Actuaries Climate Index:

Data Sources, Calculations and Possible Improvements

January 9, 2019

Agenda

- A brief history of the ACI
- Data Inputs
- Calculations
- Possible Improvements

History

- ACI 1.0 was launched in November of 2016
- Was designed with significant support from a consulting firm with expertise in the analysis of weather data and climate trends
- Has been well-received by the actuarial community
- ACI 1.1 was introduced in April of 2019. It is identical to ACI 1.0, but with an adjustment to address a deterioration in gridded data for Canada (to be discussed subsequently)
- Now in the early stages of developing ACI 2.0

Data Inputs

- GHCNDEX gridded monthly data for precipitation and temperature:
  - Rx5day, CDD, TX10p, TN10p, TX90p, TN90p
- NCEP gridded daily wind data, which we transform into a monthly wind metric
- PSMSL station data for monthly mean sea level

Downloading the Data

- GHCNDEX gridded data can be downloaded in bulk as a NetCDF file. I use an R library to extract data for North America and store it in a CSV file.
- NCEP gridded wind data is available in separate NetCDF files for each year (1961, 1962, 1963, etc.). Each file must be downloaded separately. We then convert them to CSV format.
- PSMSL station data for monthly mean sea level can be downloaded in bulk. We then extract the data for 76 stations located in North America.

The Basic ACI Calculation Consists of Two Steps

1. Across each geographic region, and separately for each ACI component, average the unstandardized local or gridded data to create a regional-level unstandardized time series. For example, if a region has 20 grid points, then the data for these points is averaged together.

2. Standardize the time series produced in step 1 by computing the average and standard deviation across the period from 1961 to 1990, and applying these values as follows:

   Standardized value(year, month) =

   [Unstandardized(year, month) – Ref Period Average] ÷ (Ref Period Stdev)

The Australian ACI Reverses the Steps

North American ACI (standardize at regional level):

Step 1: average local data to produce regional-level time series

Step 2: standardize the regional-level time series produced in step 1

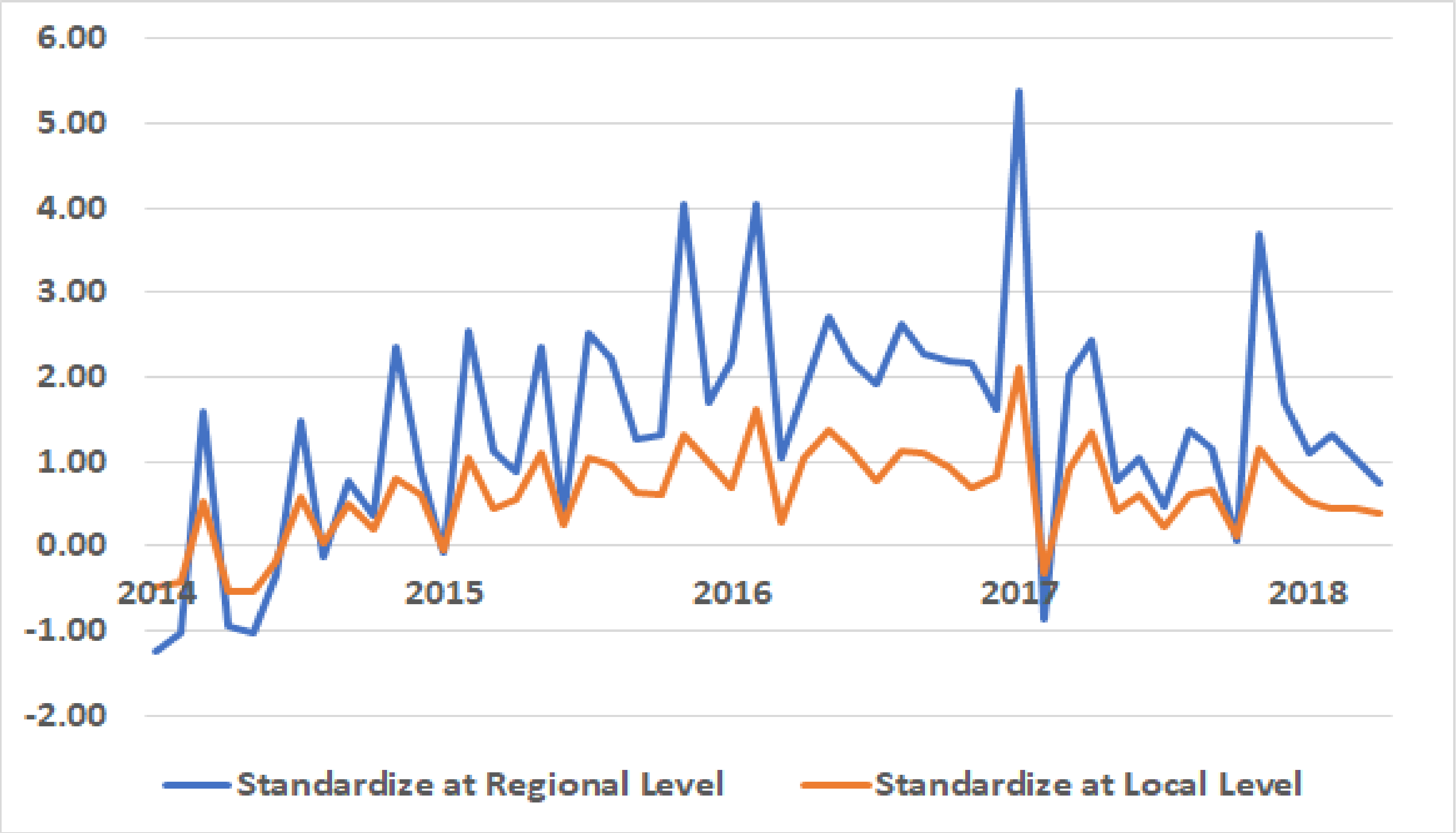
Australian ACI (standardize at local level):

Step 1: standardize local data

Step 2: average the local standardized data to produce regional-level time series

"Local" data = data for a particular weather station or grid point

Impact of Sequencing of Steps: T90, USA and Canada



Sequencing of the Two Steps

- The two approaches produce identical results if the local data points are 100% correlated, such that they move in unison like a school of fish. However, in practice, correlations are significantly less than 100%, and the lack of correlation tends to increase with the distance between local points.
- The North American ACI's calculation approach leads to regional-level time series each of which have a standard deviation of 1.0 across the reference period, while the Australians' approach leads to a set of local-level time series each of which have a standard deviation of 1.0 across the reference period.
- There is no single "right" approach, in our view.

Software Used for ACI Calculations

- The ACI calculations are programmed in Excel/VBA.

- However, the calculations are sufficiently simple that they could have been programmed using Excel cell formulas, without any VBA.

- Excel/VBA will probably not be adequate if we shift to GHCN station data because the dataset is much larger and more complicated than GHCNDEX gridded data. For our ACI 2.0 research focused on GHCN data, we have been using C++ to enable us to rapidly “digest” and tabulate the data.

Some of the Issues Under Consideration for ACI 2.0

1. Alternative approaches for aggregating data from local to regional level
2. Use station data rather than gridded data for temperature and precipitation
3. Separate nighttime and daytime temperature metrics
4. Alternative metrics/data for measuring drought
5. Alternative metrics/data for extreme wind events

How Best to Summarize Data Across a Region?

- To create regional-level time series, ACI 1.0 and 1.1 average local (grid point) data across geographic regions.
- When this approach is applied to a large region, opposing movements in weather metrics occurring in different parts of a region may “cancel out” when averaged together.
- Consider a situation in which the northern part of a region experienced a decline in average rainfall while the southern part has experienced in increase. These opposing trends can go undetected when averaged together.

Three Issues with GHCNDEX Gridded Data

1. Data Holes

2. Geographic Resolution

3. Gridded data may “mute” or compact the tails of local weather distributions

“Holes” in GHCNDEX Gridded Data

There are “holes” in the GHCNDEX data, particularly in northern Canada. By “holes”, we mean that a data field is empty for a particular grid point, year and month. Moreover, the holes increase in frequency across time, with significant deterioration in data completeness across the last five years.

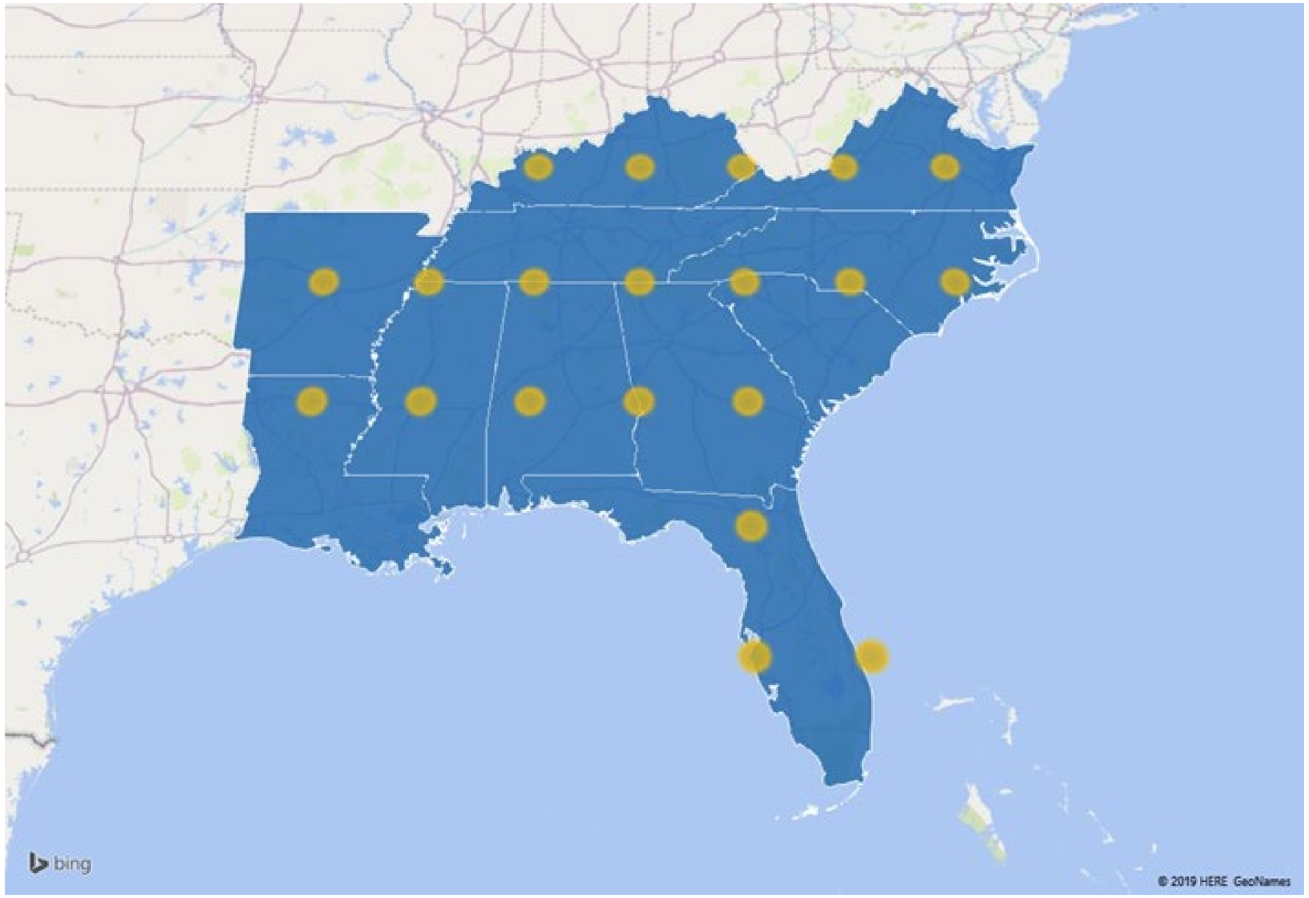
Rx5day -- # of Grid Points in Canada with Data

1961 to 1999	130
2000 to 2009	124
2011 to 2013	109
2014 to 2015	89
2016 to 2017	82
2018	65

Low Geographic Resolution of GHNDEX Gridded Data

The 2.5 by 2.5-degree gridded data doesn't provide detailed geographic resolution. Notice in the graph below that there are no grid points close to the coast of Louisiana, Mississippi, Alabama, and the Florida panhandle. Yet the Gulf Coast is perhaps the most dangerous location in the USA with respect to severe weather.

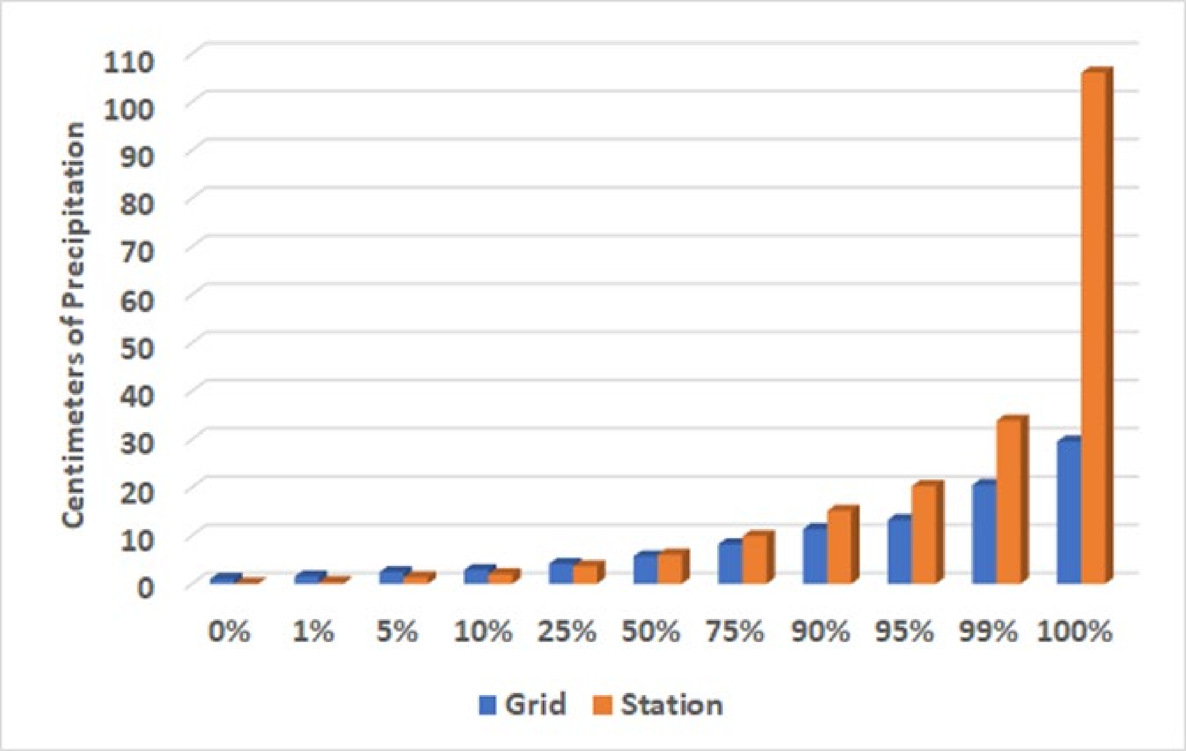
ACI Region "SEA" (Southeastern U.S.); the Red Dots Correspond to Data Grid Points



Gridded Data May Mute the Tails of Distributions

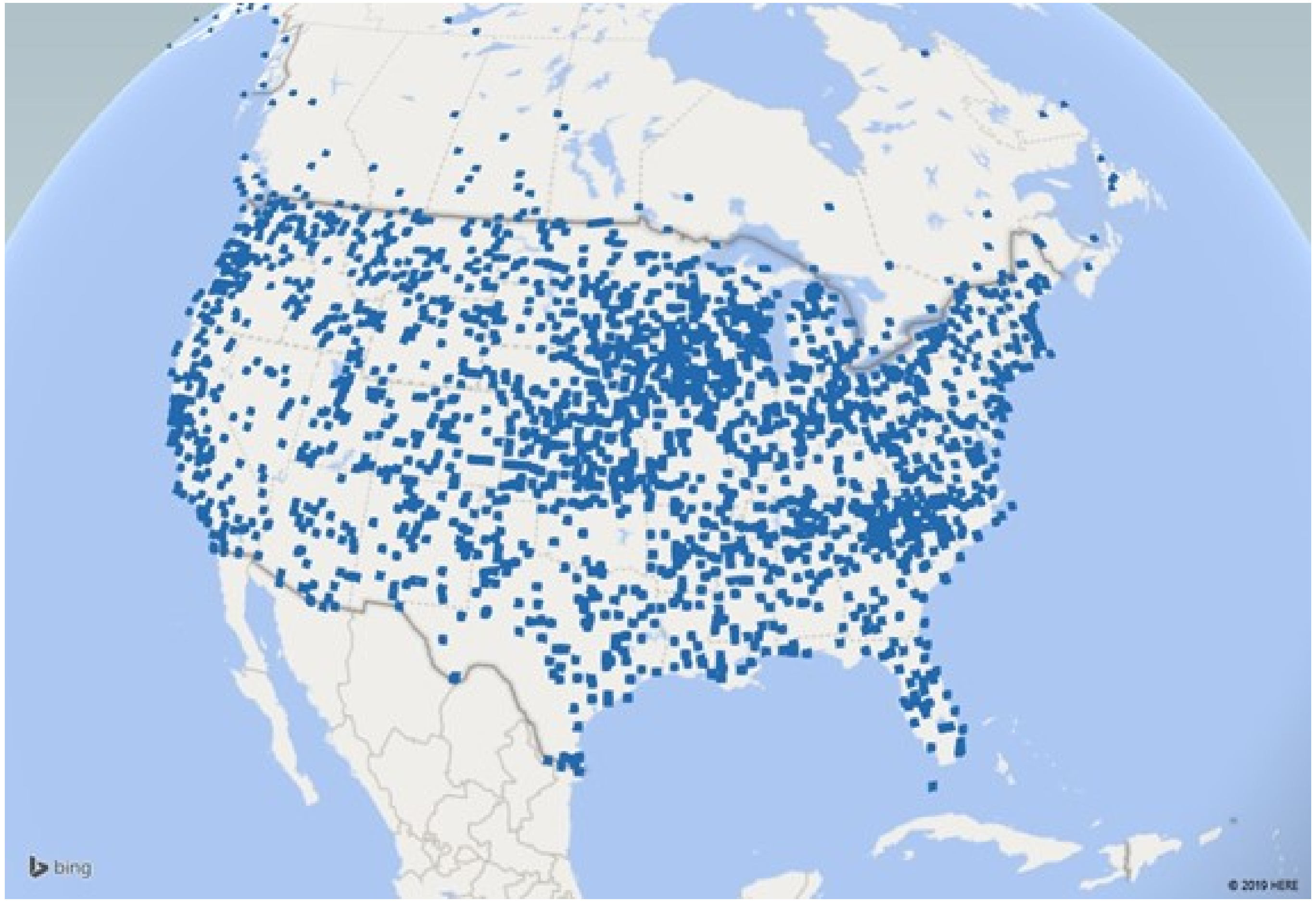
We compared gridded GHCNDEX precipitation data against corresponding data for nearby weather stations. In general, the gridded data's precipitation distribution is more compact relative to the station data, as shown in the example below. The graph shows percentiles from the Rx5day distribution, computed across the period from 1960 to the present.

Rx5day: GHCNDEX Gridded Data for 30N, 95W versus
GHCN Data for the Closest Station (Liberty, Texas at 30N, 94.7W)



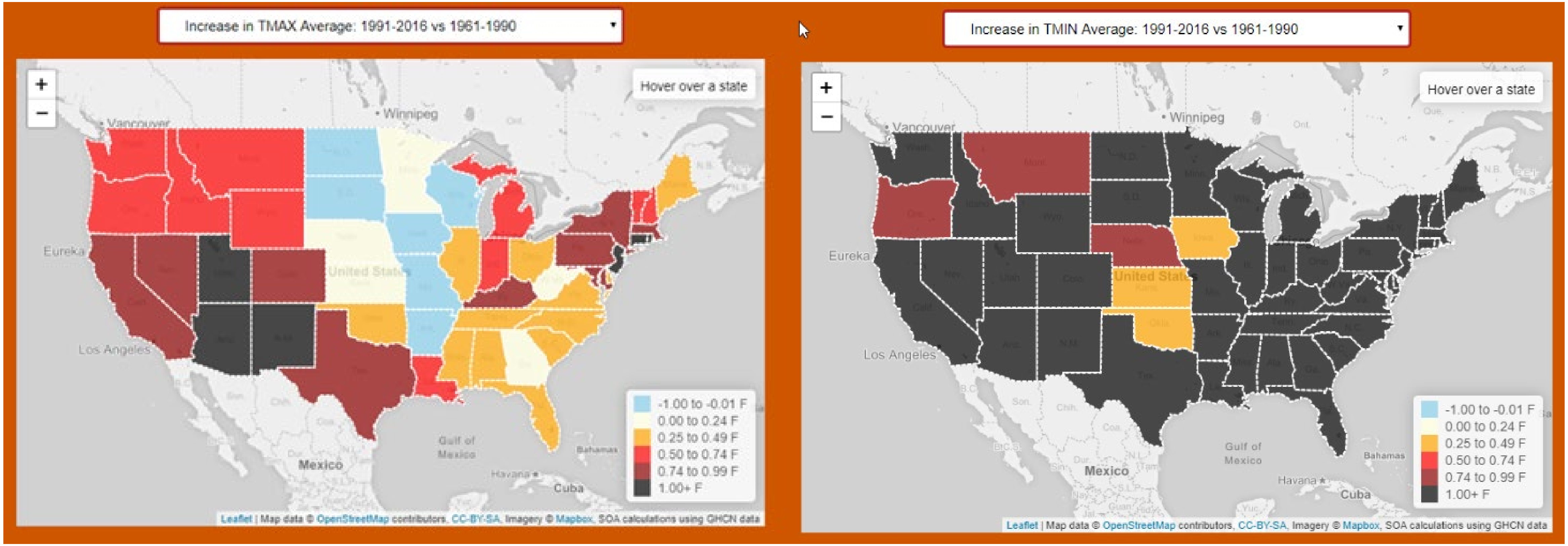
GHCN Station Data As an Alternative to Gridded Data

GHCN Weather Stations with Nearly Complete Temperature Data From 1960 to Present



Should Daytime and Nighttime Temperature Data Be Separated?

The ACI averages together daytime and nighttime data to create T10 and T90. While it is convenient to average the data together, this could potentially mask interesting information. Consider the maps below which were calculated from GHCN station data for 1960 to 2016. The graphs show that, in the USA, the upward trend in nighttime temperatures is more dramatic than for daytime temperatures.



Alternative Metrics for Measuring Drought

1. It is arguable that Consecutive Dry Days (CDD) is a poor measure of drought. In addition, the underlying data is annual rather than monthly. This limits CDD's usefulness for identifying dry months or seasons.

2. We are therefore considering other options, such as the Palmer Drought Index.

Alternative Data or Metrics for Extreme Wind Events

1. The wind data used for ACI 1.0 and 1.1 is average daily wind speed. This data may fail to distinguish between a blustery but otherwise harmless day of wind, versus a day in which a dangerous wind event occurred.

2. We are therefore searching for alternative wind data sources. Ideally, we would like data that captures peak daily wind speed (or a similar metric). However, we have not yet identified a good source for this data.