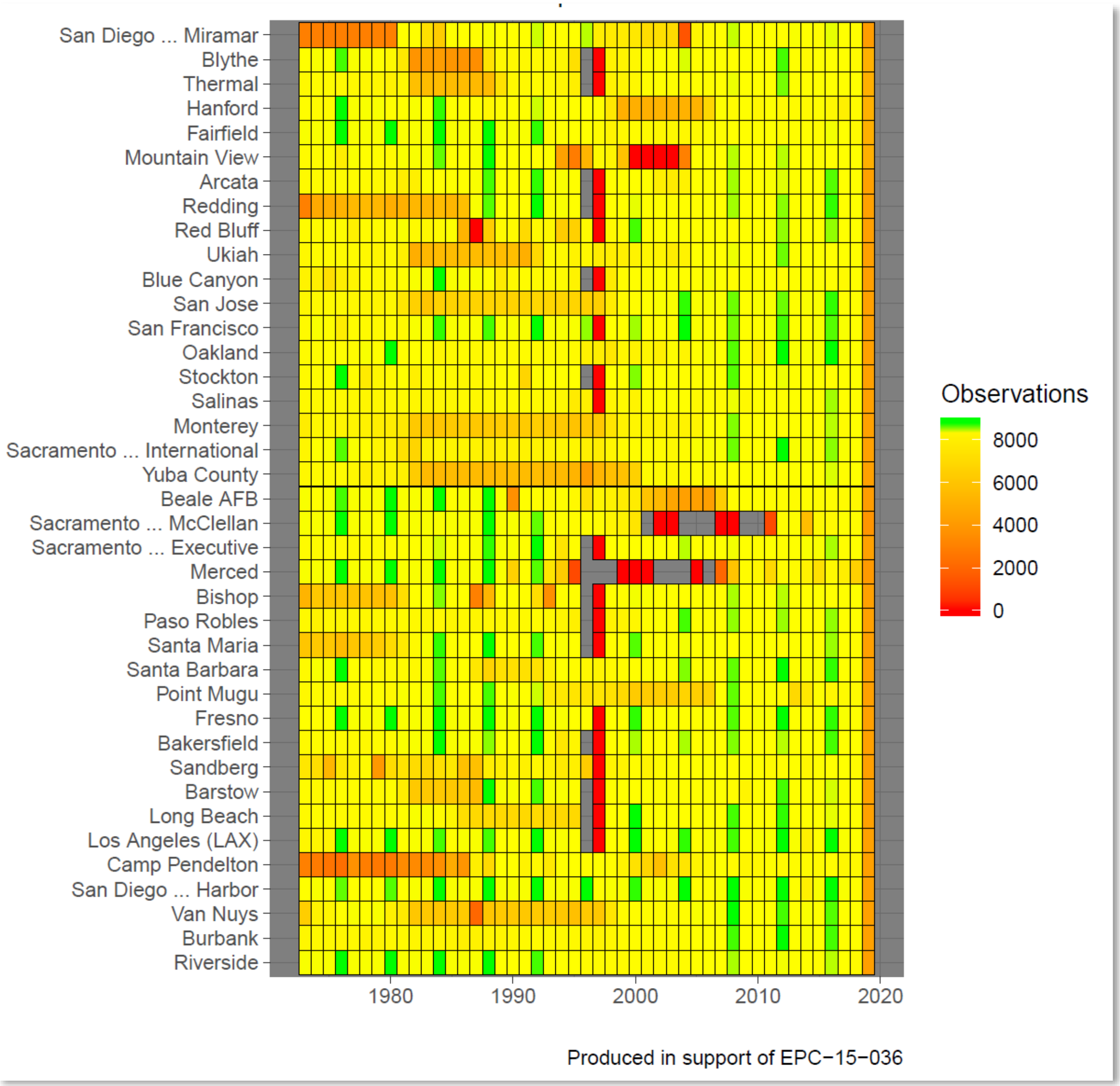
Map showing the locations of 39 hourly weather stations.

Source: Eagle Rock Analytics

Long-Term Stable Records

Stations were included in the database if and only if the station record began in 1973 or earlier and ran through present. Stations were required to have at least 30 nonconsecutive complete years in the record. Figure 7 shows the number of observations per year for the stations that were selected over the period 1973 to present. Many stations had data gaps during the period of record, particularly in the late 1990s due to instrumentation changes. Sacramento–McClellan and Merced had mid-record gaps but were included because they had more than 30 complete years of data.

Figure 7: Observations per Year



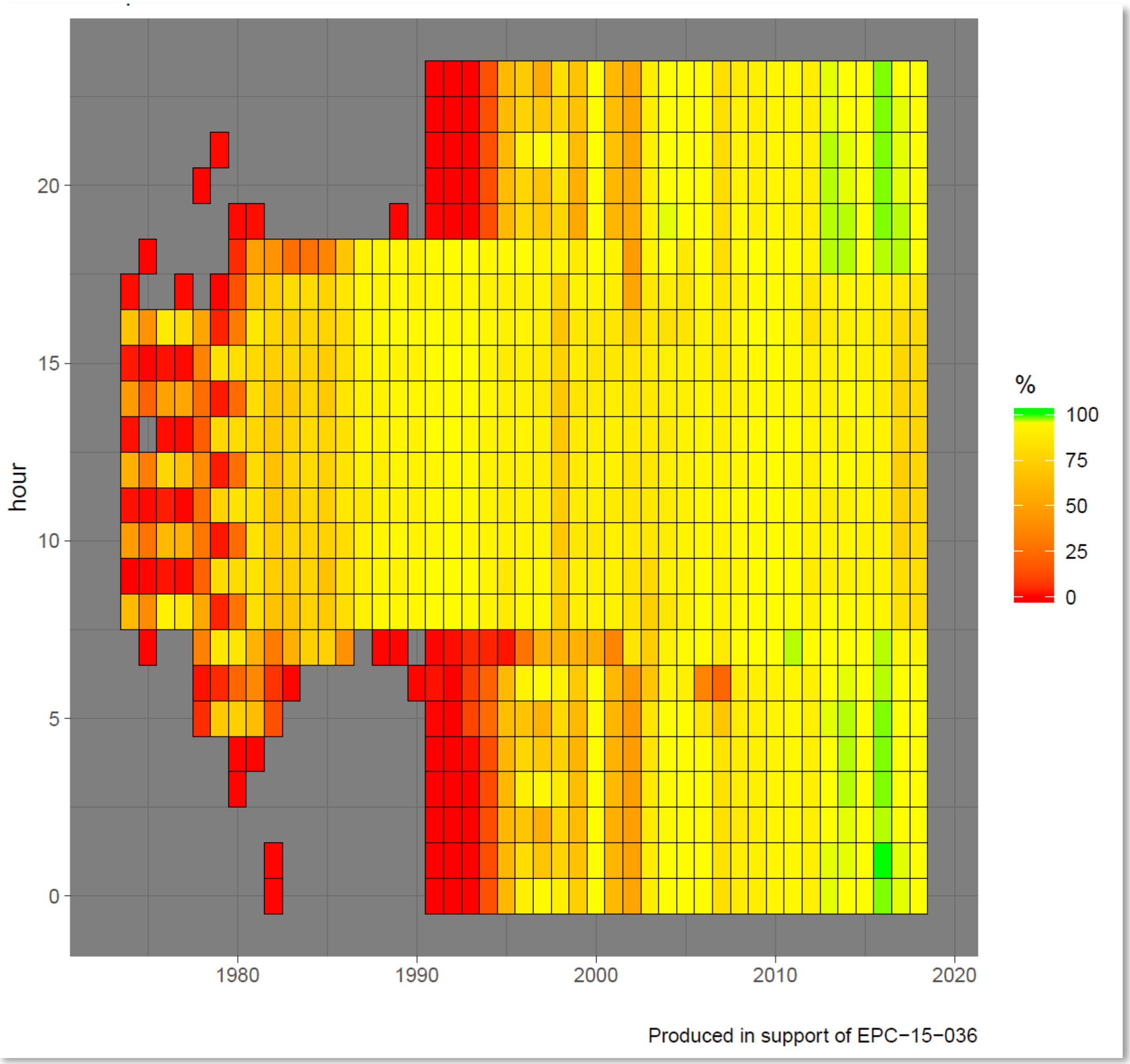
Observations per year at each station: yellow represents one observation per hour, green denotes more than one observation per hour, and orange and red represent increasing frequency of missing observations.

Source: Eagle Rock Analytics

Complete Daily Records

Many regional airports and most military airfields close during overnight periods (midnight to 6am) and prior to the automated sensor period (pre-1997), no observations were taken while the airports were closed. Such regular data gaps make creating diurnal thermal profiles and trend analysis challenging or potentially inappropriate. Regional airports that failed this test (such as Truckee-Tahoe Airport, Figure 8) were often found in rural regions, reducing the coverage in these underrepresented regions.

Figure 8: Observations by Hour at Truckee-Tahoe Airport



Percentage of maximum possible observations by hour and year at Truckee-Tahoe Airport. This sort of profile is typical of a regional airport that was closed during nighttime hours. Full diurnal reporting at Truckee-Tahoe Airport began after automation in 1995.

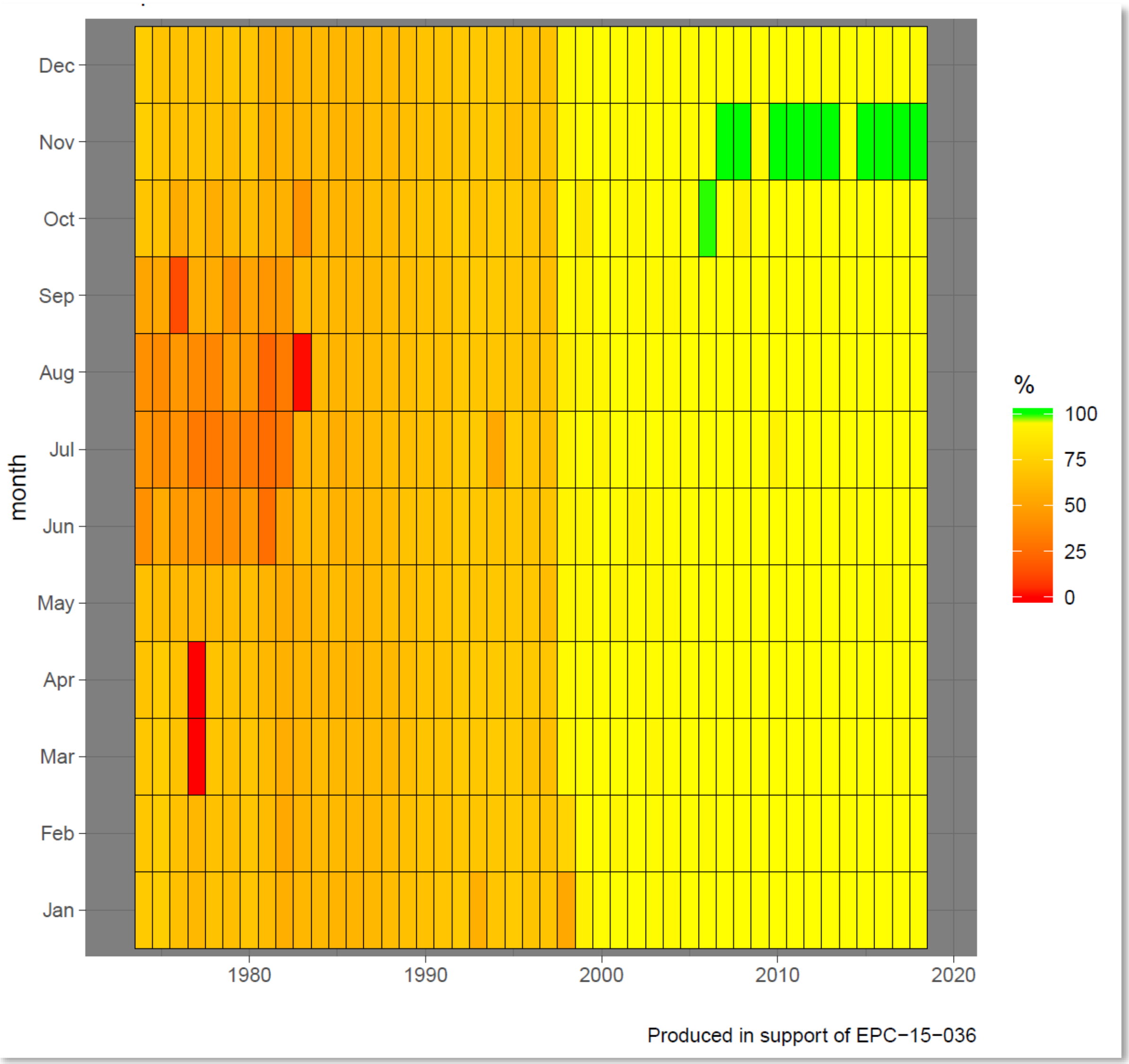
Source: Eagle Rock Analytics

Stations were required to have 24-hour coverage in at least 66 percent of days for at least 30 nonconsecutive years to be deemed to have complete daily records. Some stations, such as San Jose, have short periods of reduced or no observations during the overnight hours. Such stations may have less reliable diurnal profiles or trend analyses associated with them. Users looking at individual stations are advised to carefully review the frequency of nighttime observations (as in Figure 8) before drawing conclusions.

Stability Across Seasons

Some regional airports may see less frequent flight operations in a given season, and thus less frequent weather observations. This occurs more often in regions with seasonal tourism (for example, Palm Springs in summer; Figure 9), and typically before the deployment of automated weather stations (post 1996). Stations were required to have full 12 months of coverage (66 percent of observations in a month) for at least 30 nonconsecutive years in the record to be deemed seasonally stable. Users looking at individual stations in regions in which tourism varies seasonally should carefully review the frequency of seasonal observations (as in Figure 9) before drawing conclusions.

Figure 9: Observations by Month at Palm Springs



Percentage of maximum possible observations by month and year at Palm Springs. A decline in the number of observations during the summer season is observed in the early record, typical of regional airports where use varies seasonally due to tourism.

Source: Eagle Rock Analytics

Infrequently Missing Observations

Stations where observations are frequently missing are likely to be of lower quality, with hourly observations potentially attributed to the wrong hour in the manual observation era and indicative of lower quality, less frequently serviced equipment (AWOS vs. ASOS) in the automatic observation era. Stations that lacked observations in greater than 20 percent of potential hours were deemed to be missing too many observations to be included in the record. Users looking at individual stations for temporal subsets of the record should check to ensure sufficient information is available during that period (by, for example looking at Figure 8 and Figure 9 for their location of interest) before using the data.

Table 4: Missing Observations by Station

Station	Missing Obs	Total Obs	Missing Rate (%)
San Francisco	1542	619464	0.25
Bakersfield	1586	648696	0.24
Barstow	1395	603360	0.23
San Diego- Harbor	1005	656688	0.15
Sacramento-Executive	894	615096	0.15
Fresno	330	641400	0.05

Rate of missing observations at six stations with highest missing rates.

Source: Eagle Rock Analytics

Few Data Quality Issues

Stations exhibiting frequent removal of observations for failing QA/QC routines could indicate lower quality, less frequently serviced equipment or instrumentation that is frequently moved and thus not appropriate for analysis. Stations failing QA/QC tests more than 2 percent of the time were deemed of poor quality and removed. Chapter 5 includes a discussion of the potential for bias in globally developed QA/QC routines that may make their use in California less appropriate.

Table 5: Failure by Station

Station	Flagged Obs	Total Obs	Flagged %
Monterey	3656	358655	1.02
San Diego-Miramar	4742	555560	0.85
Oakland	3738	577870	0.65
Sandberg	2763	481973	0.57
Ukiah	2810	499171	0.56
Los Angeles (LAX)	2185	396795	0.55

Rate at which observations failed QA/QC at six stations with highest failure rates.

Source: Eagle Rock Analytics

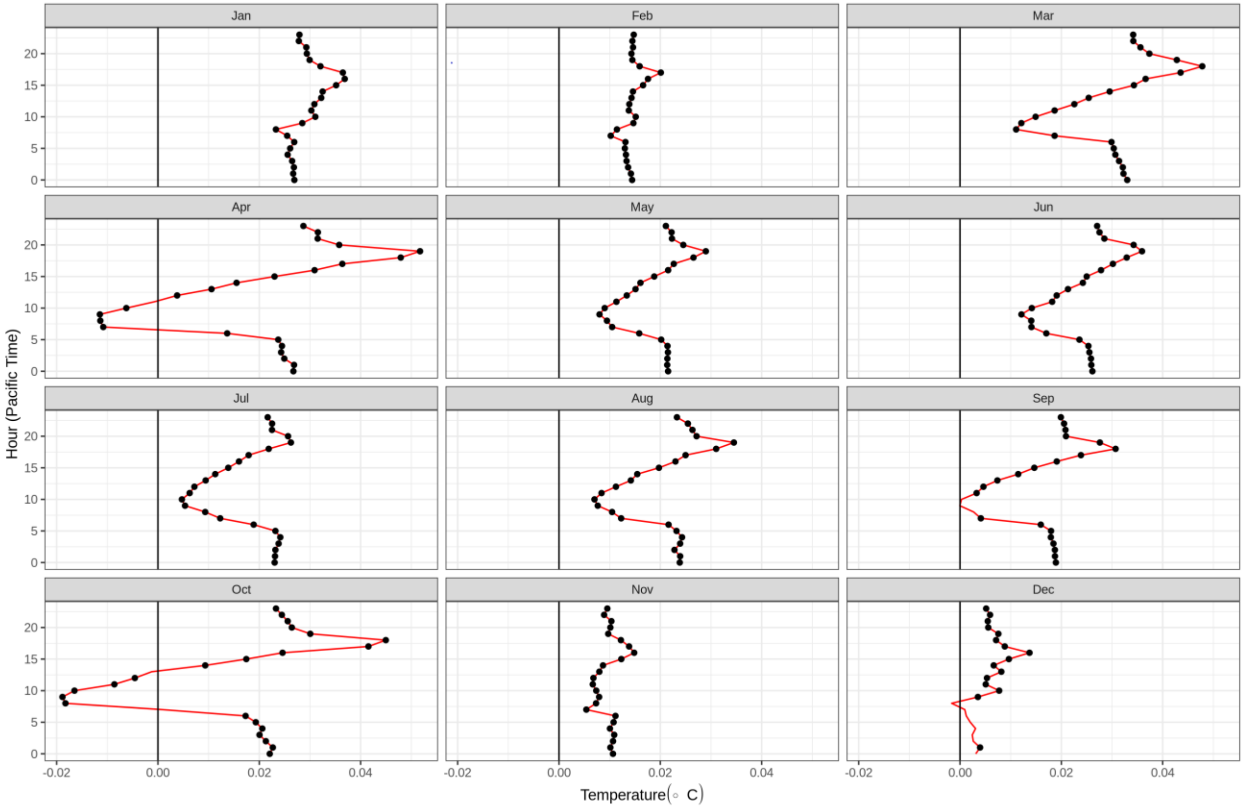
Changes in the Daily Cycles of Temperature Across California in Response to Climate Change

California is warming at a rapid rate as a response to anthropogenic climate change, but that warming is not uniform, with large differences across California’s complex topography and differences across the span of the day. The time of the day in which warming is occurring is of critical importance as California shifts to a high-renewables grid that relies heavily on renewable generation that varies in productivity over the course of the day (in addition to seasonal variation). Broadly, California is warming most rapidly in the evening and very early morning hours, with the least warming observed around or shortly after sunrise. This trend is detected in the aggregate of all observations, but regionally critical differences emerge. In this section, key statewide and regional trends are discussed, and how the observational record may (or may not) affect results.

Statewide Trends by Season

Statewide trends are calculated by averaging hourly temperature anomalies across all 39 stations, for each month (Figure 10). Because the weather stations are unevenly distributed geographically, the results are weighted toward highly populated regions where airports tend to be located. This is arguably helpful because the trend analysis is more reflective of where energy is being consumed.

Figure 10: Statewide Trend by Month



Linear trend by month.

Source: Eagle Rock Analytics

Winter (December, January, and February)

The winter period features a relatively uniform warming across the day, with significant trends of a similar magnitude across all hours. Trends are slightly larger in the late afternoon (hours 16 — 17) and reduced in the early morning (around hour 8). Trends are broadly 0.01°C to 0.03°C (0.02°F to 0.05°F) per year, with a peak warming in January and less warming in December. It is noted that the total warming over the observed period can be loosely estimated by multiplying the warming rate (in degrees Celsius per year) by the number of years (47) in the record.

Spring (March, April, and May)

Spring shows a distinct diurnal pattern in trends, a clear strong warming trend in the late afternoon (hours 15 — 20) and a lack of warming (or even cooling) in the early morning (hours 6 — 10). The shape is consistent through meteorological spring, with April showing the greatest extremes — the largest warming (more than 0.05°C (0.09°F) per year) in late afternoon and the greatest cooling (–0.01°C (-0.02°F) per year) in the hours after sunrise. The shape of the warming trend in spring was similar to that observed in fall, with April and October, and May and September, showing remarkably similar shapes and magnitudes. Peak warming rates seem to track with changes in sunset, starting at hour 16 in January, hour 17 in February, hour 18 in March and hour 19 in April to June.

Summer (June, July, and August)

Summer shows statistically significant warming at each hour across all months, with a notable peak in late afternoon (hours 18 — 20) and minimum in early morning (hours 7 — 9). This makes California's warming pattern largely different than that seen in most locations in North America during the summer, where one would expect the largest warming to occur overnight (USGCRP, 2017). The periods of greatest warming line up directly with the daily peak in residential demand for electricity and are concurrent with the period of the day in which solar production declines.

Fall (September, October, and November)

Fall features a pronounced diurnal shape in warming rates, with peak warming occurring in the late afternoon (hours 17 — 19), and the least warming (and sometimes no significant warming) occurring in the early morning hours. Peak warming rates occur before sunset, becoming progressively earlier from the summer into fall (hour 19 in August, hour 18 in September and hour 16 by November).

Across all seasons a trend emerges: peak warming occurring just before or at sunset and the least warming occurring just after sunrise. This is suggestive of a potential atmospheric response to daily heating; perhaps in some regions an enhancement of late afternoon sea breezes or marine layer intrusion in response to afternoon warming, which persist until sunrise.

Regional Trends

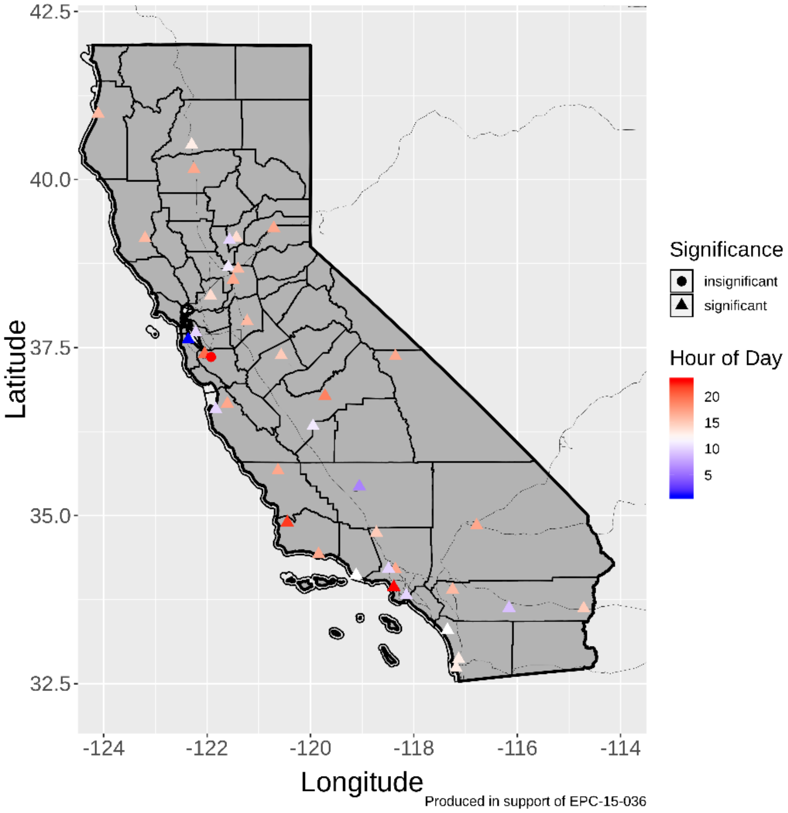
The previous discussion of statewide trends conveys the larger picture of when and how warming is occurring across the state; however, warming rates can vary locally and regionally from the statewide mean. Here, regions which conform to and differ from statewide trends are identified and individually discussed by season. Stakeholders looking at a regional view should reference station-by-station, month-by-month trend analysis and/or the linear trend model databases for a more comprehensive review of trends. Included here is a discussion of how to interpret the trend analysis, with suggestions for detecting changes due to instrumentation and siting, and the need to review station QA/QC data.

Winter (December, January, and February)

Two key regions differ from the statewide late afternoon peak (hours 16 — 17): the Bay Area and the coast from Santa Monica southward to the Mexican border. Both regions show a peak warming occurring sometime between midnight and 9am for parts of the winter season

(Figure 11 shows the hour of greatest trend in January). A closer look at these coastal stations' individual profiles shows peak warming in the afternoon and overnight, a dual peak that differs from the predominant statewide singular peak in early afternoon. That is to say, while locally the maximum warming rate occurs in the morning, a second, additional peak occurs in the late afternoon, in concert with the statewide profile.

Figure 11: Hour of Largest Warming Trend in January



Hour of greatest warming trend by station in January.

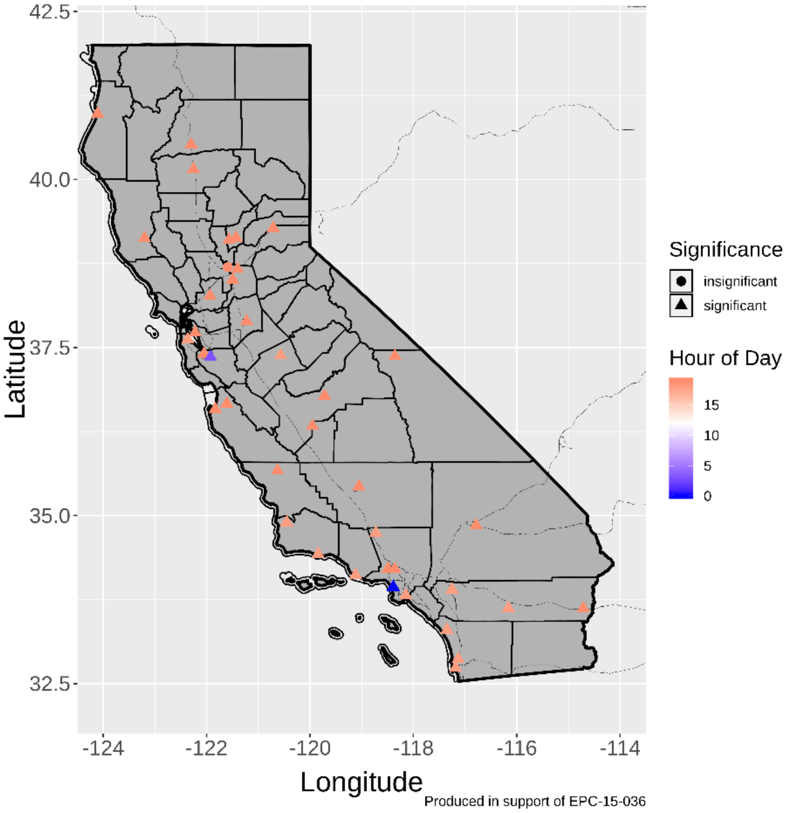
Source: Eagle Rock Analytics

Many of the same areas that showed differences from the statewide hour of largest warming also differ from the statewide hour of lowest rate of warming (or in some cases, cooling). Whereas statewide minimum warming was occurring near to sunrise, at coastally influenced regions like the Bay Area and Southern California, the temperature increases were smallest in the daytime minimum temperature — often near to noon. This is consistent with the aforementioned double peak. The North Coast region's trends are more in line with the statewide average, suggesting structural differences in thermal response along the North and South coasts.

Spring (March, April, and May)

March and April (Figure 12) are unique in that the hour of maximum increases in temperature are nearly uniform statewide, peaking in late afternoon or early evening. At coastal stations in Los Angeles, some stations peak near midnight, but the signal overall is still suggestive of the largest warming occurring in late afternoon into the early part of the night. Minimum warming occurs in late morning, again nearly uniformly across the state in March and April; in many regions, a statistically significant cooling occurs in the morning. In May, the statewide uniform response breaks down, and distinct patterns emerge similar to the summer months (discussed below).

Figure 12: Hour of Largest Warming Trend in April



Hour of greatest warming trend by station in April.

Source: Eagle Rock Analytics

Summer (June, July, and August)

In the summer, the Central Valley, Inland Empire–High Desert, and Salinas Valley–San Luis Obispo all behave similarly to the statewide mean of strongest warming trend occurring in the late afternoon and the smallest trend just after sunrise; while the North Coast, southern San Joaquin Valley, High Sierra and Los Angeles Basin feature peak warming between midnight and noon. Despite stark differences in climate, elevation, and ecological regimes, these locations typically feature a strong warming trend in all hours, except around midday. This manifests itself as two peaks, with nearly equivalent warming occurring in late afternoon, like much of the state. Stations can be roughly broken into two categories: those experiencing "two peaks" not showing significant midday warming, whereas the stations conforming to statewide averages do show significant midday warming. In the Central Valley, stations that are influenced by the Delta breeze conform to the statewide trend, but more isolated stations like Redding, Red Bluff, Hanford, and Bakersfield, which feel the breeze less frequently/intensely, exhibit significantly different warming patterns.

Fall (September, October, and November)

The shapes of diurnal trends change most dramatically through the fall season, with September behaving similarly to the summer period, October similarly to April, and November similarly to December. Overall, November exhibits the smallest trends, and at most stations is among the lowest hourly variability in trend (nearly equivalent throughout the day). The previously identified "double-peak" commonly seen in the Bay Area and coastal Southern California is seen in September but is less evident in October and November. Most stations in Northern California do not report significant warming trends in November.

Station-Based Trend Profiles

At every station, for every month, (1) visualizations of hourly trends (as pdf) and (2) the linear models with complete statistics (as R databases) are deliverables of this project. A comprehensive review of all the profiles is beyond the scope of this final report, but suggested approaches for users to interpret individual profiles are outlined below.

How to Use Station-Based Profiles

Hourly trends are generated at each of the 39 stations, for each month; the graphic profiles show the magnitude and p-value of the trend. The R database submitted with this project contains supporting statistics to help users contextualize the results. It is recommended that users looking at a highly localized level review the supporting QA/QC materials to ensure that their analysis is appropriate to the underlying data.

Users should consult adjacent stations before using these products for operational purposes. Microclimates, urbanization, and other regional conditions can affect individual stations; regional trends should consider adjacent stations. Figure 13 shows the hour of maximum temperature trend increase by station and month. In Sacramento, the trends are largely consistent, although some regional differences do occur (for example, Sacramento International Airport behaving differently than other regional stations). In the Bay Area, stations are largely consistent, with the exception of San Jose, where missing overnight observations due to airport closures in the early part of the record may affect the trend analysis. Given these missing observations, the composite of San Francisco, Mountain View, and Oakland may be more reflective of regional trends.

User-Generated Profiles

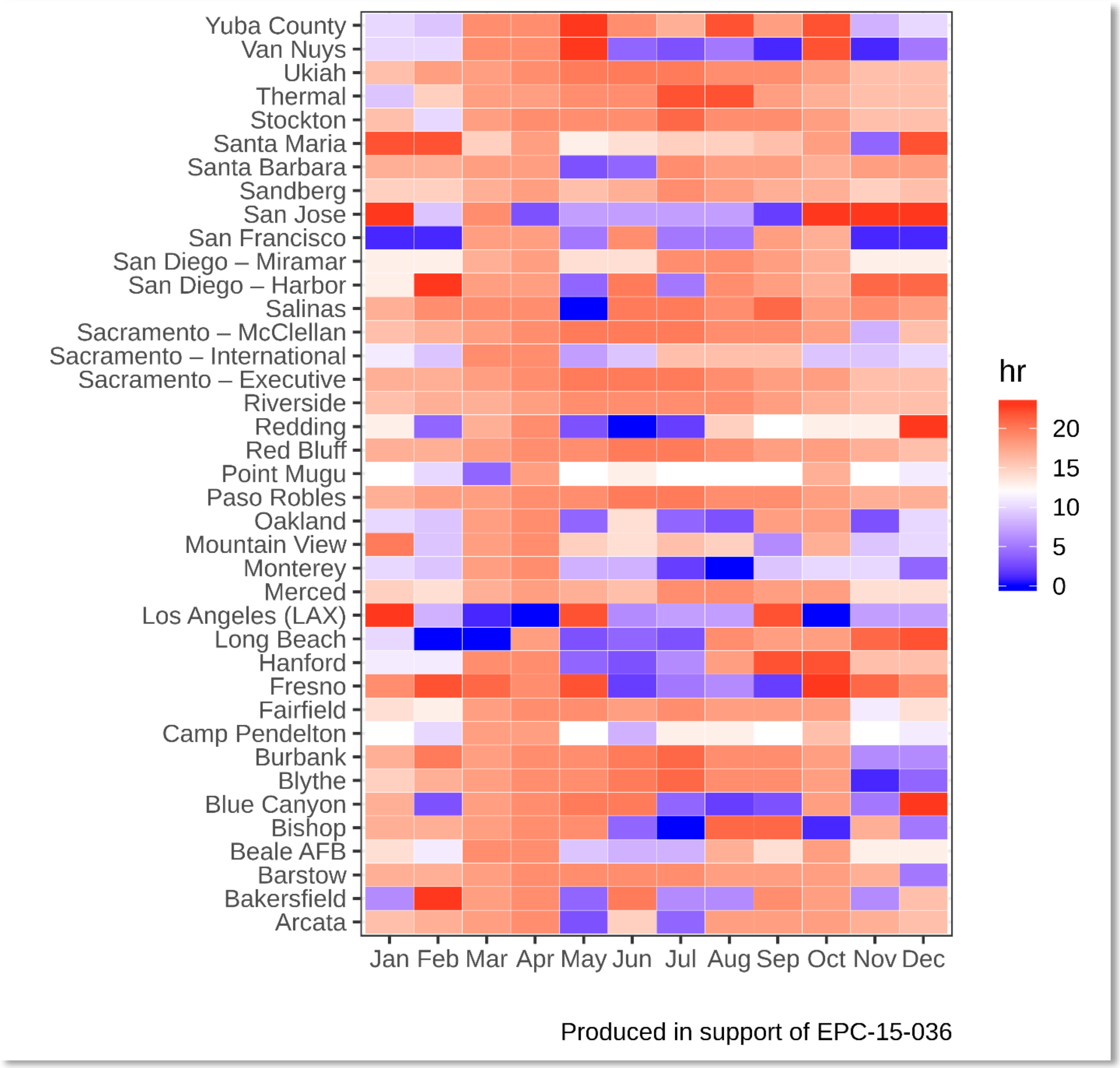
The approach used to detect changes in the diurnal shape of temperature at each station uses open-source code packages or repositories and is easily adapted to address multiple questions for the energy sector. Users are encouraged to take advantage of the products generated from this project to build on this work to address their own questions. The R database[3] containing trends includes the model structure and calls needed to replicate this analysis on different timescales, or to aggregate across differing spatial domains. Coupled with the NetCDF file of hourly observations, users can design their own trend analysis using three freely available R

---

[3] An "R database" is a digital storage mechanism for storing data objects, in this case a statistical model, for the R programming language. It is used here because stakeholders identified it as the preferred method for conveyance.

libraries, which complement the data processing structure of the database: lubridate, tidyr and dplyr.

Figure 13: Hour of Greatest Temperature Trend



Hour of greatest temperature trend by month and station.

Source: Eagle Rock Analytics

Considering Impacts of the Shift from Manual Observations to Automated Observations

Changes in instrumentation or shifts in the location of instrumentation can induce spurious trends or suppress evidence for environmental trends. The shift from manual observations to automated observations (augmented by human observers at some locations) was accompanied by a change in instrumentation, and potentially a shift in the location of the instrumentation. The HadISD dataset is tested for shifts through a pairwise-homogeneity test, comparing the record at a given location to its nearest neighbor, and data in which one station shows a spurious and sudden shift in mean state is rejected. However, this test is designed for a global analysis and may not be ideal for use in California where "nearby" stations may be in a completely different climate zone (for example, San Diego and El Centro, or Oakland and Stockton — each geographically close but nearly physically unrelated).

To address the impacts of potential jumps in the record due to station changes, the trend analysis is broken down into two periods of nearly equivalent length — the manual era consists of the period 1973 to 1996 and the automatic era from 1997 to the present. Users should ensure that the trends in each era are the same sign to confirm that overall trends are not affected by instrumentation or siting changes. The record, even after QA/QC was applied, is somewhat noisy so users are encouraged to look beyond p-values to establish significance of trend, consider trends in adjacent stations, and establish that trends are consistent over manual and automatic observation time periods. Users seeking to apply these results are encouraged to use bootstrapping and a nonparametric test (for example, Mann-Kendall) to ensure results are not a function of endpoints or bad assumptions about the shape of the underlying data.

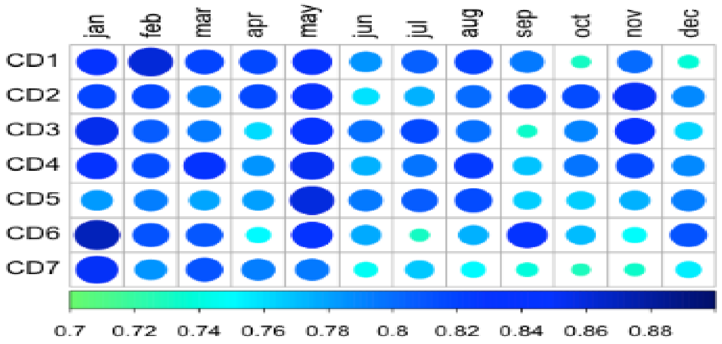
Seasonal Predictability for the Electrical Sector

The Potential for Predictability of Extreme Heat

California's electricity sector is vulnerable to widespread extreme heat events that strain the system's ability to provide reliable service (Burillo et al., 2018). Timely predictions of upcoming extreme heat events can help utilities to anticipate and plan for periods of extreme heat, such as the July 2018 Southern California heatwave that led to a peak power event and outage affecting more than 80,000 ratepayers (LADWP, 2018). The predictive system's ability to predict extreme heat events on seasonal timescales is assessed through what is known as a "perfect prog" analysis, or one in which the ocean condition is predicted perfectly and therefore can test the statistical tests' ability to predict, in this case, extreme heat.

Figure 14 shows a graphical summary of the ability of the scheme to predict extreme warm months (defined as months that were two standard deviations above the mean temperature) in each of California's climate divisions. Broadly, coastal regions were consistently better forecasted, in terms of number of extreme warm months successfully predicted, than interior regions across Northern (CD1 vs. CD2 & CD3), Central (CD4 vs. CD5) and Southern (CD6 vs. CD7) California, but even in the lowest performing regions (CD 7: Interior Southern California) the scheme predicted 70 percent of the extreme warm months.

Figure 14: Fraction of Extreme Warm Months Predicted



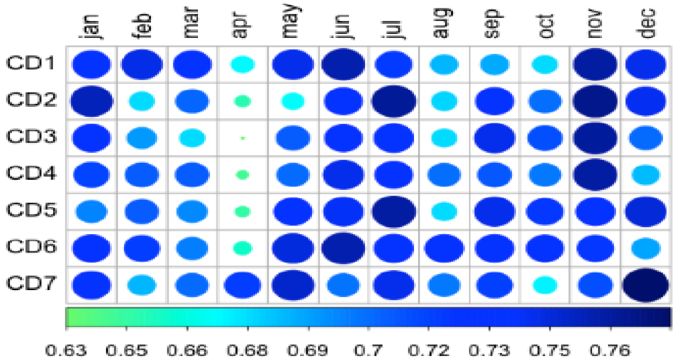
Fraction of extreme warm months (greater than 2 standard deviations above median) predicted in a perfect prog analysis of prediction scheme.

Source: Eagle Rock Analytics

Forecast skill is best understood within the context of how good is a given forecast compared to the challenge of making the forecast? Figure 15 shows the skill score of the schema compared to persistence or using the previous month’s conditions as a forecast. Skill score represents the improvement over a reference forecast (Wilks, 2011), as shown in Equation 1, where A represents the accuracy of the forecast, and *ref* refers to a reference forecast (in this case persistence) and *perf* refers to a perfect forecast.

$SS = (A – A_{ref} / A_{perf} – A_{ref}) * 100\% \ (1)$

Figure 15: Skill Score of Prediction of Extreme Warm Months



Skill score (%) of a perfect prog forecast of extreme warm months (greater than 2 standard deviations above median) relative to persistence.

Source: Eagle Rock Analytics

A forecast is perfect if the SS is 100 percent; if a forecast is only as good as its reference (persistence) it earns an SS of 0 percent; and negative values indicate the forecast is less skillful than its reference (persistence). The prediction scheme showed skill in predicting extreme warm months, with skill scores ranging from 63 to 70 percent across all regions and seasons. Skill scores were greatest across May–July and November–March. In comparison to the fraction of events well forecast, the skill score did not vary as a function of distance to the ocean, suggesting that the Pacific Ocean has significant capacity to predict extreme warm months in California.

Predictability of the Pacific Ocean

The ability of a LIM to predict ocean conditions is demonstrated by Dias (2019) and a comprehensive review is beyond the scope of this report. Here the implications for the potential for predictability are briefly discussed. Linear inverse modeling of the Pacific Ocean has insufficient skill to make temperature predictions in California accurate enough for use by utilities, but the strengths of the LIM are known and can be leveraged in a more complex predictive framework. LIMs have two key limitations: (1) LIMs struggle to perform when key regions of the ocean are neutral or lack temperature anomalies, and (2) LIMs are built upon smoothed and usually temporally aggregated training data, which can limit their ability to predict extreme conditions. Chapter 5 provides an explanation of how these limitations can be overcome with novel approaches, and through a hybrid approach for making seasonal temperature predictions for California.

CHAPTER 4: Technology/Knowledge/Market Transfer Activities

Transference of Hourly Weather Data to Demand Forecast Office

During the final meeting for this project, a plan was devised to transfer knowledge and data products to the CEC after consultation with CAM and DFO leadership. Three key data products are to be provided to the CEC: (1) hourly weather database, (2) typical day formulations, and (3) linear trend analysis models. Each of these databases contains spatially variant information for 39 locations and broadly contain information on the period 1973 to the present.

Online Repository for Data Access

All data products and supporting metadata will be hosted on a Microsoft One Drive server from which the data can be downloaded by DFO personnel to local servers. The data (~30 GB in size) will be held on One Drive for a period of six months (ending March 31, 2020) at the recipient's cost, with extensions provided by request. Data products (discussed below) to be stored in this manner include NetCDF (hourly and typical day databases) and R data structures for linear trend models.

Hourly Weather Data

This dataset consists of a curated record of hourly temperature observations at 39 locations across California, for the period 1973 to the present. The comprehensive review of data quality, HadISD QA/QC products, and metadata focused on temperature, and no claims of the quality of ancillary products delivered are made. The HadISD database does include additional weather variables such as dewpoint temperature, wind (speed, direction, and gustiness), atmospheric pressure, present weather, visibility, and cloud cover. It is formally advised that any and all additional weather variables delivered should not be used operationally without a proper assessment of data quality. Supplementary variables fail to meet the CPUC requirements for "demonstrated acceptance by the community-of-practice," through a lack of review and utilization in the field. The approach taken here with temperature represents a template that could be replicated on these variables. Additional weather variables may be of provisional use as a source for "back of the envelope" type calculations, so they are included in this data product.

NetCDF Databases

Two NetCDF databases are created as deliverables for this project; the primary database provides temperature and temperature anomalies and a secondary database with the complete unmodified HadISD records (with additional weather variables) at all stations. For both datasets, variables are stored with the dimensions of time (represented in units of hours since January 1, 1973, at midnight Pacific time) by station. Station metadata includes standard station informatics and identifying ID numbers for referencing standard atmospheric datasets (for example, WBAN or USAF Station ID numbers). Datasets are formatted and documented via metadata to CF-Compliant NetCDF and Cal-Adapt standards.

Typical Day Formulations

Typical day formulations are stored in NetCDFs, with variables representing the smoothed long-term mean temperatures and anomalies. The datafiles are structured to contain 366 days (numbered as Julian days from January 1 through December 31, inclusive of February 29) at 39 stations.

Linear Models for Hourly Trends

A comprehensive trend analysis was performed on the hourly temperature record across California. As part of this analysis, a wide suite of statistical measures (beyond p-values) were generated, and the analysis repeated across timescales. This is a large amount of data, which is challenging to summarize succinctly in a short-format report. The DFO has expressed interest in the models developed in this study for internal use. DFO staff use R in their analysis, and this model structure is stored in a way that makes it accessible and easily loaded for internal use.

Bringing Data Products to Cal-Adapt

Cal-Adapt typically works with regularly gridded datasets; most often produced from climate models. However, some point-based data (as in this product) is available from the hourly sea level rise tool and database, so Cal-Adapt possesses the technical ability to host these data.

Digital Transfer Options

This research team works closely with the Cal-Adapt production team at the Geospatial Innovation Facility at UC Berkeley and will work with CAM and Cal-Adapt personnel to ensure a smooth transfer of data to Cal-Adapt servers. Data will be hosted on a Microsoft One Drive server, from which the data can be downloaded by CEC and Cal-Adapt personnel to local servers. The data will be held on the One Drive for a period of six months (ending March 31, 2020) at recipient’s cost, with extensions provided by request. However, the technical capacity exists to amend this approach and use alternative data storage options as mandated by state or Cal-Adapt requirements.

Conforming to Cal-Adapt Standards

The hourly weather database and typical day formulations adhere to Cal-Adapt best practices for metadata and file conventions and are therefore ready for immediate transfer and dissemination via Cal-Adapt.

Knowledge Transfer at Cal-Adapt Users Needs Assessment Workshops

Here it is proposed that this work be presented at the 2019 Cal-Adapt User Needs Assessment Workshop to IOUs and other energy sector stakeholder groups, to ensure dissemination of key results, promote access to data products, and discuss how the data products may be used to support energy sector planning. The User Needs Assessment Workshop can be used to better understand formatting needs for the data to ensure that the information is accessible by IOUs and others. Such a presentation may serve as a jumping-off point to enable CEC staff to better understand which needs for hourly data are not currently met.

CHAPTER 5: Conclusions and Future Directions

Here the authors discuss the next steps needed to progress to a more granular, more temporal, more data-intensive approach to meet the needs for higher resolution weather data, as articulated by CEC Commissioner Andrew McAllister. This section provides concrete steps, background information, and context to allow the CEC to improve quality and access to hourly weather data, leverage hourly data to create additional products requested by utilities and to make seasonal predictions operational. The future directions include:

1. Expanding the current hourly weather database (>46 years of hourly observations at 39 locations in California) to cover regions of California currently underrepresented by:
    - Incorporating additional sources of weather data, identified below.
    - Relaxing the requirements for including stations in the record.
2. Improving the scope (higher spatial resolution on subdaily timescales) and the quality of the temperature record, as clamored for by utilities and energy sector stakeholders to meet climate-related goals. Recommendations to improve the temperature record include:
    - Offering varying levels of data service (that is, from raw observations through smoothed regular products) to better match stakeholder needs.
    - Using diurnal characteristics from this work to extend gridded daily temperature products to hourly frequency.
    - Understanding arguments for and against the State of California developing localized temperature records.
3. Leveraging California’s great opportunity to improve the understanding of temperature variability across the state by merging observations that are presently disparate (across various state and local agencies, as well as utilities) into a functional, uniform mesonet. The CEC can continue its historical role at the state level by providing leadership on climate through:
    - Centrally collecting, quality controlling, and assimilating weather observations across the state into a uniform data product; disseminating this information in near-real time on Cal-Adapt and providing hyperlocal weather information to utilities to improve reliability in the face of extreme weather events.
    - Using information from this collection to build a high-resolution gridded product that would meet the spatial and temporal needs requested by energy sector stakeholders.
    - Merging these disparate networks into a single network would be of great value to the electricity sector, in the protection of life and property during extreme weather events, and to weather forecasting and helping to achieve other state climate goals.

4. Developing a seasonal prediction scheme for air temperature in California, using a hybrid approach, including information from Linear Inverse Models and seasonal climate modeling efforts.

Improving Hourly Weather Data for California

The curated database produced through this project is a good starting point for providing the energy sector with a standardized high-quality product, with fair spatial coverage. In the following section, guidance is provided on how this product could be improved and augmented in the future, as well as how this product can support and inform CEC research projects. An opportunity exists for California to develop a state-of-the-art data repository for weather data relevant to the electricity sector, which could lead to the development of cutting-edge, gridded hourly temperature products that electricity sector stakeholders are requesting and could be used as a basis for ultra-high-resolution climate model downscaling efforts.

Expanding the Coverage of Hourly Stations

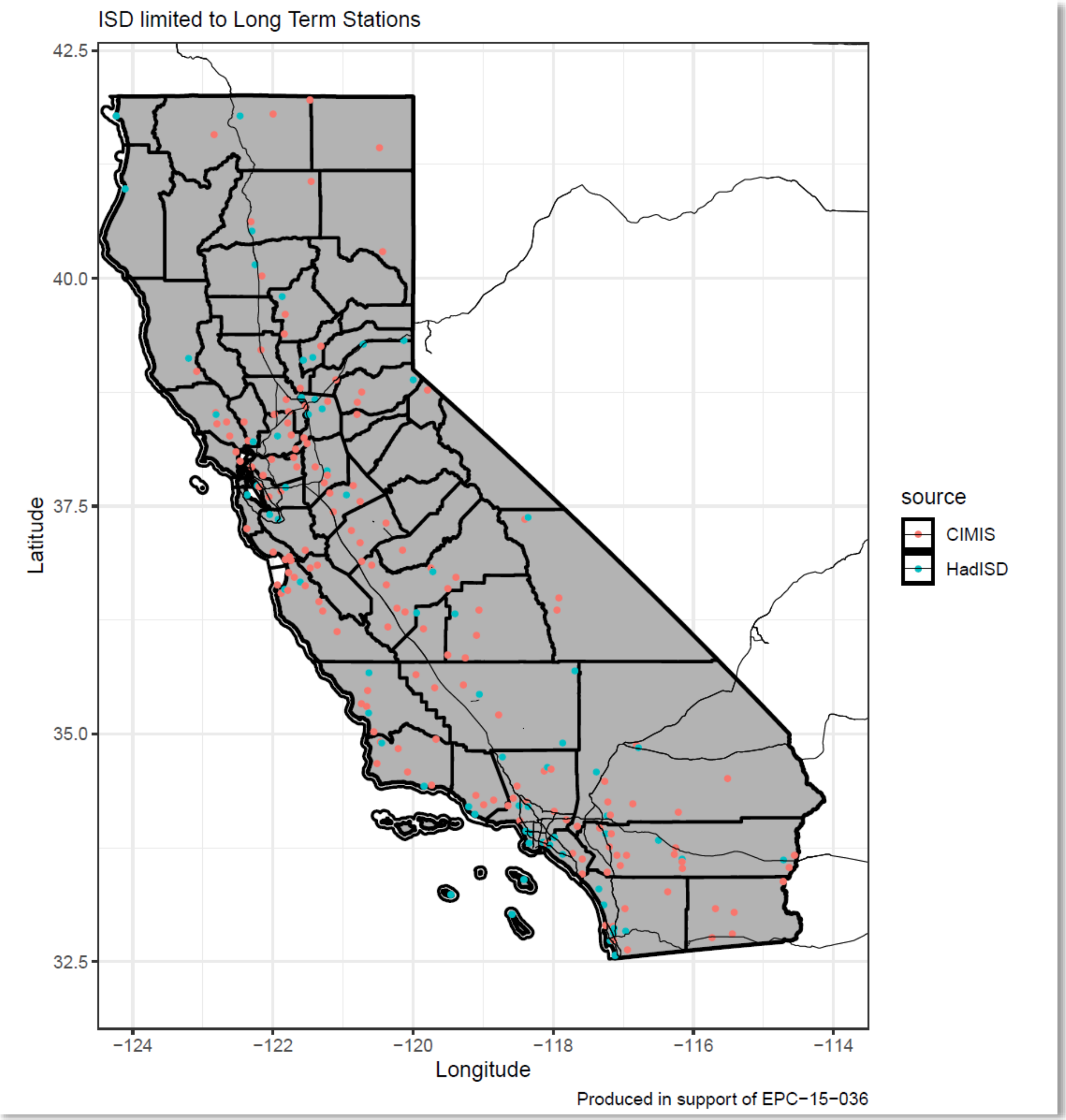
Incorporating Additional Data Sources

Other Sources of Subdaily Weather Information

The HadISD repository, used as the primary source for hourly weather data in this study, is preferred as it is likely of the highest quality; however, additional sources of hourly weather data do exist. Two major challenges arise with respect to subdaily weather data: (1) weather stations often do not stay on line for long periods of time and (2) the data collected at supplementary sites are of lower quality because instrumentation maintenance and QA/QC protocols tend to be less robust than at airports. However, as the service areas of smaller IOUs and POUs tend to be insufficiently represented by the HadISD, two potential sources of subdaily weather data are considered below.

The California Irrigation Management Information System (CIMIS) archives hourly weather observations dating back to 1982. The focus of CIMIS is on supporting agriculture and it offers a limited number of weather variables (wind speed, wind direction, humidity, temperature). Stations tend to have shorter, less stable records in comparison with airports, as these agricultural monitoring stations drop in and out of service more frequently. Further, agricultural stations are maintained less rigorously, increasing the probability of instrumentation bias and drift. Despite these limitations, incorporating CIMIS into the hourly weather dataset would be of great value because the stations are located in rural regions that are underrepresented in the current record (see Figure 16).

Figure 16: Station Locations



Location of active CIMIS in California (red) which could be used to augment the current network of stations (blue) used in this analysis. Note the expanded coverage of CIMIS stations into less densely populated and underrepresented regions of California.

Source: Eagle Rock Analytics

Beginning around the turn of the century, the development of cheap weather sensors and proliferation of wi-fi and cellular data personal and institutional weather stations led to widespread deployment of "backyard" weather sensors. From amateur weather enthusiasts to weather monitoring devices deployed by state agencies in critical locations (for example, Caltrans–Road Network Information System and the California Air Resources Board), the fraction of the state covered with subdaily observations of temperature has greatly increased. New QA/QC protocols have been developed that can process data records of varying quality and produce spatially consistent products (de Vos, 2019). Of particular interest and value are the dense and well-maintained mesonets being deployed by IOUs, with a focus on wildfire-vulnerable areas. While such data sources are presently of too short a record to produce a climatologically suitable long-term product, a uniform quality-controlled stable product could be produced on more recent time periods. Merging these disparate networks into a single network would be of great value to the electricity sector, in the protection of life and property in extreme weather events, and to weather forecasting and helping achieve other state climate goals.

Sub-Hourly Weather Data

This work uses hourly weather reports, usually transmitted between 45 minutes to 59 minutes into the hour. These reports are, in theory, transmitted in fully automatic mode (not reviewed by a person) or in manual mode in which either a weather observer (at major airports) or air traffic controllers review the information before sending. In practice, air traffic controllers are too busy or too poorly trained to properly augment or correct the transmissions. However, fully automated observations are taken and archived every two minutes. These data are lower quality, and more likely to be missing, but may be of use in future energy sector studies (for example, big data studies using high temporal resolution, individual consumer meter information). It is possible to produce a dataset using this ultra-high temporal resolution data, if needs warrant it.

A More Tolerant Data Record

This data record was curated in response to feedback from TAC members and CEC staff suggesting that subdaily information was critically needed, and no reliable standard database was available for use, whether through Cal-Adapt or other internet locations). Proposed uses focused on the need for basic climatological information about diurnal cycles across California (that is, typical day formulations), changes in the shapes of diurnal cycles in response to a warming climate (that is, trend analysis) and the potential predictability of hours in a given month above or below threshold values deemed critical for operations (for example, SMUD identified the number of nights with daily minimum temperatures above 70°F as more operationally relevant than a prediction of monthly mean temperature). For each of these applications, a long-term, stable, homogeneous record is important.

Other energy sector applications may have less stringent requirements and lowering the thresholds would increase the number of stations available. For example, the DFO mentioned that they consider the period 1989 onward in most applications. The database could be expanded to include stations that came online between 1973 and 1989. Other applications may look only at high temperatures; thus, missing overnight records (that is, the complete daily cycle test) may not be a hindrance and so stations currently excluded due to insufficient nighttime records could be added, and as previously noted, these stations tend to be located at regional airports in underrepresented regions of California.

Modernizing California’s Weather Data Collection and Access

Statewide, significant effort is going into producing weather data, but no centralized mechanism or approach presently exists to collect, process, and systematically analyze the data. Some opportunities and approaches to make subdaily weather data accessible to the energy sector are outlined below.

Multiple Data Processing Levels as a Service

The needs of the energy sector are wide and varied, and no single product meets the needs of all potential studies. For example, modeling approaches often struggle with data products containing large gaps (sometimes requiring non-ideal assumptions to be made), whereas some engineering applications looking at hyperlocal, granular data or trend analysis prefer raw observations with no infill of data. To address this, an ideal temperature product would feature multiple data processing levels (similar to those used at NASA), as shown below.

- Level 0: Unprocessed Instrumentation Record.
- Level 1: QA/QC'ed Data Record, with information about removed values and corresponding tests.
- Level 2: Data record with temporal gaps filled in through geophysical or statistical modeling.
- Level 3: Complete 3-D record (latitude, longitude, time) with underrepresented spatial regions informed through geophysical and/or statistical modeling.

By leveraging observed diurnal characteristics at each location for each day, sparse observations (such as a daily high or low temperature at a nearby daily temperature station, of which there are many more available) can be taken and used to recreate missing hourly temperature profiles. This can be done using traditional statistics (such as quantile mapping, or analogues) or through machine-learning techniques. This would allow us to recreate periods in which airport observations were not taken (such as nighttime periods at regional airports) and periods in which observations were removed due to QA/QC or instrumentation failures. Further, such an approach would allow us to significantly expand the spatial coverage of the dataset by including many more stations in which nighttime or seasonal depletions in the record are evident.

Toward a High-Resolution Gridded Product (Level 3)

Energy sector stakeholders have expressed interest in a standardized, ultra-high-resolution (that is, approximately 2 km) gridded, high temporal resolution (ideally hourly). Such data are necessary to complement the "big data" revolution in electricity and natural gas, combining with such sources as per-meter-user utilization rates and highly advanced building-level modeling work. Further, in order to downscale coarse output from global climate models to resolve the complexity of California's diverse microclimate zones, observations at a high spatial resolution are needed as inputs to downscaling approaches. High resolution, ideally improving on the current resolution of approximately 7 km, is a critical component of California's future statewide Climate Assessments. Presently, IOU technical staff and other stakeholders are left to guess where to find hourly weather data and may or may not have the technological and subject-level expertise to understand the limits and quality issues of the data.

The challenge is, therefore, to produce a continuous (in space and time), regularly spaced dataset from irregularly spaced point observations which vary in record length and quality. Traditional approaches to solving this problem include interpolation (weighted — inverse distance weighting or kriging, and unweighted — spline or regression) and modeling (such as adjoint, reanalysis, or fluid dynamical models). Hybrid approaches use some sort of interpolation, limited or informed by physical characteristics (like elevation, slope orientation, distance from ocean, or other bio-geophysical parameters). Recent advances in machine learning (for example, random forest as in Hashimoto [2019]) suggest that it may represent a path forward for increasing the spatial and temporal resolution of temperature products.

Diurnal characteristics from this dataset could be combined with daily reporting stations (that is, instruments that record minimum and maximum daily temperature) to produce estimates of hourly conditions. Daily reporting sites are significantly more common than hourly reporting sites, and many of these stations have extensive (50+ years) stable records.

Similarly, remote sensing snapshots of surface temperature, available infrequently due to polar orbits of satellites or frequent cloud cover, can be merged with known diurnal characteristics to create a high-resolution gridded product, with reasonable estimates for hours for which no remote sensing products are available. The current 39-station record may be insufficient to represent diurnal characteristics across the entire state, but inclusion of CIMIS or other observational sites could make this approach feasible and address a long-standing need of the energy sector.

The Case for a California-Focused Temperature Product

California's unique and complex topography and preponderance of microclimates represent an often insurmountable challenge for globally designed products to accurately represent. As seen in Figure 17 and Table 5, the globally developed QA/QC protocols can misdiagnose QA/QC conditions in coastal and semi-arid regions. Failure rates are high in areas frequently affected by marine airmasses and low in interior regions, despite these stations receiving similar maintenance. Many of these coastal stations have full-time weather observers (for example, Los Angeles, San Diego, Santa Ana, Oakland, San Francisco, San Jose) checking each observation for mistakes before transmission. It is likely that many of these "failures" are actually accurate measurements of California's marine-influenced nighttime weather (steady invariant temperature) that are erroneously flagged as "streaks" and removed. A product designed specifically for California could make better assumptions and estimates appropriate for California's diverse environments.

Figure 17: Failure Rates by Station

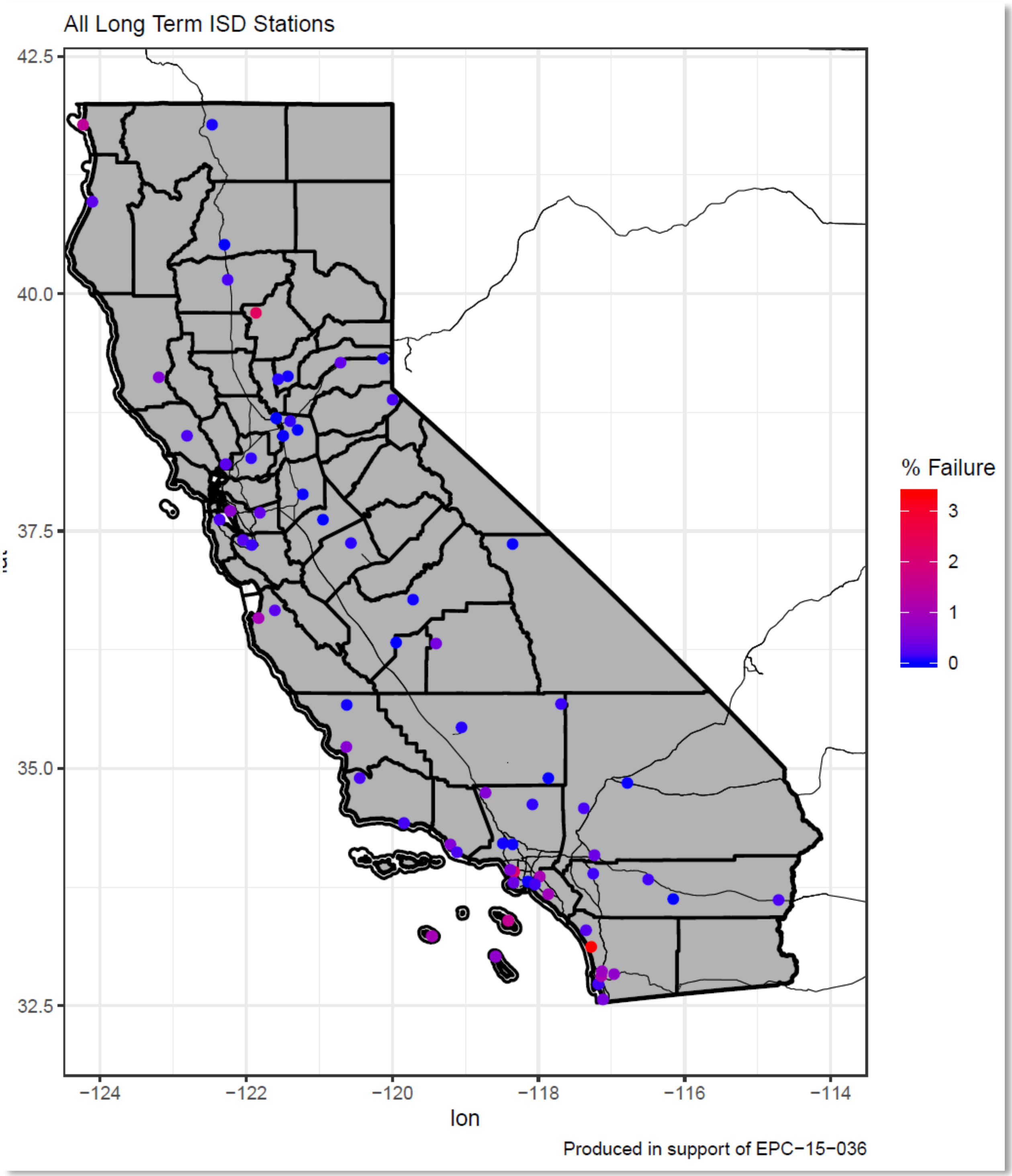


Figure 17: Rate of frequency of observations that fail QA/QC tests. Failure rates are high in marine-influenced locations and low in interior arid and semi-arid regions.

Source: Eagle Rock Analytics

Many states have mesonets, dense networks of weather stations continuously producing near-real-time weather data. Even economically limited states like North Dakota, Nebraska, and Oklahoma have mesonets, some dating back to the early 1980s. Despite being a leader on many climate issues, and suffering a disproportionate amount from climate change, California does not have a functional mesonet. Mesonets are excellent sources of high-resolution, high-frequency (sub-hourly) weather data and are natural components of the sort of high-resolution gridded products needed by energy sector stakeholders. With IOUs deploying extensive instrumentation, disparate state, county, and municipal agencies collecting information, and private citizens collecting weather data, it could be worth exploring a state-led effort to collect, analyze and manage these data with the aim of producing a long-term, ultra-high-resolution temperature (and other variables) product. IOUs would benefit from having an external QA/QC protocol applied to their mesonets, and ratepayers would benefit from associated improvements to grid reliability. Further, POUs and smaller utilities who lack the resources to deploy extensive mesonets and the technical expertise to use such observations would benefit most from the centralized collection of these data. Cal-Adapt would be a natural home for such a real-time monitoring network.

The Case Against California Going It Alone

The United States government (and other national governments) spends millions of dollars each year producing surface temperature datasets, so it is reasonable to think that California would be better off taking advantage of external efforts. Such products are typically frequently updated, and often incorporate new techniques and data over time, leading to higher quality products. Some examples of products to consider include:

- PRISM (developed at Oregon State University, supported by USDA) has daily maximum and minimum temperatures at a resolution of less than a kilometer. The modeled product uses such parameters as distance from the ocean and elevation, variables critical to represent California microclimates. However, the product is not suitable for trend analysis because the amount of data that goes into the observations fluctuates with time, leading to an unstable (non-homogenized) time series. Available at: http://www.prism.oregonstate.edu/
- Daymet (developed by Oak Ridge National Laboratory, supported by NASA) is a pseudo-modeled, pseudo-extrapolated set of algorithms that take daily observed station-based minimum and maximum temperature observations and extrapolate to regions where no observations are available. Daymet iterates over 2-degree by 2-degree regions (~220 km by ~220 km), applying various statistical tests to weigh and ultimately interpolate surface observations over a subregion, an appropriate approach for regions where climates change dramatically over relatively small areas. Available at: https://daymet.ornl.gov/
- Livneh is the currently used temperature product for the CEC, featured on Cal-Adapt and used in the downscaling techniques employed for California’s Fourth Climate Change Assessment. It produces maximum and minimum temperatures on a very high, one-sixteenth-degree resolution grid (7 km x 7 km), but may struggle to resolve conditions in urban areas and coastal areas, or during periods of extreme heat, requiring significant preprocessing before use. In this study, use of the Livneh data (Livneh et al., 2015, via Cal-Adapt.org) was attempted make seasonal temperature projections and the data were found to be too noisy to be of use in that application.
- WorldClim produces near 1 km estimates of maximum and minimum temperature, primarily to support biological modeling. It is only produced on monthly timescales and

lacks a publication and peer-review of methods for utilization at this time. Available at: https://www.worldclim.org/.

Reanalysis products, which synthesize observations by passing them into fluid dynamic climate or weather models and then processing (via assimilation or modeling) the input observations into a gridded product, represent another potential source of hourly weather information. While such products are not observations, they do have superior spatial products and typically long-term records. There are many reanalysis products, available at various timescales and spatial resolutions, each of which has its own strengths and weaknesses. Here, two of the highest-resolution reanalysis products available are noted: (1) initializations of the North American Mesoscale Forecast System (NAM), available four times per day (at six-hour intervals) at 12 km resolution, are often used like a reanalysis product and available from 2004–present; and (2) ERA5 has hourly values at a resolution of one-quarter degree, covering the period 1979 to the present. Reanalysis products may not be suitable for engineering purposes but could represent a "first guess" product for an iterative approach to generate a Level 3 product.

Subdaily Weather Data on Cal-Adapt

The NetCDF file containing hourly weather data is designed to Cal-Adapt specifications and metadata requirements, to facilitate easy transfer to Cal-Adapt. Below are some suggestions on how these data could be used on Cal-Adapt.

Dissemination of Hourly Data

The hourly data files are structured to be easily parsed for dissemination via the Cal-Adapt download tool and/or API. There is no guidance available for data producers for Cal-Adapt to reference in generating files, so CF-compliant naming and metadata practices are used herein to ensure that files are easy to handle by the Cal-Adapt team. It is recommended that Cal-Adapt carefully consider how best to name the two anomaly products (based on daily means and hour of day in year means, respectively). To avoid confusion, daily anomalies and hourly anomalies are recommended for use.

Incorporation into Existing Products

Incorporating the hourly weather dataset directly into existing Cal-Adapt tools is not recommended. The limited spatial coverage of the products makes this a challenging proposition, and although having "real observations" to reference during the historical time period would provide utility, the effort needed to code this is likely too high to justify the production cost.

New Tool Development

A standalone tool highlighting changes in the structure of the daily cycle and presenting "typical day" formulations or showing linear trends may be of interest to energy sector stakeholders, and represent a direct path for users to download the data. Recommendations include (1) presenting the results as an average across stations regionally, where sufficient coverage exists, to reduce single station noise; (2) passing along comprehensive statistical measures of significance (beyond a simple p-value) to end users; and (3) presenting the trend results on monthly timescales and typical day formulations on daily timescales.

Developing a new tool around the hourly data repository is not recommended; however, these data should be made available to end users via the API and download tools functionality.

Automation

This work identifies 39 stable, long-term stations that are most appropriate for representing subdaily weather conditions across California. These stations are taken from data updated monthly by the Hadley Centre in the United Kingdom, representing an opportunity for a near real-time continuously updated product. This is desirable because, for example, DFO staff annually update products to incorporate recent year data. There are two paths forward to approach this: (1) scripts that scrape the FTP site to download the new month's data and append it to the current NetCDF dataset, or (2) redownload and process the entire record each month. The database is structured so that time is the unlimited variable, and new data can be appended to NetCDF via CDO command line tools through a simple shell script or similar code. The choice of which approach to use might depend on the backend structure of Cal-Adapt.

While the upsides to a regularly updated product are clear, a dynamic dataset may lead to complications in practice. Infrequently, every couple of years, the QA/QC protocols applied in HadISD are updated, leading to a new version of the product with the potential for differences in the temperature record emerging. One can imagine the worst-case situation in which different stakeholders are using different versions of the product and coming up with different answers. For this reason, if a plan is implemented for automation, Cal-Adapt would also need to enact a plan for version control and archiving of prior versions. Metadata should be automatically appended/included, with each download or API instance tagging the temperature data with the current version. Specifically, attaching the metadata in a passive manner that users receive automatically, without prompting, is recommended. Further, each version of the dataset produced should be archived and made accessible, so users can make comparisons between prior versions.

Beyond Temperature: Additional Hourly Information for the Energy Sector

Weather data transmitted on an hourly basis have changed extensively during the transition from manual to automated observations. Key variables such as temperature, dewpoint, atmospheric pressure, wind (speed, direction, and gustiness) and present weather have been less affected by the transition, and a long-term record of these variables from this source is possible. In this work, the temperature record is carefully reviewed and 39 stations are identified that are appropriate for long-term analysis. No other supplementary weather variables were examined for quality as part of this study. Based on the instrumentation and QA/QC protocols employed, it is likely that dew point and atmospheric pressure are of similar quality to temperature. All other variables stored and presented here should be used with caution, and only after appropriate QA/QC.

Incorporating Changes in Diurnal Characteristics into Energy Sector Planning

This work has shown that warming is asymmetric in California, with a majority of stations showing greatest warming in late afternoon, and the least warming in early morning. These results vary by location and season, with coastal and inland regions showing differences in the timing of maximum warming rates. Broadly, these results are inconsistent with the widely accepted theory that nighttime lows are increasing more quickly than daytime highs, and studies in California suggesting the same. This work focused on mean conditions and trends; an hour-by-hour analysis of trends in extreme events may yield different results.

Maximum warming in the late afternoon during peak heat months has implications for the electricity sector. As more renewable sources of electricity come online that are diurnally variant, asymmetric warming should be considered in resource planning activities. Lower, and in some cases negative, warming rates in early morning at some locations have implications for the natural gas system, potentially offsetting anticipated drops in home heating demand due to climate change in the early morning hours.

Making Seasonal Predictions Operational for California Ratepayers

Combining LIMs with Other Forecasts

LIM predictions of ocean conditions provide useful information and predictability in certain conditions. The ocean conditions associated with high levels of predictability for a LIM (that is, periods of strong anomalies), and conversely the ocean conditions for which predictability is low (that is, periods of weak anomalies), are known. Predictability could be enhanced through a hybrid approach, in which ocean forcing informs temperature predictions when the conditions are apt, and predictions from other sources (for example, modeling or statistical) as conditions are less favorable. In practice, this would mean using a LIM-based forecast during periods when large structural anomalies exist in the Pacific Ocean and alternative forecast products when such structural anomalies are weak or nonexistent.

Because data that train LIMs are usually temporally aggregated and sometimes smoothed, LIMs can struggle to generate large anomalies that lead to extreme heat or cold in California. LIMs can be made to produce larger anomalies by forcing them with an ensemble of observations with increased noise and treating the output as a probabilistic prediction rather than a deterministic forecast. Such probabilistic information would propagate through the statistical models to generate air temperature, increasing utility of the forecasts.

Forecasts Beyond the Mean

In this work, the ocean has been shown to have significant potential for predictions of periods of high heat in California, which are critically important to the energy sector through increases in electricity demand and elevated fire risk. Presently, modeling efforts have been predicated on traditional assumptions that predicting the monthly mean temperature is important, but perhaps this is not the question that needs answering. A forecast model designed to answer more specific questions such as "what is the likelihood of a month with extreme heat/cold/precipitation" will be more successful, and of greater utility to energy sector stakeholders. Such a framework could rely on a single LIM run, with branches of statistical models taking ocean state predictions and making predictions on different variables as a function of location (for example, number of hours above a low-temperature threshold).

Who Is the Forecast Written For?

Conversations with IOUs suggest a lukewarm interest in CEC–supported seasonal prediction framework. At the first TAC meetings, IOUs described current practices, which ranged from informal ("we use some teleconnection indices") to formal (that is, relying on commercially available forecasts). No IOU expressed interest in collaborating on a seasonal product or vocalized a need for making one operational. The DFO ostensibly would have interest in such a product, but at present does not consider seasonal-scale predictability. In contrast, POUs seemed more receptive and hosted follow-up meetings to discuss project results and the potential for collaborations. Ratepayers in California would benefit from a standardized CEC seasonal prediction, particularly those in POU or smaller utility domains where there is no weather or climate prediction team. Moreover, having a standardized forecast would give IOUs a metric to report on to shareholders or regulators — for example, what was their skill score, false alarm rate, or bias — compared to the CEC forecast. Such a target forecast could still have value if the predictions were only marginally better than a reference forecast such as persistence or climatology.

CHAPTER 6:
Benefits to Ratepayers

Project Benefits to California Ratepayers

California’s energy systems are incorporating and will incorporate increasing contributions from renewable sources of energy, such as solar and wind. Concurrently, California’s peak demand for electricity during extreme heat events will increase as the state continues to warm at an alarming rate. Historically, IOUs and DFO personnel have used daily climate information (that is, daily minimum and maximum temperatures) to assess and plan for use response to weather. With growth in the use of renewables, and the rising availability of per-user subdaily user consumption data, subdaily weather information is increasingly needed. This work provides a stable quality-controlled record curated for California’s energy sector. Providing this information to energy system stakeholders through Cal-Adapt will enable a more stable energy system for California ratepayers, by allowing utilities to better understand how diurnal weather variability and electricity use are related. Improved weather data reporting has the potential to allow for reduced cost of electricity if utilities can leverage information provided therein to improve supply and electricity purchase decisions. Further benefits to ratepayers are likely to emerge as this information can significantly better inform infrastructure developments compared to historical data. Last, this work will reduce costs to ratepayers by providing a centralized repository of subdaily weather data, reducing utilities’ expenses in acquiring expert guidance to produce their own records.

California is warming asymmetrically, with different warming rates across the day as a function of location and month in year. *When* during the day and *how* temperature is changing at a location is becoming more important as utilities seek to implement peak demand shifting. This information will help utilities anticipate and plan for near-term future changes in demand, leading to a more stable electrical system at a lower cost. This information is likely to be of greater benefit to POUs and smaller utilities, where staffing with expertise in climate and weather is limited.

Improvements in energy use projections, through improved weather and climate forecasts, will increase electricity stability and reduce rates paid by ratepayers by allowing utility companies to make informed decisions about infrastructure and market purchases. Electricity consumption in California peaks during extreme heat events during the warm season. Accurate predictions of temperatures will ensure sufficient electricity is available to meet peak demands. As California moves toward more diverse sources of energy in the future, including a push toward renewables such as solar energy (which depends on cloud cover), wind power (which depends on wind speed), and hydropower resources (which depend on precipitation intensity, timing, frequency, and type — snow vs. rain), accurate weather forecasts are going to become increasingly necessary for the CEC and energy suppliers to accurately predict energy consumption and generation. Through more accurate seasonal and decadal predictions, utility companies can ensure adequate zero-carbon electricity supplies are available to ratepayers.

LIST OF ACRONYMS

Term	Definition
CAM	Commission Agreement Manager
CIMIS	California Irrigation Management Information System
DFO	Demand Forecast Office
DJF	Winter Season (December, January, February)
EOF	empirical orthogonal function
ERSST	extended reconstructed sea surface temperature
IOU	investor-owned utility
ISD	integrated surface database
JJA	summer season (June, July, August)
LIM	linear inverse model
LOESS	locally-estimated scatterplot smoothing
MAM	spring season (March, April, May)
POU	publicly-owned utility
SON	fall season (September, October, November)
SST	sea surface temperature
SVD	singular value decomposition
TAC	technical advisory committee

REFERENCES

Burillo, D., M. Chester, S. Pincetl, E. Fournier, K. Reich, A. Hall (2018). University of California Los Angeles. Climate Change in Los Angeles County: Grid Vulnerability to Extreme Heat. California's Fourth Climate Change Assessment, CEC. Publication number: CCCA4-CEC-2018-013

Dias, D.F., Subramanian, A., Zanna, L. et al. *Clim. Dyn.* (2019). 52: 3183. https://doi.org/10.1007/s00382-018-4323-z

Dunn, R.J.H. (2019). HadISD version 3: monthly updates, Hadley Centre Technical Note

Dunn, R.J.H. et al. (2016). Expanding HadISD: quality-controlled, subdaily station data from 1931, Geoscientific Instrumentation, Methods and Data Systems, 5, 473-491

Dunn, R.J.H. et al. (2014). Pairwise homogeneity assessment of HadISD, Climate of the Past, 10, 1501-1522

Dunn, R.J.H., K.M. Willett, P.W. Thorne, E.V. Woolley, I. Durre, A. Dai, D.E. Parker, and R.S. Vose, 2012. HadISD: a quality controlled global synoptic report database for selected variables at long-term stations from 1973-2010. *Clim. Past*, **8**, 1649-1679 (DOI: 10.5194/cp-8-1649-2012).

Hashimoto, H., W. Wang, F.S. Melton et al. High-resolution mapping of daily climate variables by aggregating multiple spatial data sets with the random forest algorithm over the conterminous United States. *Int J Climatol*. 2019; 39: 2964– 2983. https://doi.org/10.1002/joc.5995

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

LADWP, 2018. July 2018 Heat Storm Outage Event Summary. https://www.ladwpnews.com/weekend-of-july-6-2018-heat-storm-related-power-outages-and-response/

Livneh B. 2015. National Centers for Environmental Information. doi:10.7289/V5X34VF6

Smith, A., N. Lott, and R. Vose, 2011. The Integrated Surface Database: Recent Developments and Partnerships. *Bull. Amer. Meteor. Soc.,* **92**, 704–708, https://doi.org/10.1175/2011BAMS3015.1.

Wilks, D.S., 2011.. *Statistical methods in the atmospheric sciences* (3rd ed.). Oxford ; Waltham, MA: Academic Press.

USGCRP, 2017. Climate Science Special Report: Fourth National Climate Assessment, Volume I [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 470 pp., doi: 10.7930/J0J964J6.

Vos, L. W., Leijnse, H., Overeem, A., & Uijlenhoet, R., 2019. Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. *Geophysical Research Letters*, 46, 8820– 8829. https://doi.org/10.1029/2019GL083731