Daily Gridded Fields and Area Averages for the Contiguous United States (nClimGrid-Daily)

User Guide

1. Intent of This Document and Points of Contact

1a**)** This document is intended to provide basic information, references with additional information, and points of contact for the collection of 1951-present daily gridded fields and area averages of temperature and precipitation, hereafter referred to as nClimGrid-Daily.

1b**)** Technical point of contact: ncei.grids@noaa.gov

1c) Customer service contact: ncei.orders@noaa.gov

2. Product Overview and Intended Uses

NCEI's nClimGrid-Daily product contains gridded fields and area averages of daily maximum, minimum, and average temperature (Tmax, Tmin, and Tavg) and daily precipitation amount (Prcp) for the Contiguous United States (CONUS) between January 1, 1951, and the present. The gridded fields cover the land area between approximately 24°N and 49°N and between 67°W and 125°W at a resolution of 1/24 of a degree (0.0417°). Thus, the gridpoint spacing is 4.6 km between latitudes and ranges from 3.0 km in the Northern CONUS to 4.2 km in the South between longitudes. Area averages are available for nine commonly-used types of regions. These include census tracts, counties, Climate Divisions, states, major hydrologic units, Weather Forecast Office regions of the National Weather Service (NWS), the nine regions used in NCEI's climate monitoring reports, the seven regions used in the Fifth National Climate Assessment (NCA5), and the Contiguous U.S. as a whole.

An auxiliary product, nClimGrid-Daily-Auxiliary, provides daily gridded fields of counts of the number of Tmax, Tmin, and Prcp observations within 48.3 km (30 miles) of each grid point.

These counts were designed according to specifications provided by the U.S. Army Corps of Engineers and provide an indication of the observation density, which is one of the key factors in determining the uncertainty in the grid point estimates.

The primary purpose of these products is to support applications such as drought monitoring that require time series of spatially and/or temporally aggregated grid point values. Reliance on single-day values and individual points is discouraged due to the significant uncertainty that is inherent in such a product, as a result of the spatial distribution of the underlying observations, differences in observation time between neighboring stations, and interpolation errors. Spatial and temporal averaging tends to reduce the effect of these uncertainties, and time series of such aggregated values can be shown to be suitable for climatological applications.

3. Data Field Description

3a.) Overview

Within the primary nClimGrid-Daily product, gridded fields are organized into one Network Common Data Format (NetCDF) file per year-month, each containing gridded fields for the 4 variables (Tmax, Tmin, Tavg, and Prcp) for each day of the month. Area averages are stored in a total of 36 comma-separated value (CSV) files per month (four variables times nine types of regions). In addition to the gridded fields and area averages in nClimGrid-Daily, nClimGrid-Daily-Auxiliary contains one netCDF file per month, with each file providing the observation counts for Tmax, Tmin, and Prcp at each grid point for each day of the month.

3b.) filenames

File names are of the form ncdd-YYYYMM-grd-SSSSSS.nc for gridded data files, ncddsupp-YYYYMM-obcounts.nc for the grids of auxiliary observation counts, and EEEE-YYYYMM-RRR-SSSSSS.csv for area-average files. In these names, ncdd stands for "nClimGrid-Daily data", ncddsupp stands for "nClimGrid-Daily data supplement", and the capital letters have the following meanings.

EEEE is the meteorological element, or variable. It has four possible values:

prcp = Precipitation (mm)

tavg = Average temperature (°C), calculated as the average of TMAX and TMIN

tmax = daily maximum temperature (°C)

tmin = daily minimum temperature (°C)

YYYY is the year and ranges from 1951 to the current year.

MM is the current month and ranges from 01 for January to 12 for December.

RRR is the type of region to which the values in the file apply. It can take the following values:

cen = census tracts

cns = Contiguous U.S.

cty = Counties

div = Climate divisions

hc1 = Hydrologic Unit Code (HUC) 1 regions

nca = NCA regions

reg = NCEI regions

ste = state

wfo = Weather Forecast Office (WFO)

SSSSSS is the file status. It can take one of two values:

prelim = Preliminary, indicating that the values in the file have not been scaled to match the corresponding values from the monthly nClimGrid or nClimDdiv products.

scaled = Scaled, indicating that the values in the file have been scaled to match corresponding monthly values

For example, ncdd-201712-grd-scaled.nc contains the gridded fields for each day of December 2017, ncddsupp-201712-obcounts.nc contains the corresponding observation counts, and tmax-202209-ste-prelim.csv contains the preliminary average of daily maximum temperature for each of the 48 contiguous states for each day in January 1951.

3c.) Format of the CSV Files of Area Averages

Each of the CSV files contains one record per geographic entity. For example, the file containing state averages for one variable, year, and month has one record for each of the 48 states in the contiguous U.S.

The first three columns identify the region type, region code, and region name. For example, corresponding entries for the state of North Carolina are ste, 31, North Carolina. Numeric identifiers for U.S. states, which are also contained in the county and climate division identifiers, follow the NCEI numbering system that is used in the station identifiers of sites within the U.S. Cooperative Observer Network. For convenience, a supplementary cross-reference between these NCEI state codes and the Federal Information Processing System (FIPS) state codes is provided.

The fourth and fifth columns identify the four-digit year and two-digit month. Example: 1951,01.

The Sixth column identifies the variable: PRCP = precipitation (mm); TAVG, TMAX, and TMIN = average, maximum, minimum temperature (degrees Celsius), respectively.

The remaining 31 columns provide the values of the specified variable on each day of the given month. These values are provided in fixed-width format of the form %8.2f, i.e., as a real number with two significant digits. Values on nonexistent days such as February 30 are set to -999.99.

4. Data Sources and Processing

The daily gridded fields are generated by interpolating morning and midnight observations from the Global Historical Climatology Network - Daily (GHCNd) dataset (Menne et al. 2012), using thin-plate smoothing splines as implemented by the Australian National University (ANUSPLIN, Hutchinson 2007). Additional processing steps improve the identification of grid points without precipitation and ensure internal consistency between daily maximum and minimum temperatures as well as consistency with NCEI's monthly nClimGrid product (Vose et al. 2014). In the case of temperature, consistency with the monthly products also means that month-to-month inhomogeneities in the monthly mean have effectively been removed since the monthly products are based on monthly-mean temperatures that have been adjusted using the pairwise homogenization algorithm (PHA) of Menne and Williams (2009). The details of these processing steps, as well as the rationale behind them, are provided in Durre et al. (2022a).

Each final grid point estimate is assigned to the centroid of a grid box and, due to the smoothing effect of the thin-plate smoothing splines, can be treated as an average value for the entire grid box. The area averages then are computed as means of the cosine-latitude-weighted values at grid points within the desired region. Census tract values are based on the tract boundaries associated with the 2020 census. For census tracts that have an area smaller than a grid box, the value at the nearest grid point is used.

The auxiliary counts were obtained from the set of input observations used to construct the gridded fields. The count at each grid point represents the number of Tmax, Tmin, or Prcp observations located within 48.3 km (30 miles) of the coordinates of the grid point. Higher counts indicate a greater density of observations in the vicinity of the grid point, and the accuracy of a grid point estimate typically increases with increasing density of the nearby observations. Note, however, that the radius of 48.3 km does not imply that observations outside of that radius did not influence the estimate; it merely serves as a proxy for estimating the density among what are likely to be the most influential observations.

Although nClimGrid-Daily fields and area averages frequently become available two to three days after the observation date, the accuracy of the estimates will improve over subsequent days and months as additional data are received. The amount of input data for a given day can increase by an additional 50% to 100% after the first day on which data become available, with most of that increase taking place in the first few days. It is also worth noting that occasional interruption in the data streams can result in greater latency in the production of nClimGrid-Daily. In addition, estimates for days in the current month will not be fully consistent with either the corresponding monthly products or the daily estimates for previous months and years until approximately the fourth day of the following month. Thus, users are cautioned to assess the impact of using the preliminary estimates on their particular application.

The current version of nClimGrid-Daily and nClimGrid-Daily-Auxiliary is v1.0.0. The incorporation or additions of further input data for the existing variables will not result in a version change. The last version digit will be incremented to signify minor changes in contents or processing software, such as if area averages for an additional area type are included or there has been an update to underlying third-party software that has no significant effect on the output (e.g., a change from Python 2.7 to Python 3.7). The addition of new variables, the use of a different interpolation method, or other significant change in the contents of nClimGrid-Daily or the underlying methodology will be denoted by incrementing the primary version number (e.g., v1.0.0 to v2.0.0). Other changes to the input data or software that have a broad impact on the output will be designated by a change in the middle digit of the version number (e.g., v1.0.0 to v1.1.0).

5. Validation

The performance of nClimGrid-Daily was assessed in two primary ways. First, in order to assess the degree to which the gridded fields can be expected to accurately represent large spatial variability, the spatial patterns of the gridded fields were first qualitatively compared to the corresponding patterns in the input data on days of significant weather events. Second, spatial and temporal analyses of the differences between observations and collocated interpolation estimates characterize the magnitude and variability of the interpolation errors.

The nClimGrid-Daily fields largely agree with patterns of station observations. During 1991-2020, half of the stations tested during cross-validation analyses had an average absolute interpolation error of 1°c or less for both Tmax and Tmin, and there were roughly equal numbers of positive and negative differences between estimates and observations. For precipitation, half of the stations had a median absolute error of of 2 mm or less, and larger errors are more common when precipitation amounts themselves are large.

However, significant uncertainties are associated with estimates at individual grid points on specific days, particularly in the context of significant spatial and temporal variability. For both temperature and precipitation, strong gradients in any terrain can result in decreased accuracy in the interpolated field, particularly when the observations are unevenly or sparsely distributed. Even though grid point estimates generally exhibit reasonable variations with elevation in mountainous regions, the largest interpolation errors are found in the mountains of the Western United States. Further, due to the smoothing effect of the interpolation technique, observations that are warmer, colder, wetter, or drier than neighboring observations may be “smoothed out” such that the nearby gridpoint estimates may be more reflective of the general conditions of the area. This smoothing effect also introduces a slight dry bias of the gridded precipitation fields, a characteristic that is common among gridded datasets.

These analyses imply that nClimGrid-Daily is best suited for applications that require time series of spatially aggregated data. Single-day values at individual points should not be interpreted as an indication of the true conditions at that particular location on that day, but rather as an estimate of conditions that is derived from the observations in the vicinity of the point of interest. Additional details on the validation analyses and examples of climate monitoring applications of nClimGrid-Daily can be found in Durre et al. (2022a).

6. References

Durre, I., A. Arguez, C. J. Schreck III, M. F. Squires, and R. S. Vose, 2022a: Daily high-resolution temperature and precipitation fields for the Contiguous United States from 1951 to Present. Journal of Atmospheric and Oceanic Technology, DOI:10.1175/JTECH-D-22-0024.1.

Durre, I., M. F. Squires, R. S. Vose, A. Arguez, W. S. Gross, J. R. Rennie, and C. J. Schreck, 2022b: NOAA's nClimGrid-Daily Version 1 – Daily gridded temperature and precipitation for the Contiguous United States since 1951. NOAA National Centers for Environmental Information, since 6 May 2022, enter suggestion enter commenthttps://doi.org/10.25921/c4gt-r169.

Hutchinson, M. F., 2007: ANUSPLIN Version 4.37 User Guide, Centre for Resource and Environmental Studies/Australian National University, 54 pp., https://vdocuments.mx/anusplin-version-437-user-guide.html.

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

Menne, M. J., and C. N. Williams, 2009: Homogenization of temperature series via pairwise comparisons, J. Climate, 22, 1700–1717, doi:10.1175/2008JCLI2263.1

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., Jr., Fenimore, C., Gleason, K., & Arndt, D. (2014). Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions, Journal of Applied Meteorology and Climatology, 53(5), 1232-1251. doi:10.1175/JAMC-D-13-0248.1

7. Dataset and Document Revision History

Rev 0 – 09 Sep 2022 - This is a new document/dataset

Rev 0.5 - 17 Sep 2022 - This is the revision after feedback from one internal reviewer.

Rev 0.5 - 23 Sep 2022 - Version published after additional internal review.