

Ted Dang  
Final Project  
GEOG 375  
Spring 2019  
[teddang@gmail.com](mailto:teddang@gmail.com)

## **PROJECT SUMMARY**

The objective of this project is to produce temperature data where no data exist and store such data as thematic layers in the form of heating degree days (HDD) and cooling degree days (CDD). Mainly used in energy consumption studies, HDD and CDD are major determinants of energy demand, especially with the portion used for heating and cooling of living spaces. Historical daily temperature from over 500 weather stations scattered throughout California are used as input data. This project performs spatial interpolation turning input point data into gridded temperature datasets. The aim is to produce reasonably accurate annual temperature, as raster structured data, at 10 x 10 km spatial resolution for the entire state. The final products contain smooth contours of data showing heating and cooling patterns independent of political boundaries. This makes it possible for aggregation of temperature data at any geographic units greater than 10 km<sup>2</sup> using zonal statistics procedures.

## **PURPOSE**

The purpose of this project is to produce temperature data in the space between weather stations and to store such data as thematic layers in the form of heating degree days (HDD) and cooling degree days (CDD). The final products are raster structured data which can be aggregated into any geographic units using zonal statistics procedures. Aggregated temperature data can be used as determinants for heating and cooling needs of living spaces in energy consumption studies.

## **GEOPROCESSING TASKS**

To achieve the intended results, this script was set up to carry out multiple tasks. First, several modules such as arcpy, sys, traceback, datetime, time, and spatial analyst must be imported. Spatial Analyst extension must be checked out due to licensing restrictions. Overwriting of output was invoked to accommodate reusing of name for intermediate output to minimize use of system memory. The remaining tasks in this script can be loosely grouped into the following activities: (1) Data Input, (2) Loop Creation, (3) Spatial Interpolation, and (4) Raster Calculation and Output. Belows are detailed descriptions of the four task groups.

## DATA INPUT

As temperature data are point-data generated by weather stations (with longitude and latitude), they are brought in using MakeXYEventLayer function. As this is a processor intensive task, steps are taken to pre-process outside of the script. Maximum and minimum temperature values are cleaned and used to calculate average daily temperature. Saved in comma delimited value format, only average temperature along with the most essential spatial attributes are added to the map.

CopyFeatures tool is then used to convert the added XY comma delimited formatted data for use in downstream tasks. This process adds necessary object IDs to the input data set making it possible for saving to a geodatabase. Next, MakeFeatureLayer function is applied to turn it into a feature layer so select statements can be used to subset the dataset. The purpose is to produce feature layers containing only one day of temperature values.

## LOOP CREATION

To produce annual HDD and CDD data from daily temperature recorded over many years, the input data must be first subset by the year. Within that subset dataset, another selection must be made to produce another dataset containing temperature data for the day. This task is achieved by producing a “date\_loop” nested inside a “year\_loop.”

Index values produce by the nested loops are used to produce a query statement. This statement is then used as a parameter in SelectLayerByAttribute function to produce a new feature layer containing only one day of average temperature values. This looping structure allows the process to repeat every day of the year regardless of how many days there are in a year. Once December 31<sup>st</sup> is reached, the process is repeated until last day of the last year is reached.

GetCount function is also used inside the date\_loop to generate and print the total number of records selected. This is to ensure that SelectLayerByAttribute function is executed correctly. Especially with a script that runs slow like this one, having a visual cue helps track progress of code execution.

## INTERPOLATION

Use of spatial interpolation technique to construct data, such as for areas in between sparsely located weather stations, is a reasonably accurate and inexpensive approach to spatial data acquisition. Accuracy of different interpolation methods has been a focus of research for a long time (Hughes, 1982; Philips et al., 1992; Price et al., 2000; Cao et al., 2009). More recent studies show accuracy of estimates is improved with minor variations. Willmot and Matsuura (1995) proposes incorporation of spatially high-resolution digital elevation information. Samanta et al. (2011) suggests inclusion of distance from the coastline and seasonal wind pattern. Yang et al. (2004) tested four methods including Inverse Distance Weighting, Spline, Kriging, and Cokriging. Kriging is recommended. Kriging is the method being used in this project.

According to ESRI's technical documents, Kriging is a, "geostatistical technique used to create surfaces incorporating the statistical properties of the measured data. To make a prediction for an unknown value for a specific location, kriging will use the fitted model from variography, the spatial data configuration, and the values of the measured sample points around the prediction location." Without going into detailed technical aspects, Kriging function can be called with `gp.kriging_sa` clause. For this project, parameter calling for input from 12 nearest weather stations is used. Relying on recorded temperature data from over 500 weather stations scattered across California, this function takes XY point temperature data and returns estimated average temperature value for every 10 km<sup>2</sup> covering the entire state.

## RASTER CALCULATION AND OUTPUT

Since end products are gridded raster data of annual values expressed as HDD and CDD, interpolated surface data need to be accumulated. Raster calculator function makes this possible. Intermediary raster datasets are produced with a "Con" statement. Once done, simple math algebra, in conjunction with the looping structure, is used to accumulate a year worth of raster data into one final raster for the year. Starting with a raster of pixels with zero value called "Raster\_hdd", each daily produced "Raster\_interpolated" is added to it using the following mathematical expression:

$$\text{Raster\_hdd} = \text{Raster\_interpolated} + \text{Raster\_hdd}$$

Once the last day of the year is reached at the end of the inside loop, the last "Raster\_hdd" is saved to disk with a different name including a year index such as "HDD\_1985". Before going through the next loop, "Raster\_hdd" is re-initialized with zero valued pixels and the process repeats.

## **DIFFICULTIES**

Creating a query statement involving time and date indexes was a difficult process. Besides having to adhere to strict Python syntax, it also must agree with MS Windows' date and time format. Hence date and time attributes as downloaded from the source must be reformatted in a way that Windows would accept and can be parsed for use as string query statement.

Due to the lack of experience, figuring out which module works with which function in what Arcgis Desktop version was frustrating. Relying on code output from ESRI's ModelBuilder was a major source of the culprit. Upgrading from Arcgis Desktop 10.6 to 10.7 during the middle of production process was a huge mistake.

Working with raster data was also a humbling experience. Due to the lack of authoritative documentation, consistent instructions, and/or clear explanations via on-line resources, most of knowledge gain for the author came from trials and errors. This was a painful process but for the time being, one must learn to take it in stride.

## REFERENCES

- Cao, W., Hu, J., & Yu, X. (2009, August). A study on temperature interpolation methods based on GIS. In *2009 17th International Conference on Geoinformatics* (pp. 1-5). IEEE.
- Hughes, D. A. (1982). The relationship between mean annual rainfall and physiographic variables applied to a coastal region of southern Africa. *South African Geographical Journal*, *64*(1), 41-50.
- Phillips, D. L., Dolph, J., & Marks, D. (1992). A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and forest meteorology*, *58*(1-2), 119-141.
- Price, D. T., McKenney, D. W., Nalder, I. A., Hutchinson, M. F., & Kesteven, J. L. (2000). A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest meteorology*, *101*(2-3), 81-94.
- Samanta, S., Pal, D. K., Lohar, D., & Pal, B. (2012). Interpolation of climate variables and temperature modeling. *Theoretical and Applied Climatology*, *107*(1-2), 35-45.
- Willmott, C. J., & Matsuura, K. (1995). Smart interpolation of annually averaged air temperature in the United States. *Journal of Applied Meteorology*, *34*(12), 2577-2586.
- Yang, J. S., Wang, Y. Q., & August, P. V. (2004). Estimation of land surface temperature using spatial interpolation and satellite-derived surface emissivity. *Journal of Environmental Informatics*, *4*(1), 37-44.