

Environmental data sources for the Dengue Project

1. Introduction

NOAA data sources are compiled for easy comparison with dengue health data. These environmental data are from Iquitos, Peru and San Juan, Puerto Rico to correspond with the available dengue data. All data are provided in a csv format for easy access. All data are maintained and quality controlled by the programs that manage these data sources. Please be aware that data may still potentially contain quality issues and any user should inspect these data before use.

Environmental data provided for this study are from a variety of sources (ground observations, remote sensing, and reanalysis). Each source has different limitations and quality issues. Users of these data should be careful to utilize data that provide the most confidence for the actual representation of the surrounding conditions. Ground observations are generally an optimal representation of the actual local conditions. Remotely sensed observations are generally an excellent observation of precipitation and vegetation conditions for a location. Reanalysis data are a good representation of the conditions of a given area, especially when other data sources are not available.

Users should also be aware that these environmental data are at varying spatial scales. Station data are at a point-based estimate of the local climate conditions; whereas, the reanalysis and remotely sensed data are grid cells. For more information, please read the following descriptions to learn more about the differences.

Below is a description of the environmental data sources that are provided for analysis with dengue health data.

1.1. NOAA's GHCN daily climate data.

The GHCN stations with daily temperature and precipitation data are available for both locations.

Additional data access and full description of data can be found here:
<https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/>

Station descriptions that are inside or near the city are listed here:

San Juan, Puerto Rico: RQW00011641; Lat = 18.4325; Long = -66.0108; Elevation = 2.7; Start = 1956; End = 2015

Iquitos, Peru: PE000084377; Lat = -3.783; Long = -73.3; Elevation = 126; Start = 1973; End = 2015

Temperature values are in Celsius; Precipitation values are in mm.

Values include Maximum Temperature, Minimum Temperature, Daily Average Temperature, Diurnal Temperature Range, and Daily Precipitation.

*Be aware that some stations may have missing days and/or missing values. Missing values are identified as -9999. Missing days are not identified in the record.

note: for us, missing values have been put in as "NA" - the driven data folks kindly did that for us; they weren't left as -9999

Station List:

ghcnd_stationsLOCATION.csv

Station climate data:

CityNameIDENTIFICATION_NUMBER.csv

*Station observations for San Juan are complete and have few missing values. However, there are multiple missing station observations for Iquitos and users should refer to other data sources provided.

1.2. NOAA's CDR PERSIANN Precipitation Product

PERSIANN is a global climatological data record of precipitation from remote sensing information using an artificial neural network. These data are available for each city on a daily basis from 1983-present.

The resolution of this data product is 0.25x0.25 degree.

Additional data access and full description of data can be found here:

<http://www.ncdc.noaa.gov/cdr/operationalcdrs.html>

Precipitation data are from the grid surrounding the station located in the city or from the city centroid for Rio de Janeiro, as Rio de Janeiro does not have an active climate station.

Precipitation values are reported as daily sums in mm.

*Missing values are listed as -9999.

Precipitation data file:

LOCATION_Precip.csv

1.3. NOAA's NCEP Climate Forecast System Reanalysis

Climate Forecast System Reanalysis is global, high-resolution, coupled atmosphere-ocean-land surface-sea ice system to provide the best estimate of the state of these coupled domains over the period of record.

Data are available from 1979-present.

Additional data access and full description of data can be found here:

http://rda.ucar.edu/datasets/ds093.0/#metadata/detailed.html?_do=y

CFSR provides a variety of data that are not easily accessible from other data sources. This includes relative humidity, specific humidity and dew point.

Temperature data values are available in Kelvin.

Resolution of the grid is 0.5 degree.

Data files are labeled:

LOCATION.csv

1.4. NOAA's CDR Normalized Difference Vegetation Index

NDVI CDR is a global climatological data record of vegetation. These data are available for each city on a weekly basis from 1981-present.

The resolution of the product is at 0.05x0.05 degree.

Additional data access and full description of data can be found here:

<http://www.ncdc.noaa.gov/cdr/operationalcdrs.html>

Four pixels closest to the city centroid are provided for evaluation of vegetation change.

30-year climatologies were also provided for these sites.

Raw Data:

LOCATION_noqc.csv

NDVI Climatologies

LOCATION_clim_noqc.csv

*NDVI data are provided without quality flagged data removed. Quality flagged data are provided at the CDR webpage.

2. Example code to convert the daily values to weekly

Provided below is SAS code to convert the daily environmental observations to weekly averages to remain consistent with the weekly dengue data.

```
PROC IMPORT OUT= WORK.daily
            DATAFILE= "H:\FILENAME.xlsx"
            DBMS=EXCEL REPLACE;
            RANGE="RanalysisTemp$";
            GETNAMES=YES;
            MIXED=NO;
            SCANTEXT=YES;
            USEDATE=YES;
            SCANTIME=YES;
RUN;

proc contents data=daily; run;

data daily; set daily (rename=( _S=month));
run;
data daily; set daily (rename=( _3W=day));
run;
data daily; set daily (rename=(STATION=year));
run;
```

```
data daily; set daily;
date = mdy( month, day, year );
      format date mmddy.;
run;

PROC EXPAND DATA=daily
  FROM=day TO=week OUT=weekly;
ID DATE;
CONVERT tavg=tavg_week;
RUN;

proc print data=weekly;
run;
```