Explanatory Guide to Hourly Observed Historical Data on Cal-Adapt

This guide is an introduction to the *Hourly Observed Historical Data* product on Cal-Adapt. It seeks to inform users as to what the data is, what applications it is appropriate for and references to supporting documentation for further information.

Table of Contents

What is the Dataset?

The *Hourly Observed Historical Data* product consists of 39 stations across the state (Figure 1), each with an observation period of greater than 30 years (1973 to present) from the [HadISD global record]. Stations identified for use in this data product were chosen based on being considered high quality for temperature. Due to observing techniques, instrumentation used, and similarities in QA/QC protocols, it's likely that data for dew point and mean sea level pressure will be of similar quality, however this has not been assessed fully.

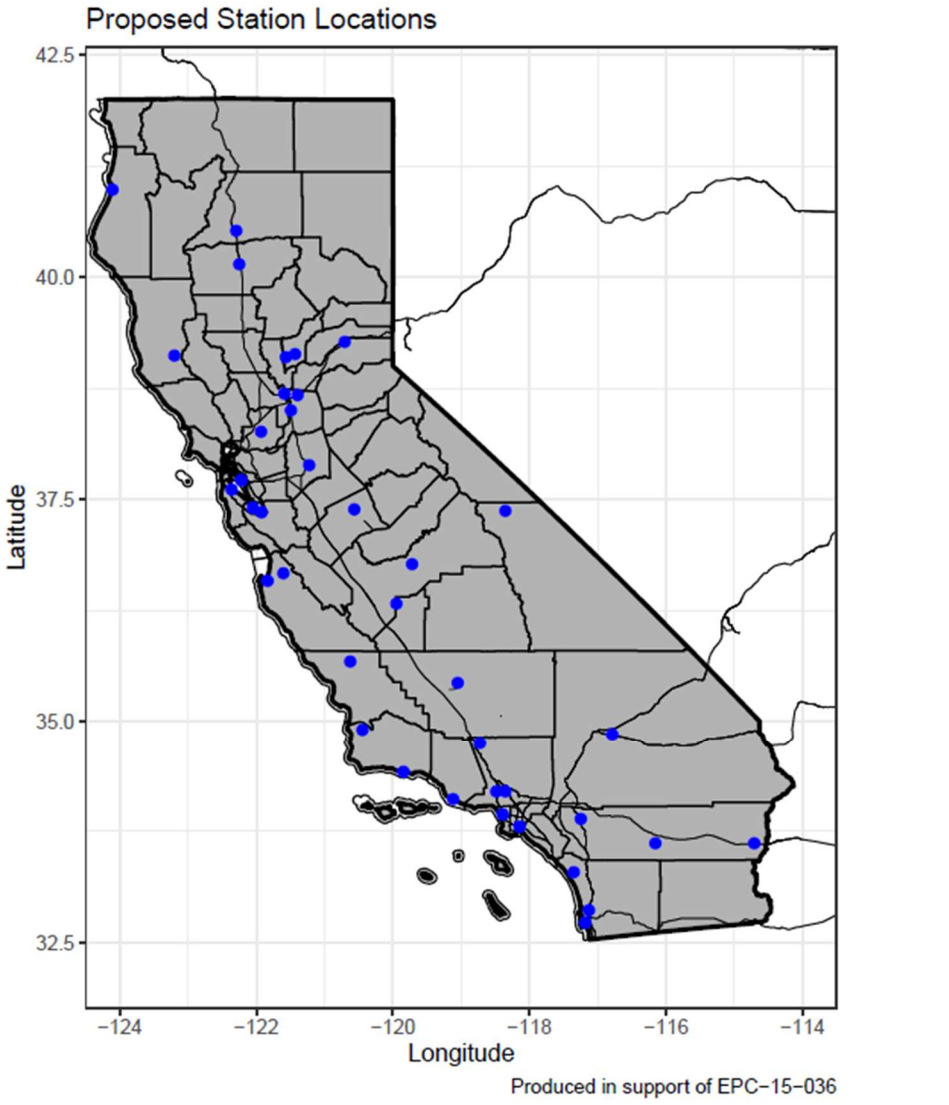

Figure 1. Location of all 39 hourly weather stations across California.

The *Hourly Observed Historical Data* will be updated periodically, with versions noted in file metadata. We anticipate updates occurring monthly.

A version of *The Hourly Observed Historical Data* will be archived annually. Archiving should be performed each spring, as the QA/QC protocol is updated in January -- with the potential for

adjustments to historical values. Version control is important so users can monitor and account for potential changes. Archiving of monthly updates is not necessary as monthly data is just appended to prior months, with no changes to previous values, outside of annual updates.

Data files in each update and archiving will be available in both netcdf and csv format. The csv files are simplified – to include only base metadata, temperature, dewpoint, and sea level pressure.

Observations are taken in the ten minutes preceding each hour. So, for example the 11am observation could have been taken at 10:53am. The precise time of observations is staggered, a historical artifact when observations were transmitted via phonelines, with limited bandwidth capacity on the receiving end. At each observing site this time is consistent (i.e. Sacramento International always reports on :53 before the hour, San Francisco International always reports at :56, etc). Information on the exact time of observation is included in the metadata under the "time" variable.

The following metadata is available for each station, as listed in the table below. Users can access NCDC's station lookup website for additional information about each station, which can be helpful in interpreting inhomogeneity test results.

Table 1. Metadata, units and variables names within netcdf files.

Common Name	Units	Variable Name
Longitude	degrees_east[1]	longitude
Latitude	degrees_north[1]	latitude
Station Elevation	meters	elevation
Station ID Number	ID number	station_id
Time of Measurement	Hours since 1931-01-01 00:00 UTC	time

[1] Units of degrees for Longitude and Latitude are in decimal degrees.

Variables, units, and variable name within the netcdf are listed below:

Table 2. Primary variables within each netcdf file.

Common Name	Full Name	Units	Variable Name
Temperature	Dry bulb air temperature at screen height (~2m)	degrees C	temperatures
Dew Point	Dew point temperature at screen height (~2m)	degrees C	dewpoints
Sea Level Pressure	Reported Sea Level Pressure at screen height (~2m)	hPa	slp
Wind Speed	Wind speed at mast height (~10m)	m/s	windspeeds
Wind Direction	Wind Direction at mast height (~10m)	degree	winddirs
Cloud Cover	Total cloud cover (oktas)	oktas	total_cloud_cover
Low Cloud Cover	Low cloud cover (oktas)	oktas	low_cloud_cover
Mid Cloud Cover	Mid cloud cover (oktas)	oktas	mid_cloud_cover
High Cloud Cover	High cloud cover (oktas)	oktas	high_cloud_cover

Supplemental variables included in the netcdf, but not evaluated by Cal-Adapt or California Energy Commission staff are:

Table 3. Supplemental variables within each netcdf file.

Common Name	Full Name	Units	Variable Name
Pressure	Reported Station Level Pressure at screen height (~2m)	hPa	reported_station_level_pressure_at_screen_height
1 hourly precipitation	Depth of Precipitation Reported in 1 hour (from all four ISD fields)	mm	precip1_depth
2 hourly precipitation	Depth of Precipitation Reported in 2 hour (from all four ISD fields)	mm	precip2_depth
3 hourly precipitation	Depth of Precipitation Reported in 3 hours (from all four ISD fields)	mm	precip3_depth
6 hourly precipitation	Depth of Precipitation Reported in 6 hour (from all four ISD fields)	mm	precip6_depth
9 hourly precipitation	Depth of Precipitation Reported in 9 hour (from all four ISD fields)	mm	precip9_depth
12 hourly precipitation	Depth of Precipitation Reported in 12 hour (from all four ISD fields)	mm	precip12_depth
15 hourly precipitation	Depth of Precipitation Reported in 15 hour (from all four ISD fields)	mm	precip15_depth
18 hourly precipitation	Depth of Precipitation Reported in 18 hour (from all four ISD fields)	mm	precip18_depth
24 hourly precipitation	Depth of Precipitation Reported in 24 hour (from all four ISD fields)	mm	precip24_depth
Cloud Base	Cloud base of lowest cloud layer	meters	cloud_base
Wind Gust	Wind Gust Speed at mast height (~10m)	m/s	wind_gust
Present Weather	Reported past significant weather phenomena	text	past_sigwx1

Not all Variables Are of Equal Credibility

Variables contained in this archive undergo different types of QA/QC protocols, the application of which changes by variable type.

Table 4. Number of QA/QC tests per primary variable.

Variable	Number of Tests
Dew point	19
Sea Level Pressure	16
Temperature	14
Wind Speed	13
Precipitation	3
Wind Direction	3
Cloud	3 to 7

A subjective analysis of likely data quality is provided below to help guide users in appropriate use of the data.

Table 5. Subjective analysis quality of primary variables.

Variable Name	Degree of QA/QC	Notes
Temperature	Very High	Stations chosen based on analysis of temperature
Dew Point	Very High	Likely to be of similar quality as temperature
Mean Sea Level Pressure	Very High	Likely to be of similar quality as temperature
Wind Speed	High	
Wind Direction	Medium	Highly spatially variant – no nearest neighbor test
1-Hour Precipitation	Medium	Highly spatially variant – no nearest neighbor test
Total Cloud Cover	Low	Observing Techniques Changed

When Should I Be Cautious About Using This Data?

Extreme Value Analysis. Automated QA/QC approaches seek to remove erroneous values in a historical weather record. In such an approach large jumps in values can be flagged as suspect and removed, as unexpected large deviation in current weather are likely to be caused by instrumentation or recording errors. In California, particularly in areas of rugged terrain and in coastally influenced regions, large jumps in values could be results in change in weather conditions and incorrectly identified as erroneous. For applications where sudden large shifts are of critical operational relevance (i.e. wildfire) users are encouraged to use the QA/QC flags (which include removed values) and carefully examine historical records.

Trend Analysis. Weather stations are periodically moved, and the methods for observing and recording weather conditions have changed over time. Such changes can introduce inhomogeneities or "jumps" in the record. Such inhomogeneities have been identified and noted for temperature, dew point, sea level pressure and wind speed. Users should assess if their variable of interest and station of interest have such inhomogeneities and perform appropriate bias correction before performing trend analysis. Users

should not perform trend analysis on precipitation, cloud cover or wind direction data without performing their own homogeneity test on the data.

Using Precipitation and Wind Direction. Precipitation data undergoes a different, more simplistic QA/QC protocol than the other critical variables (temperature, dewpoint and mean sea level pressure). This is because, in part, precipitation is highly spatially variant and comparing between neighboring stations is not appropriate. Further, manual weather observers at airports are less likely to provide backup measurements of precipitation. We anticipate more frequent missing hourly precipitation estimates.

Wind direction is included in this database and has limited QA/QC tests (3) to identify errors. Near surface wind conditions vary strongly from location to location, as surface topography and microclimates greatly impact the variability. As such, tests for wind direction quality that compare wind results to neighboring stations are problematic. Broadly this means wind direction data may contain more frequent errors and/or removed values. Wind direction data can be used after careful review by users.

When Should I Not Use This Data?

Using Cloud Cover. Use of cloud cover data (low-cloud, middle-cloud, high-cloud, total-sky cover) is not suggested at this time. Systematic changes in observation techniques has reduced the stability of cloud observations over the record. Early in the period, observations were taken manually by observers, through visual inspection supported by ground based estimated. In the late 1990s many stations were automated, with recording devices that could not reliably detect clouds above 12,000 feet. Complicating things further, major airports continued recording clouds manually (including clouds above 12,000 feet) while regional airports did not. All stations are likely to exhibit inhomogeneities from observational practices. Advanced users who wish to use this data should do so with extreme caution, and only after diligent quality control.

Supplemental Weather Data. Supplemental weather variables (as listed above) have not been evaluated for accuracy and are not advised for use in technical analysis at this time.

References.

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

Smith, A., et al. (2011): The Integrated Surface Database: Recent Developments and Partnerships. Bulletin of the American Meteorological Society, 92, 704-708

Acknowledgements. Data used in this archive is produced by the Met Office, www.metoffice.gov.uk/hadobs. We thank and acknowledge the helpful comments submitted by CEC's EAD/DAO, PG&E and other natural gas sector stakeholder used in informing the producing this dataset.