Impact of heat waves on electricity consumption and outages

Diana Rodenberger

Manuel Moreno Cabrera

Olivier Zimmer

## GOAL

* Impact of heat waves on electricity consumption and more specifically on power outages.
* Characteristics above which a heat wave is likely to be harmful to the electrical grid
* Relations between power outages and different sectors (residential vs. transportation…)
* Probabilistic risk assessment of future outages based on multiple weather scenarios.

## Electricity

The U.S. Energy Information Administration (EIA) has launched an open data initiative for researchers and policy analyst. The source is being managed by the governmental agency and is well maintained (last update June 2016). The data set contains up to 1.2 million energy series and is accessible online including through an API. The electricity data is broken down in the following set:

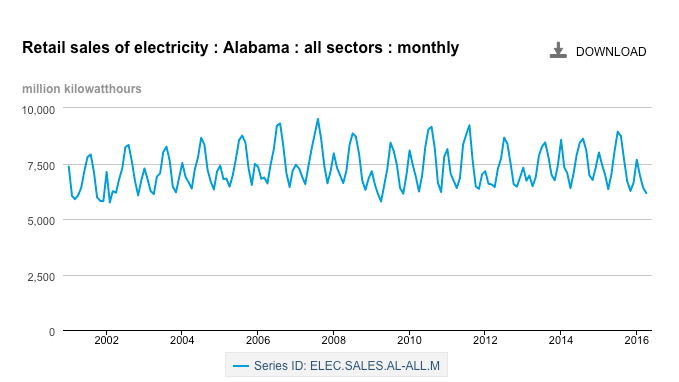
* 408,000 electricity series organized into 29,000 categories
* 30,000 State Energy Data System series organized into 600 categories

### Scope

To understand better the issue of power outages, we will be looking for datasets that permit the inference of demand, supply and price of electricity (the latter being a function of the two first plus market conditions such as source and importation). This data is available on a monthly and state level at the following locations:

1. [Retail sales](http://www.eia.gov/opendata/qb.cfm?category=38) of electricity
2. [Generation](http://www.eia.gov/opendata/qb.cfm?category=1) of electricity
3. [Price](http://www.eia.gov/opendata/qb.cfm?category=40&sdid=ELEC.PRICE.AL-ALL.M) of electricity

*Eg. Retail sales of electricity : Alabama : all sectors : monthly*



Overall, the size of the data to be acquired through the EIA API will depend on the dimensions studied.

### Dimensions

* Time range: starts on January 1st 2001 to December 2014 - 156
* States: 51
* Sector: residential, industrial, commercial, transportation - 4
* Variables: retail sales, generation, price - 3

Total data size being pulled: 112,608 rows

Estimated API calls: 612 calls

### Access - API

The advantage of using the API, is that all the data sets can be programmatically called by looping through a list of right parameters

Eg. *http://api.eia.gov/series/?api\_key=API\_KEY&series\_id=ELEC.SALES.AL-ALL.M*

Currently, we estimate that 612 discrete API calls will be required, though this may change depending on the ability to combine such calls. Our data-gathering process involves the use of an iterative Bash or Python script to save the data to an AWS instance. Once the data is uploaded onto the instance, it can be joined with the other datasets for further analysis.

## Power outage data

Inside Energy’s grid disruption dataset is a compilation of annual power outage data from the Department of Energy. This data encompasses disruptions from years 2000 to 2014. While the initial intent of this dataset was to provide insight into weather-related power outages, it includes power outages from all causes. The primary source of this data is OE-417 regulatory filings to the Department of Energy.

Limitations of this dataset include:

* Different levels of geographical granularity due to unstandardized reporting requirements
* Possible reporting gaps due to jurisdictional requirements and manual processes

### Dimensions

* 1652 discrete events, of which 866 are severe-weather related
* Time stamps for the events and resolutions
* Impact as Demand Loss (MW)
* Number of Customers affected

The dataset can be downloaded from the following location: <http://insideenergy.org/2014/08/18/data-explore-15-years-of-power-outages/>

## Weather Data

Weather data will be downloaded from the National Center for Environmental Information (NCEI) website. NCEI is the largest provider of weather and climate data. For our analyses, we will focus on the land-based database collected daily by Global Historical Climatology Network (GHCN). This integrated database contains weather data for 80,000 stations, and it covers over 180 countries and territories dating back more than 100 years.

Weather data is organized as follows:

* Metadata

o Station metadata including coordinates

o List of country codes and names

o List of U.S. state codes used in the dataset

o Periods of recorded data for each station and element

* Data

o Historical daily weather data for each station.

* Files

All files will be downloaded from ftp://[ftp.ncdc.noaa.gov/pub/data/ghcn/daily/hcn/](http://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/hcn/)

|  |  |
| --- | --- |
| File | Description |
| ghcnd-stations.txt | Station metadata including coordinates |
| ghcnd-countries.txt | List of country codes and names |
| ghcnd-inventory.txt | Periods of recorded data for each station and element |
| gncnd-states.txt | List of U.S. state and Canadian province codes used in the dataset |
| USxxxxxxxxx.dly | Daily weather data for each station. ‘xxxxxxxxx’ represents the station id. |

### Dimensions

States: 50 U.S. states, the District of Columbia, and U.S territories.

Time range: 1763 - 2016.

Weather elements: Precipitation, snow accumulation, max temperature, min temperature.

Location: State, FIPS country codes, latitude and longitude coordinates.

### Size of Data and Access

Data will be downloaded via FTP from ftp://[ftp.ncdc.noaa.gov/pub/data/ghcn/daily/](http://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/).

There are five metadata files that combined contain 30 MB of data.

All daily weather data is contained in a compressed file “ghcnd\_hcn.tar.gz” with a size of 277 MB.

Uncompressed, U.S. daily weather data consists of 1218 files with a total of 2.35 GB of data.

### Scope

For the purpose of our analysis, we will only consider weather stations in the continental U.S. for the period 1990-2014 (same period as the energy data plus 10 previous years). Metadata files will be used to identify weather stations of interest.

Each station collects daily data for multiple weather elements such as maximum temperature, minimum temperature, precipitation, snow depth, among others. Only maximum temperature and minimum temperature will be considered for the analysis.

### Calculating Heat Waves

The calculation of a heat wave requires a baseline period. In our analysis, we will use a baseline period of 10 years prior to the period of interest which is 2001- 2014. For example, to identify heat waves in 2010, we will use weather data from 1990-2000 as a baseline period and then compare temperatures in 2010 to the period 1990-2000 by week and region (zip code, city, etc). We will only consider daily maximum temperature, minimum temperature and average temperature to calculate heat waves.