Climate Reference Metadata.

1. Introduction

Metadata was collected from a variety of sources for the Climate Reference Network (CRN) stations. The latitude and longitude datum for the stations was used to reference various GIS datasets to extract information related to the topography of the stations, the population surrounding the stations and the land cover, including data pertaining to the urbanity of the sites. The Latitude and Longitude of the stations is reported to 4 significant digits which indicates a precision of close to 10 meters at the equator. The accuracy of these datum in conjunction with the GIS data will tend to be less than this as it is dependent on the registration accuracy of the GIS data and its inherent resolution. Unless otherwise stated all projections are assumed to be WGS84, a standard Lat/lon projection. The code for processing the metadata is located at <https://github.com/stevenmosher/Ghcn_V4_Metadata> and the script for processing the data is CRNmetadata.R. Where feasible the data sources are located in github, otherwise their URL is given in the references. All data sources are open, however, some require registration to acquire. Consequently, there is no code to automate the download of the data and so the scripts do not run turnkey. The user is expected to download the data for themselves and alter the “filenames.R’ source code accordingly.

1. Filename and format.

The filename is CRN\_Metadata\_Final.csv. It is a comma separated file. It is available here: https://github.com/stevenmosher/Ghcn\_V4\_Metadata/blob/master/CRN\_Metadata\_Final.csv

1. Column names
   1. Station\_Id

Station\_Id is a unique station identifier take from the site inventory. It is an integer value that represents the WBAN of the station.

* 1. Name

The station name is a asci record of the station name with spaces replaces by underlines, for example; “Murphy\_10\_W”

* 1. Longitude

The station’s longitude in degrees to 4 significant digits. The projection is WGS84

* 1. Latitude

The station’s latitude in degrees to 4 significant digits. The projection is WGS84

* 1. Elevation

The station’s elevation in meters taken from the station inventory

* 1. DEM1km

In addition to the elevation at the station’s exact latitude and longitude, the average elevation in meters surrounding the station is supplied from a Digital Elevation Model (DEM). The DEM data is supplied for cases where the elevation data may be missing, as is the case with a small percentage of GHCN stations. In addition, the DEM data may indicate that the station is in an area where the surrounding terrain is at a higher elevation and could be subject to cold air drainage. Finally, since the data includes Bathymetry the data can be used to identify stations that are located on ships at sea, oil platforms and other ‘land’ stations that may actually be located in the ocean. The DEM data source –ETOP01-- is described here: <https://www.ngdc.noaa.gov/mgg/global/relief/>. The data source has both elevation and bathymetry at a 1 arc minute resolution, or roughly 2km at the equator. Over land the elevation supplied is the elevation at the top of the ice sheet, which is significant for those stations in Greenland and Antarctica. URL <https://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/data/ice_surface/cell_registered/georeferenced_tiff/ETOPO1_Ice_c_geotiff.zip>.

* 1. DistancetoCoast

Distance to coast represents the distance of the station to a continental coast. The units is kilometers. Negative values represent a station that is inland and positive values a station that is located in the ocean. Stations can be located in the ocean for several reasons: A) slight inaccuracies in the stated location. B) stations located on small islands or atols not in the coastline database. C) stations located on ships or platforms. For CRN stations none of these conditions exist. The dataset is a WGS84 grid spanning -180 to 180 and -90 to 90. The cellsize is .01 degrees or roughly 1km at the equator. The dataset is described here: <https://oceancolor.gsfc.nasa.gov/cms/DOCS/DistFromCoast>

* 1. LCCOwnLabel

The column contains the landcover for the station lat lon as determined by a lookup into the ESA Climate Record Data Product (CRDP) built from SPOT and MERIS data. The time period covered is the 2008-2012 epoch. The data is in a WGS84 projection with a resolution of 300 meters at the equator or 10 arc minutes. The data can be obtained after registration here

<http://maps.elie.ucl.ac.be/CCI/viewer/download.php>. Overall accuracy of the classification is ca. 70%, however for the most relevant classes ( Urban, Water, Croplands) User accuracies exceed 85%. Validation of the classification described here: <http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-PUG-v2.5.pdf>.

* 1. EF\_LF\_Desc

This column contains a description of the topology of the site. The landform data was extracted from the Global USGS dataset using the lat and lon the site. The dataset is described here:

<https://rmgsc.cr.usgs.gov/ecosystems/global.shtml> and more detail here

<https://rmgsc.cr.usgs.gov/ecosystems/docs/AAG_Global_Ecosystems_Booklet.pdf>. The dataset covers the entire world in a WGS84 format at a 250 meter resolution. The landforms are classified based on the underlying DEM data. They include:

Flat Plains, Smooth Plains, Irregular Plains, Escarpments, Low Hills, Hills ,Breaks, Low Mountains, High Mountains/Deep Canyons, Surface Water. The data can be downloaded here: https://rmgsc.cr.usgs.gov/outgoing/ecosystems/Global/

* 1. WaterArea,

The water area is the total sq kilometers of water within a 10km radius of the site’s latitude and longitude. The ESA Landcover product ( see 3.8) is used to determine the amount of area that is water surface within 10km of the site. Water area is provided for the following reasons. Presence of water bodies is known to diminish the UHI effect and it can be used to identify stations that are located on ships and platforms in certain cases. The product used here is the same as for item 3.8. From every location a 10km buffer is interrogated and the area of the cells that are classified as water is summed. The maximum figure is roughly 314sq km.

* 1. UrbanArea10K

The Urban area is the total sq kilometers of urban landcover within a 10km radius of the site’s latitude and longitude. The ESA Landcover product ( see 3.8) is used to determine the amount of area that is urban landcover within 10km of the site. From every location a 10km buffer is interrogated and the area of the cells that are classified as urban is summed. The maximum figure is roughly 314sq km.

* 1. GPwV4\_Area

Population figures for each site are retrieved from a 30 arc minute (1km) gridded dataset. Since the population count is retrieved the population density will be a function of the area of the cell. Since area of the cell changes with latitude a record of the cell area at each site location is recorded. This provides the necessary data for calculating a population density on per sqkm basis. The population dataset is described in 3:13

* 1. GPwV4\_00

Population counts for the site location are extracted from the gridded GWP version 4 dataset. This column represents the population count in the year 2000. The dataset has a 30 arc minute resolution (1km at the equator) and spans from -60 lat to 85N. For locations not over land and outside the coverage area, NA ( Not Available) is recorded. At this stage we do not impute a zero population for those locations and that decision is left to the user.

The methodology is described here <http://beta.sedac.ciesin.columbia.edu/data/collection/gpw-v4> The data uses areal-weighting and maintain fidelity to the input census data. So for example it does not take into consideration the landcover of the pixel or other factors such as water area or slope that other more highly modelled products use. The dataset is available here <http://beta.sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev10> and can be downloaded after registration. The projection is WGS84.

* 1. GPwV4\_05

Population counts for the year 2005 from GPW V4

* 1. GPwV4\_10

Population counts for the year 2010 from GPW V4

* 1. GPwV4\_15

Population counts for the year 2015 from GPW V4

* 1. GPWV4\_Area10

The column is a constant representing the approximate area of a circle with a radius of 10km from the site location. The units are sq km

* 1. GPwV4\_15\_10km

The population counts around every site are aggregated out to a radius of 10km.

* 1. Hyde\_Area

In addition to the GPW v4 data we also provide population counts from Hyde3.1 datasets. The Hyde dataset has population counts every decade back to the beginning of the recording of temperatures. The gridded data is provided in a 5 arc minute resolution ca. 9km at the equator.

* 1. Hyde1970
  2. Hyde1980
  3. Hyde1990
  4. Hyde2000
  5. Hyde2005
  6. GpwV4\_density00
  7. Hyde\_density00
  8. GPW10km\_15\_Density
  9. EST\_POP2000
  10. EST\_POP2000\_50K
  11. DistanceToPlace
  12. DistanceToPlace50K
  13. PopulatedPlace
  14. PopulatedPlace50K
  15. Populated\_Lon
  16. Populated\_Lat
  17. Populated\_Lon50K
  18. Populated\_Lat50K
  19. Urban\_ES00Pop
  20. UrbanArea
  21. UrbanAreaName
  22. UrbanLon
  23. UrbanLat
  24. UrbanDailyMean
  25. BufferDailyMean
  26. UrbanNightMean
  27. BufferNightMean
  28. UrbanRuralDailyDiff
  29. UrbanRuralNightDiff
  30. Airport\_Dist,
  31. Airport\_Dist2,
  32. Airport\_Name,
  33. Airport\_Name2,
  34. Airport\_Type,
  35. Airport\_Type2
  36. Airport\_Lon,
  37. Airport\_Lat,
  38. Airport\_Lon2,
  39. Airport\_Lat2,
  40. Lights

1. References
   1. DEM

Amante, C. and B.W. Eakins, 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M [Access date:12/26/2016]

4.2 Gridded Population of the World Center for International Earth Science Information Network - CIESIN - Columbia University. 2017. Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 10. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).<http://doi.org/10.7927/H4PG1PPM>. Accessed 22 12 2106.

1. Supplementary Information