We have a mix of file formats (xls, dat, and txt). Below is a summary of the various datasets and associated notes below:

* Monthly Data
  + Temperature: 20.dat files, 4 xls files
    - Most of the data is contained in the .dat files, each of which has a header that contains descriptive information about the station (lat, long, elev, etc).
      * Missing data are indicated as “-9.9M”
      * Some of the temperature values have a letter (e.g., “I”) following the number, which needs to be deleted when imported
      * The dat files just have average temperature values, while the xls have min/max values so we need to capture all three sets of data in the final output
    - The Excel files are updates since 2010 for the .dat files so this data needs to be incorporated into the final formatted files for monthly temperature; some have station\_ID #s so they should be easy to cross-reference
      * If it’ll be easier, I can just reformat these outside of Python and append them at the end
  + Precipitation: 42 .dat files, 5 .txt files, and 3 xls files
    - Again, most of the data is contained in the .dat and .txt files, while the .xls files contains updated info from 2012 onwards
    - Missing data are listed as “-9.9M”
    - If it’ll be easier, I can reformat the xls files so that they can be imported/manipulated more easily for appending the formatted .dat monthly precip files
* Daily Data
  + Temperature: 12 .dat files
  + Precipitation: 114 .dat files
  + Jonathan reformatted the original txt files into these dat files so that should cut down on some of the work for these files
  + There is still a data header like in the monthly text files
  + A comma separates the 2 columns and missing values are noted as ‘-9.9’
  + I would like the stations’ max and min temperature files collapsed together into one dataset

We would like all of this data to be in the following format:

YEAR MONTH DAY STATION\_NAME STATION\_ID LAT LONG ELEV MEASUREMENT(UNITS)

A few notes about these columns:

* The day would be blank for monthly data
* Station\_IDs can be identified in the “Station List.doc” file
  + If python has a hard time reading this word document
* Lat Long Elev and Measurement info are in the individual data files
  + This information needs to be cross-referenced with the information in the “Station List.doc” to make sure the stations haven’t moved
  + If they have moved, then the lat, long, and elev needs to be updated accordingly
  + If it’s hard to automate this QC process, let me know and I can have this done manually and provide you with a summary of changes that need to be made
* “Measurement” would be one of the following:
  + Precipitation(mm)
  + Average Temperature (C) Maximum Temperature (C) Minimum Temperature (C)
* Please have all missing data formatted as “NaN” (right now, some are -9.9 or -9.9M as noted below)
* For your reference, I included two excel files (one for monthly and one for daily) that contains a summary of the various stations and associated dates of data
* Preferred final output is one file per station (so if a station has both precip and temp data, please collapse them together)
  + I am open to suggestions for whether these resulting text output files should be CSV or tab-separated. We will probably continue to update the monthly temperature and precipitation data periodically so is there a format that will facilitate updating these files periodically?
* Can you also create two files at the end that compile all of the monthly data into one and all of the daily data into another?