

Daymet: Station-Level Inputs and Cross-Validation for North America, Version 4 R1

Get Data

Documentation Revision Date: 2023-03-06

Dataset Version: 4.5

Summary

This dataset reports the station-level daily weather observation data and the corresponding cross-validation results for three Daymet model parameters: minimum temperature (tmin), maximum temperature (tmax), and daily total precipitation (prcp) across continental North America (including Canada, the United States, and Mexico), Hawaii, and Puerto Rico. Each data file contains the daily observations and cross-validation results for one parameter for each modeled region and each year, that is, from 1980 to the current calendar year for stations across continental North America and Hawaii and from 1950 to the current year for Puerto Rico. Also included are corresponding station metadata files listing every surface weather station used in Daymet processing for each parameter, region, and year and containing the station name, station identification, latitude, and longitude. The data are provided in netCDF and text formats. In Version 4 R1 (ver 4.4), all 2020 and 2021 files were updated to improve predictions especially in high-latitude areas. It was found that input files used for deriving 2020 and 2021 data had, for a significant portion of Canadian weather stations, missing daily variable readings for the month of January. NCEI has corrected issues with the Environment Canada ingest feed which led to the missing readings. The revised 2020 and 2021 Daymet V4 R1 files were derived with new GHCNd inputs. Files outside of 2020 and 2021 have not changed from the previous V4 release.

These data are provided to facilitate station-level cross validation analyses that can offer insights into the regional accuracy of the Daymet model for the three station-level input parameters. Regression statistics and mean absolute error can be derived to assess temporal and spatial patterns of uncertainty.

This dataset contains 954 data files. There are 477 netCDF (*.nc) files that contain the daily input observations and cross-validation results. Each variable-region-year combination is a separate file. There are 129 files each for North America and Hawaii, and 219 files for Puerto Rico. Provided as additional files, are 477 corresponding text (.txt) files that contain metadata for every surface weather station used in Daymet processing for the variable-region-year combinations.

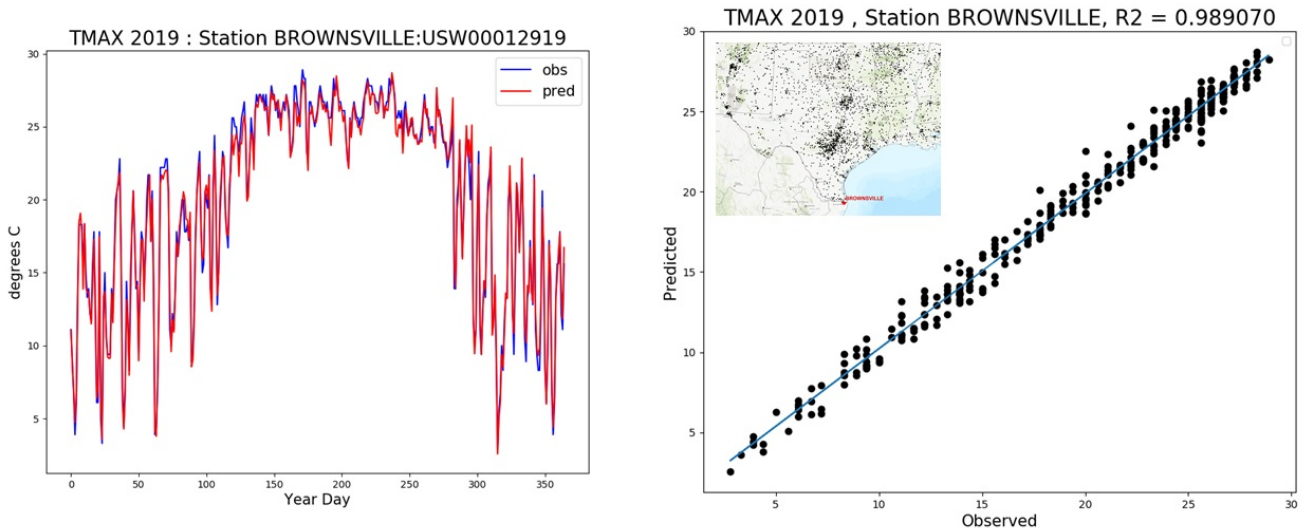


Figure 1. Plot of daily maximum temperature observed vs predicted using Daymet's cross-validation protocol (left) for one station from the Daymet 2019 cross-validation dataset. The right plot shows those data plotted on a 1:1 line with an R2 of 98.9%. The station location (Southern Texas on the Gulf Coast) is shown in the inset.

Citation

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S-C. Kao, and B.E. Wilson. 2022. Daymet: Station-Level Inputs and Cross-Validation for North America, Version 4 R1. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2132>

Table of Contents

1. [Dataset Overview](#)
2. [Data Characteristics](#)
3. [Application and Derivation](#)
4. [Quality Assessment](#)
5. [Data Acquisition, Materials, and Methods](#)
6. [Data Access](#)
7. [References](#)
8. [Dataset Revisions](#)

1. Dataset Overview

This dataset reports the station-level daily weather observation data and the corresponding cross-validation results for three Daymet model parameters: minimum temperature (tmin), maximum temperature (tmax), and daily total precipitation (prcp) across continental North America (including Canada, the United States, and Mexico), Hawaii, and Puerto Rico. Each data file contains the daily observations and cross-validation results for one parameter for each modeled region and each year, that is, from 1980 to the current calendar year for stations across continental North America and Hawaii and from 1950 to the current year for Puerto Rico. Also included are corresponding station metadata files listing every surface weather station used in Daymet processing for each parameter, region, and year and containing the station name, station identification, latitude, and longitude. These data are provided to facilitate station-level cross validation analyses that can offer insights into the regional accuracy of the Daymet model for the three station-level input parameters. Regression statistics and mean absolute error can be derived to assess temporal and spatial patterns of uncertainty. The data are provided in netCDF and text formats. In Version 4 R1 (ver 4.4), all 2020 and 2021 files were updated to improve predictions especially in high-latitude areas. It was found that input files used for deriving 2020 and 2021 data had, for a significant portion of Canadian weather stations, missing daily variable readings for the month of January. NCEI has corrected issues with the Environment Canada ingest feed which led to the missing readings. The revised 2020 and 2021 Daymet V4 R1 files were derived with new GHCNd inputs. Files outside of 2020 and 2021 have not changed from the previous V4 release.

Project: [Daymet](#)

Daymet provides long-term, continuous, gridded estimates of daily weather and climatology variables by interpolating and extrapolating ground-based observations through statistical modeling techniques. The Daymet data products provide driver data for biogeochemical terrestrial modeling and have myriad applications in many Earth science, natural resource, biodiversity, and agricultural research areas. Daymet weather variables include daily minimum and maximum temperature, precipitation, vapor pressure, shortwave radiation, snow water equivalent, and day length produced on a 1 km x 1 km gridded surface over continental North America and Hawaii from 1980 and over Puerto Rico from 1950 through the end of the most recent full calendar year.

Daymet is a research product of the [Environmental Sciences Division](#) at Oak Ridge National Laboratory, Oak Ridge, TN. Daymet is supported by [NASA](#) through the Earth Science Data and Information System (ESDIS) and the [Terrestrial Ecology Program](#). Daymet algorithm and processing development is also supported by the [Office of Biological and Environmental Research](#) within the U.S. Department of Energy's Office of Science.

Related Publication

Thornton, P. E., R. Shrestha, M. Thornton, S. Kao, Y. Wei, and B. E. Wilson. 2021. Gridded daily weather data for North America with comprehensive uncertainty quantification. *Scientific Data* 8:190. <https://doi.org/10.1038/s41597-021-00973-0>

Related Datasets

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2022. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2129>

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2022. Daymet: Annual Climate Summaries on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2130>

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2022. Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2131>

Acknowledgments

The continued development of the Daymet algorithm and processing is supported by the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

Puerto Rico data for Daymet V4 is available for an extended temporal period starting in 1950. Support was provided by the DOE NGEE Tropics Phase 2 funding to extend the temporal period.

2. Data Characteristics

Spatial Coverage: Continental North America (Mexico, United States, Canada), Hawaii, and Puerto Rico

Spatial Resolution: Point locations

Temporal Coverage:

Continental North America (Mexico, United States, Canada) and Hawaii: 1980 to 2022

Puerto Rico: 1950 to 2022

Temporal Resolution: Daily

Study Area: (All latitude and longitude given in decimal degrees)

Site	Westernmost Longitude	Easternmost Longitude	Northernmost Latitude	Southernmost Latitude
Continental North America, Puerto Rico, Hawaii	-178.13	-53.06	82.91	14.07

Data File Information

Observation and Predicted Values Files

There are 477 netCDF files that contain the daily input observations and predicted values. Each variable-region-year combination is a separate file. There are 129 files each for North America and Hawaii, and 219 files for Puerto Rico. These station-level files contain daily data from Daymet input surface weather stations as separate files for each of the three input variables, for each region, and each year of available data.

Filename format: daymet_v4_stnxval_<pppp>_<region>_<yyyy>.nc, where

<pppp> is the respective meteorological variable (tmax, tmin, and prcp)

<region> is 'na' (for continental North America), 'hi' (for Hawaii), or 'pr' (for Puerto Rico), and

<yyyy> is year (1950 through 2022).

Table 1. Parameters included in all netCDF files.

Parameter	Abbreviation	Units	Description
Station Identification	station_id		The ground weather station identification (as a string value) as extracted from the GHCN-Daily input files.
Station Name	station_name		The ground weather station name (as a string value) as extracted from the GHCN-Daily input files.
Station Elevation	stnz	m	The station elevation reported in the metadata readme as extracted from the GHCN-Daily input files.
Days Since 1950-01-01	time	day	The day number since the beginning of the dataset. Data are in a daily time step, defined as 24-hour day based on local time from mid-night to mid-night.
Beginning/End of Time Range	time_bnds	day	The start and end time points for each day (24-hr period based on local time).
Station Index	stns		An integer station index within the netCDF file.
Station Longitude	stn_lon	decimal degrees	The station longitude in WGS 84 (EPSG:4326) reported in the metadata readme as extracted from the GHCN-Daily input files.
Station Latitude	snt_lat	decimal degrees	The station latitude in WGS 84 (EPSG:4326) reported in the metadata readme as extracted from the GHCN-Daily input files.

Table 2. Parameters specific to precipitation netCDF files.

Parameter	Abbreviation	Units	Description
Observation	obs	mm	Station observed daily total precipitation in millimeters, sum of all forms converted to water-equivalent. Data are in a daily time step, defined as 24-hour day based on local time from mid-night to mid-night.
Predicted	pred	mm	Daymet model predicted daily total precipitation in millimeters, using Daymet's cross-validation protocol, sum of all forms converted to water-equivalent. Data are in a daily time step, defined as 24-hour day based on local time from mid-night to mid-night.

Table 3. Parameters specific to maximum and minimum temperature netCDF files.

Parameter	Abbreviation	Units	Description
Observation	obs	degrees C	Station observed daily maximum (or minimum respective of file) 2-meter air temperature in degrees Celsius. Data are in a daily time step, defined as 24-hour day based on local time from mid-night to mid-night.
Predicted	pred	degrees C	Daymet model predicted daily maximum (or minimum respective of file) 2 m air temperature in degrees Celsius using Daymet's cross-validation protocol. Data are in a daily time step, defined as 24-hour day based on local time from mid-night to mid-night.

Station Metadata Files

Supplementary information for weather stations is provided in 477 text (.txt) files. These files contain metadata information for every surface weather station included in the corresponding netCDF files for each variable, region, and year data file. These are the surface weather stations used in Daymet processing for the variable-region-year combinations. Each file represents a unique variable-region-year combination and is in a comma-separated value format.

Filename format: stnxvalmeta_<pppp>_<region>_<yyyy>.txt, where

<pppp> is the respective meteorological variable (tmax, tmin, and prcp)

<region> is 'na' (for continental North America), 'hi' (for Hawaii), or 'pr' (for Puerto Rico), and

<yyyy> is year (1950 through 2022).

Table 4. Fields in the station metadata files.

Parameter	Abbreviation	Units	Description
Station Identification	stn_id		The ground weather station identification (as a string value) as extracted from the GHCN-Daily input files.
Station Name	stn_name		The ground weather station name (as a string value) as extracted from the GHCN-Daily input files.

Station Longitude	stn_lat	decimal degrees	The station latitude in WGS 84 (EPSG:4326) reported in the metadata readme as extracted from the GHCN-Daily input files.
Station Latitude	snt_lon	decimal degrees	The station longitude in WGS 84 (EPSG:4326) reported in the metadata readme as extracted from the GHCN-Daily input files.
Station Elevation	stnz	m	The station elevation reported in the metadata readme as extracted from the GHCN-Daily input files.

The Daymet Calendar

The Daymet calendar is based on a standard calendar year. All Daymet years, including leap years, have 1–365 days. For leap years, the Daymet data includes leap day (February 29) and December 31 is discarded from leap years to maintain a 365-day year.

Version Information

The data are stored and distributed as an individual CF-Compliant netCDF file for each year, region, and variable. The most current Daymet data are being delivered to the user in both Daymet software and Daymet data versions. Version information is recorded in the header file of each of the CF-netCDF files within the Global Attribute fields; Version_software and Version_data. All Daymet data are provisional and subject to revision.

3. Application and Derivation

The purpose of the cross-validation analysis is to provide users with details needed to evaluate the fitness of the Daymet data products for each particular application.

The cross-validation analysis treats each variable-station-year of data from the input station lists as a unique cross-validation record (CVR). Each primary variable (tmax, tmin, and prcp) is handled separately at each station location and for each year. For each such record, estimates of the primary variable are made by dropping that record from the input station list and using the estimation methods to make estimates for the primary variable on each day of the year as described in Thornton et al. (2021). From each CVR, only days with non-missing observed data are used to calculate error statistics. The number of missing days in each CVR is used to provide appropriate weights when reporting multi-station averages or time series summaries for the cross-validation results.

These CVRs include all of the relevant metadata associated with each station, to allow users to assess patterns of error in relation to station location, observation type, station network, year of observation, seasonal patterns, or other analyses as deemed appropriate by the users.

Many different summary statistics can be computed from these paired daily observations and cross-validation results, such as mean absolute error, root mean squared error or bias. Different time periods can also be evaluated, according to the user's needs.

Overall, the Daymet data have broad applications over a wide variety of research areas, including hydrology, terrestrial vegetation growth models, carbon cycle science, and regional to large scale climate change analysis. Measurements of near-surface meteorological conditions are made at many locations, but researchers are often faced with having to perform ecosystem process simulations in areas where no meteorological measurements have been taken. The gridded surfaces of the Daymet data were developed to overcome these limitations.

4. Quality Assessment

Within the Daymet model algorithm, the Daymet cross-validation analyses were used to test the sensitivity of Daymet model methods to the variation of parameters and to estimate the prediction errors associated with the final selected parameters. The general cross-validation protocol is to withhold one observation at a time from a sample, generate a prediction error for the withheld case by comparison with observed values within the sample, and repeat over all observations in the sample to generate an average prediction error. The sample, in this case, was derived from the distribution of nearby stations based on the moving Gaussian distribution of stations for tmax, tmin, or prcp on a given day. Interest is in both the absolute value and the sign of prediction errors generated in this manner.

5. Data Acquisition, Materials, and Methods

All versions of the Daymet algorithm, use as its core input daily observations of minimum temperature, maximum temperature, and precipitation from a network of ground-based weather stations. The surface weather input observations for processing Daymet were obtained from the NOAA National Centers for Environmental Information's Global Historical Climatology Network (GHCN)-Daily data set (Menne et al., 2012a; Menne et al., 2012b). Three separate input station files were generated for the continental North America, Puerto Rico, and Hawaii study area, and these 3 spatially distinct areas were applied separately through the Daymet algorithm.

GHCN-Daily Version Download Dates:

- Daymet V4 for years 1980 - 2019 (or 1950 - 2019 (Puerto Rico only)) used GHCN Daily version 3.27 downloaded on February 5, 2020.
- Daymet V4 for year 2020 used GHCN Daily version 3.28 downloaded on February 14, 2021.
- Daymet V4 for year 2021 used GHCN Daily version 3.28 downloaded on February 16, 2022.
- Daymet V4 R1 for year 2020 used GHCN Daily version 3.29 downloaded on October 6, 2022.
- Daymet V4 R1 for year 2021 used GHCN Daily version 3.29 downloaded on September 18, 2022
- Daymet V4 R1 for year 2022 used GHCN Daily version 3.29 downloaded on February 11, 2023

The Daymet approach to estimating daily surface weather parameters at locations lacking instrumentation is based on a combination of interpolation and extrapolation, using inputs from multiple instrumented sites and weights for each site that reflect the spatial and temporal relationships of the estimation location to the instrumental observations. The approximate number of instrumental observations to use for each estimation is defined as a parameter for each of the primary Daymet variables. As part of a series of algorithm modifications intended to improve robustness in regions of very low station density, the Daymet V4 algorithm drops the iterative station density calculation and instead defines a search radius for each estimation location which is sized to capture exactly the average number of input stations, based on pre-calculated arrays of station distances. Given the pre-processed input station observations and the pre-calculated station lists and interpolation weights for each location in the estimation grid, two separate workflows are used to produce the primary Daymet output variables: one for the daily temperature variables (tmax and tmin) and another for the daily precipitation variable (prcp).

The Daymet Version 4 methodology and dataset is described in Thornton, et al. (2021).

In addition to daily maximum and minimum temperature and daily total precipitation, the Daymet data record includes estimates of other important surface

weather quantities that are not routinely observed or are available as observations from only a small fraction of the temperature and precipitation observing stations. These secondary output variables are daily total shortwave radiation (srad), daily average water vapor pressure (vp), duration of the daylight period (daylength), and a simple estimate of accumulated snowpack, measured as snowpack water equivalent (swe). The daylength estimate is based on geographic location and time of year. Estimates for the other secondary variables (srad, vp, and swe) are derived from the primary temperature and precipitation variables on the basis of theory and empirical relationships, as further described in Thornton et al. (2021).

6. Data Access

These data are available through the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC).

[Daymet: Station-Level Inputs and Cross-Validation for North America, Version 4 R1](#)

Contact for Data Center Access Information:

- E-mail: uso@daac.ornl.gov
- Telephone: +1 (865) 241-3952

7. References

Menne, M.J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E. Gleason, and T.G. Houston 2012a. Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. NOAA National Climatic Data Center. Version 3.27-upd-2020020523. <https://doi.org/10.7289/V5D21VHZ>

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

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2022. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2129>

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2022. Daymet: Annual Climate Summaries on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2130>

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2022. Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2131>

Thornton, P. E., R. Shrestha, M. Thornton, S. Kao, Y. Wei, and B. E. Wilson. 2021. Gridded daily weather data for North America with comprehensive uncertainty quantification. *Scientific Data* 8:190. <https://doi.org/10.1038/s41597-021-00973-0>

8. Dataset Revisions

ORNL DAAC Version Record for Version 4

Daymet Product Version	ORNL DAAC Release Date	Description
Version 4.5	March 5, 2023	ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America for year 2022.
Version 4.4	November 1, 2022	In this Version 4.4, all 2020 and 2021 files (36 total) were updated to improve predictions especially in high-latitude areas. It was found that input files used for deriving 2020 and 2021 data had, for a significant portion of Canadian weather stations, missing daily variable readings for the month of January. The missing readings have been verified by the National Centers for Environmental Information (NCEI) who maintain the Global Historical Climatology Network daily (GHCNd) database. NCEI has corrected issues with the Environment Canada ingest feed which led to the missing readings and has built in more protections to prevent the loss of data. The revised 2020 and 2021 Daymet V4 files were derived with new GHCNd inputs. Files outside of 2020 and 2021 have not changed.
Version 4.3		Version 4.3 of the monthly and annual climatologies was an update to cloud optimized GeoTIFF. Because the crossval and daily data do not have any GeoTIFF files, there was no version 4.3 issued for these data products.
Version 4.2	April 7, 2022	ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America for year 2021.
Version 4.1	April 7, 2021	ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America for year 2020.
Version 4.0	December 15, 2020	ORNL DAAC released daily gridded mosaics for Continental North America (Mexico, United States, Canada) and Hawaii for 1980–2019 and for Puerto Rico for 1950–2019 for each Daymet Version 4 variable. ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America and Hawaii for years 1980-2019 and for Puerto Rico for 1950–2019.

ORNL DAAC Version Record for Version 3

Daymet Product Version	ORNL DAAC Release Date	Description
Version 3.4	March 31, 2020	ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America for year 2019.

Version 3.3	April 8, 2019	ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America for year 2018.
Version 3.2	July 6, 2018	ORNL DAAC released Station-Level Inputs and Model Predicted Values for North America for year 2017.
Version 3.1	May 25, 2017	ORNL DAAC archived and released Version 3 of Station-Level Inputs and Model Predicted Values for North America for 1980–2016



[Privacy Policy](#) | [Feedback](#) | [Help](#)



Home

About Us

- Mission
- Data Use and Citation
- Policy
- User Working Group
- Partners

Get Data

- Science Themes
- NASA Projects
- All Datasets

Submit Data

- Submit Data Form
- Data Scope and Acceptance
- Data Authorship Policy
- Data Publication Timeline
- Detailed Submission Guidelines

Tools

- MODIS
- THREDDS
- SDAT
- Daymet
- Airborne Data Visualizer
- Soil Moisture Visualizer
- Land - Water Checker

Resources

- Learning
- Data Management
- News
- Earthdata Forum [↗](#)

Contact Us

