M-E PDG Climate File Generation

Daryl Herzmann, [akrherz@iastate.edu](mailto:akrherz@iastate.edu) 515 294 5978

Goal:

Generate historical climate files for each county in Mississippi for input into M-E PDG for a period from 1970 through 2009. Since not every county has a site with an observational record for that period, an interpolation method in space and time is used to fill in data gaps.

Procedure Overview:

1. Collect and quality control an archive of daily and hourly observations.
2. Produce an hourly gridded analysis using an interpolation method of available data.
3. Grid-point sample the hourly analysis to produce a .hcd file for M-E PDG.
4. Quality control the .hcd file to ensure no processing errors occurred.

This procedure is presented as a flow chart in Figure 1.

Figure Process for Generating Historical Climate Files

Collect observations hour-by-hour, create a gridded analysis using natural neighbor interpolation method

Step-2

Collect climate data from Climate Archive sources

Quality control check of data – omit gross errors

Quality control the gridded output to ensure data meets physical bounds

Identify missing data and create interpolated values

Create HCD file from best grid point for each County

Step-3

Step-4

Quality Control the .hcd file

Step-1

**FIGURE 1 - Process for Generating Historic Climate Files**

Detailed Procedure:

*Step 1: Acquire and quality control archived data*

The most difficult part about assembling climatic data files for M-E PDG is the requirement of hourly data. The variables needed for each hour are:

* *Air Temperature* (units: Fahrenheit), the measured air temperature at approximately 2 meters above the ground surface. The value is typically measured at the top of the hour and valid for the minute interval prior to measurement.
* *Wind Speed* (units: miles per hour), the measured speed of the air at approximately 10 meters above the ground surface. This value is typically an averaged wind speed over a two minute duration, instead of an instantaneous value.
* *Percent Sunshine* (0% is cloudy and 100% is clear), while M-E PDG labels this as “sunshine”, it can be thought of as the opposite of percent cloud cover. The technique used to measure cloud cover with the automated ASOS/AWOS sensors changed during the 40 year time interval of this study. The legacy technique, used prior to the late 1980s, was to divide the sky into octants and then count the sections covered by clouds. The present-day sensors attempt to estimate a bulk value of sky coverage producing four distinct categories of sky coverage at three altitudes. For this study, the altitude level with the most dense cloud coverage was used to produce the analysis.
* *Precipitation* (units: inches), the accumulated precipitation amount for the previous hour. For example, the value at 6 AM would represent the period of time between 5 and 6 AM.
* *Relative Humidity* (units: percent), the measured concentration of water vapor in the atmosphere versus potential capacity at the measured air temperature. This measurement is valid at the same 2 meter height of the air temperature and for the same one minute period prior to observation time. A representative physical bound of 12 to 100 percent were chosen for quality control purposes.

Some of these variables are more problematic than others considering the possible data sources available. For the time period and domain of interest for our study, there are only two types of data archives available. These are:

1. Automated Surface Observation System (ASOS): The observation platforms in this archive are primarily automated equipment located at airports. Prior to modernization efforts in the 1980s and 1990s, these observations were reported by airport personal on an hourly basis, but perhaps not during the night time hours at all locations. The larger airports had overnight manual weather observations. Since 1996, a similar type of observation system called Automated Weather Observation System (AWOS) is also used. Typically, the ASOS system is maintained by United States government entities, while the AWOS system is operated by the local state. While AWOS sites are technically different than ASOS sites, the ASOS term is broadly applied to both systems.
2. Cooperative Observer Program (COOP): The COOP observations are once daily climate observations administered by the National Weather Service. These observations are the backbone of climate monitoring in the United States. Reliable and relatively (compared with ASOS/AWOS) dense daily observations of high and low temperature and also rainfall exist for our period of interest. Observations are quality controlled by such places as National Climatic Data Center and the National Weather Service.

The Iowa Environmental Mesonet (IEM) is a data collection project within the Department of Agronomy at Iowa State University in Ames, Iowa. The IEM maintains archives of the aforementioned observation datasets. These local archives were expanded in space and time to support the domain of interest for this work by downloading data from available archives found on the Internet. The primary source of these archived datasets was the National Climatic Data Center (NCDC).

Of course, with any observation dataset, issues of data quality and quantity are of concern. While the COOP dataset provides a very high quality observation record, the daily time interval presents a challenge for use on hourly time steps. The COOP network also does not record percent sunshine, relative humidity, and wind speed. While the ASOS dataset provides all of the variables needed, it contains errors and gaps of missing observation. To clarify the data sources used and how it contributed to variable formulation, Table 1 is presented:

|  |  |  |
| --- | --- | --- |
|  | ASOS/AWOS | COOP |
| Observation Interval | Hourly | Daily |
| Air Temperature | Hourly functional form | Daily high/low conserved |
| Wind Speed | Verbatim use | N/A |
| Relative Humidity | Verbatim use | N/A |
| Percent Sunshine | Verbatim use | N/A |
| Precipitation | Hourly functional form | Daily precip conserved |

**Table 1: Variables required and their data sources**

While not exhaustive, some crude data quality checks were made to remove any data outside of reasonable physical bounds for Mississippi (i.e. temperature of 150 Fahrenheit, wind speed of 150 mph, or a precipitation amount over 200 millimeters). Since they are percent values, relative humidity and percent sunshine were bounded by values of 0 and 100.

Since the COOP network data is relatively free of data gaps and high spatial density, only observations for Mississippi were acquired. On the other hand, observations from neighboring state’s ASOS/AWOS stations were acquired to help with the analysis routine.

*Step 2: Produce an hourly and daily analysis*

Figures 2 and 3 present a map of sites in the ASOS and COOP observation networks.

|  |  |
| --- | --- |
| asos.png  Figure Present day ASOS/AWOS sites, sites used from surrounding states for the analysis are shown as well. | Figure NWS COOP sites with data between 1970-2009. |

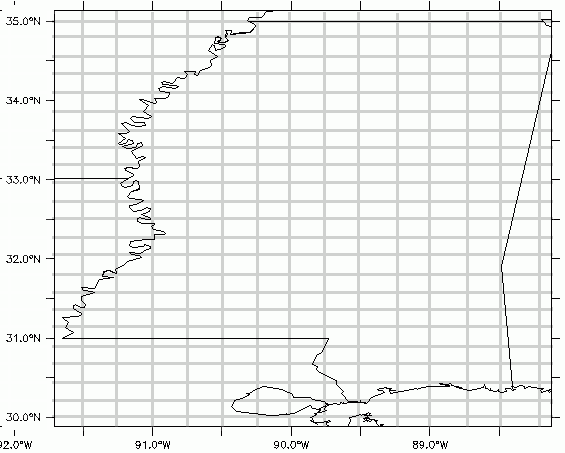
****Since not all counties of interest have both an ASOS and COOP observation point within them, a gridding technique is utilized to interpolate values spatially. A rectangular grid was constructed covering Mississippi with a grid point spacing of approximately 25 km and an exterior buffer of around 25 km. Such a grid has roughly one grid point per county in the state and is shown in Figure 4.

Figure Depiction of analysis grid. Each crosshair represents a point in space where interpolation was done on.

The gridding procedure employed a natural neighbor interpolation method (Watson 1994) which is commonly used in meteorological applications of producing a grid analysis. An illustration of this interpolation technique can be found in the Appendex Figure A. The routine works by considering the relative contribution (based on distance and directional density) of observation sites on a prescribed analysis point (or can be thought of as a cell) of interest.

So for the purposes of this study, code was developed to step through the temporal domain hour by hour and produce an analysis using the natural nearest neighbor interpolation method for the variables of interest. The code required at least four observations valid for the hour for the routine to work. If four observations were not found, the analysis was marked as missing (more on this later in the text). With the current archive available to the IEM, this error condition was met approximately 25 times (25 missing hours out of 350,640 hours). During these missing hours, all variables were typically missing.

The COOP data was gridded as well on daily time steps. The observation record of COOP sites is of very high quality and little work was necessary to account for incorrect data. The same technique as the ASOS data was used to grid the daily high and low temperature along with precipitation onto the common data grid. There were no days with less than 4 observations, so the gridding technique did not produce any missing values. In fact, each day had over 120 observations available.

*Step 3: Extract time series from gridded analysis*

With the hourly and daily gridded analysis complete, code was written to extract the values from the grid for the grid point nearest to the centroid of the county of interest. While a more sophisticated areal weighting could be done, for the purposes of this study it would not provide more accurate data due to the coarseness of the data supplied and the analysis technique’s tendency to smooth out fine-scale details.

As shown in Table 1, certain variables were constructed based on a combination of data from the ASOS and COOP archives. Since the COOP data is of much better quality than the ASOS and while the ASOS provides hourly values, an approach was utilized to take advantage of this situation. For air temperature and precipitation, the higher quality COOP data was used to adjust the lesser quality hourly ASOS data while maintaining the hourly variability. Figure 5 presented here illustrate an example result of this adjustment.

Figure An example downscaling of daily values of high and low temperature to hourly values. The observed temperature curve is stretched and compressed to match the provided daily high and low temperature.

The gridding procedure detailed in step 2 was not able to produce an analysis for each hour for the period of interest. Since the hours without data were not persistent in time and sparse in nature, a simple linear temporal interpolation was done between the hour before and the hour after with a valid analysis. More sophisticated things could have been done in place of this linear interpolator, but the suspected accuracy gain is probably not worth the effort at this time.

The result of this extraction step is an intermediate file that can be processed by M-E PDG’s climate data processor. This file has an .hcd suffix and also requires an associated entry in the station.dat file distributed with M-E PDG.

*Step 4: Quality control the .hcd file*

The .hcd file resulting from step 3 is run through a final script to ensure the data file is well formed and does not contain invalid values. The following checks are performed on the file:

1. A set of values is present for each hourly time step between 1970 and 2009. There should be 350,640 lines in the file.
2. The values in the file are free of any out of bounds values.
   1. No percentage values above 100 or below 0.
   2. No negative wind speeds or precipitation.
3. For each variable, the maximum and minimum values are reported. These numbers are visually inspected for being reasonable.
   1. Maximum air temperature should be around 100 F and minimum around 0 F.
   2. Peak wind speeds shouldn’t be much higher than 50-70 mph (typical value of a severe thunderstorm). Some sites may have hurricane data included in them, so locally higher values may happen.
   3. Maximum precipitation values should be around 2-3 inches.
4. Monthly summaries are generated and visually inspected for any suspicious values.

If errors were found, the data was investigated and the procedure started with step 2 again to correct any errors.

*The production of one .hcd file per county and the station.dat file M-E PDG uses to reference the .hcd files conclude the data processing covered in this document.*

*References:*

National Climatic Data Center, http://www.ncdc.noaa.gov

Watson, Dave, nngridr - An Implementation of Natural Neighbor Interpolation, Dave Watson Publisher, Claremont, Australia, 1994.

*Document Revision History:*

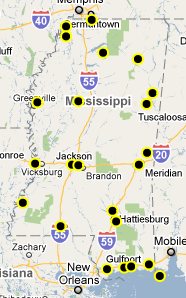
10 March 2010: Initial release for review.

13 April 2010: Second draft, removed step 4 as our end product is no longer the .icm file, but the .hcd file from step 3. This change was made to allow the local model user to specify the water table height, otherwise this value was hard coded in the .icm file delivered from this processing procedure.

14 April 2010: Implement and document a step 4 to ensure the .hcd file is free of egregious errors.

23 April 2010: Updated to include information about sites used from neighboring states.

**Appendix Figure A: Illustration of Interpolation Technique**



Analysis Grid

Neighbor

Interpolation

Grid Node